



# Some problems in graph theory and graphs algorithmic theory

Stéphane Bessy

## ► To cite this version:

Stéphane Bessy. Some problems in graph theory and graphs algorithmic theory. Discrete Mathematics [cs.DM]. Université Montpellier II - Sciences et Techniques du Languedoc, 2012. tel-00806716

**HAL Id: tel-00806716**

**<https://theses.hal.science/tel-00806716>**

Submitted on 2 Apr 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Habilitation à diriger des recherches**  
présentée devant  
**L'UNIVERSITÉ MONTPELLIER II**  
**Ecole Doctorale I2S**

**M. Stéphane BESSY**

**TITLE**:- Some problems in graph theory and graphs algorithmic theory -



# Contents

<b>1</b>	<b>Overview of my research works</b>	<b>5</b>
	Foreword . . . . .	5
1.1	Introduction . . . . .	5
1.1.1	Basic definitions on graphs . . . . .	5
1.1.2	Directed graphs . . . . .	6
1.1.3	Some graphs invariants . . . . .	7
1.1.4	Algorithmic basics . . . . .	7
1.2	Problems on circuits in digraphs . . . . .	8
1.2.1	Covering by directed cycles . . . . .	8
1.2.2	Cyclic order of strong digraphs . . . . .	9
1.2.3	Packing of directed cycles . . . . .	11
1.3	Coloring and partitioning problems . . . . .	13
1.3.1	Arc-coloring in digraphs . . . . .	13
1.3.2	WDM . . . . .	14
1.3.3	Substructures in colored graphs . . . . .	16
1.4	Algorithmic problems on graphs . . . . .	18
1.4.1	Kernelization for some editing problems . . . . .	19
1.4.2	Counting edge-colorings of regular graphs . . . . .	23
	Bibliography . . . . .	24
<b>2</b>	<b>Materials</b>	<b>33</b>
2.1	Spanning a strong digraph with alpha cycles . . . . .	34
2.2	Bermond-Thomassen conjecture for regular tournaments . . . . .	40
2.3	A proof of Bermond-Thomassen conjecture for tournaments . . . . .	48
2.4	Arc-chromatic number for digraphs . . . . .	64
2.5	Wavelengths Division Multiplexing Networks . . . . .	81
2.6	A proof of Lehel's Conjecture . . . . .	105
2.7	A poly-kernel for 3-leaf power graph modification . . . . .	110
2.8	A poly-kernel for proper interval completion . . . . .	134
2.9	A poly-kernel for FAS in tournaments . . . . .	161
2.10	Counting the number of edge-colourings . . . . .	172



# Chapter 1

## Overview of my research works

### Foreword

This document is a long abstract of my research work, concerning graph theory and algorithms on graph. It summarizes some results, gives ideas of the proof for some of them and presents the context of the different topics together with some interesting open questions connected to them. This is an overview of ten selected papers which have been published in international journals or are submitted and which are included in the annex. This document is organized as follow: the first part precises the notations used in the rest of the paper; the second part deals with some problems on cycles in digraphs, a topic I am working on for almost 10 years; the third part is an overview of two graph coloring problems and one problem on structures in colored graphs; finally the fourth part focus on some results in algorithmic graph theory, mainly in parametrized complexity. I mainly worked in this last field with A. Perez who I co-supervised with C. Paul during his PhD Thesis.

To conclude, I would like to mention that this work is the result of different collaborations and each result is then a collective work, with me as common link. Doing research in graph theory is for me a great pleasure, and a job, and meeting people from various place to work with them is an also great pleasure. I would like to thank them for the nice moments we spent working together: J. Bang-Jensen, E. Birmelé, F.V. Fomin, S. Gaspers, F. Havet, C. Lepelletier, N. Lichiardopol, C. Paul, A. Perez, S. Saurabh, J.-S. Sereni and S. Thomassé.

### 1.1 Introduction

Almost all the definitions given below are standard and can be found in classical books on Graph Theory (see [47], [30] or [8]) or Parametrized Complexity Theory (see [55], [63] or [105]). We give them to precise the notations used in this document.

#### 1.1.1 Basic definitions on graphs

##### Graphs

For a set  $X$ , we denote by  $[X]^2$  the set of 2-element subsets of  $X$ . A *graph*  $G$  is a pair  $(V(G), E(G))$  consisting of a finite set  $V(G)$ , called the *vertex set* of  $G$ , and a set  $E(G)$ , subset of  $[V(G)]^2$ , called the *edge set* of  $G$ . Classically, the cardinality of  $V(G)$  and  $E(G)$  are respectively denoted by  $n(G)$  and  $m(G)$ . For notational simplicity, we write  $uv$  an unordered pair  $\{u, v\}$  of  $E(G)$ . Two vertices  $x$  and  $y$  which belong to an edge  $e$  are *adjacent*, and  $x$  and  $y$  are the *ends* of  $e$ . Furthermore, we say that  $e$  is *incident* to  $x$  and  $y$ .

The set of vertices which are adjacent to a specified vertex  $x$  is the *neighborhood* of  $x$  and is denoted by  $N_G(x)$ . So, we call a *neighbor* of  $x$  a vertex which is adjacent to it. The *degree* of a vertex  $x$ , denoted by  $d_G(x)$ , is the cardinality of its neighborhood. Finally, when two adjacent vertices  $x$  and  $y$  have the same neighborhood in  $V \setminus \{x, y\}$ , we say that  $x$  and  $y$  are *true twins*. When no confusion can occur, we will forget the reference to the background graph in all the previous notations.

A graph  $H = (V(H), E(H))$  is a *subgraph* of a graph  $G = (V(G), E(G))$  if we have  $V(H) \subseteq V(G)$  and  $E(H) \subseteq E(G)$ . If  $H$  is a subgraph of  $G$  with  $V(H) = V(G)$ , we say that  $H$  is a *spanning* (or *covering*) subgraph of  $G$ . And, if  $H$  is a subgraph of  $G$  with  $E(H) = E(G) \cap [V(H)]^2$ , we say that  $H$  is an *induced* subgraph of  $G$ . For  $X$  a subset of the vertex set of  $G$ , the *induced subgraph of  $G$  on  $X$* , denoted by  $G[X]$ , is the induced subgraph of  $G$  which has  $X$  as vertex set. We denote also by  $G \setminus X$  the induced subgraph of  $G$  on  $V(G) \setminus X$ . And, if  $F$  is a subset of the edge-set of  $G$ , we denote by  $G - F$  the subgraph of  $G$  with vertex set  $V(G)$  and edge set  $E(G) \setminus F$ . Finally, we say that two graphs  $G$  and  $H$  are *isomorphic* if there exists a bijection  $f$  from  $V(G)$  to  $V(H)$  such that for every vertices  $x$  and  $y$  of  $G$ ,  $xy \in E(G)$  if, and only if,  $f(x)f(y) \in E(H)$ . And, an *homomorphism* from  $G$  to  $H$  is a mapping  $f$  from  $V(G)$  to  $V(H)$  such that if  $xy$  is an edge of  $G$ , then  $f(x)f(y)$  is an edge of  $H$ .

In Section 1.4.2, we will deal with multi-graphs (i.e. graphs with a multiset for edge set), but anywhere else in this document, the considered graphs are simple.

### Some special graphs

The following definitions deal with some remarkable subgraphs. The *complete graph* on  $n$  vertices, denoted by  $K_n$  is the graph on vertex set  $V(K_n) = \{v_1, \dots, v_n\}$ , and with edge set  $[V(K_n)]^2$ , meaning that  $K_n$  contains all the possible edges on its vertex set. A *clique* of a graph  $G$  is a subset of its vertex set which induces on  $G$  a graph isomorphic to a complete graph. Similarly, the *empty graph* on  $n$  vertices (which is not totally empty) is the graph on  $n$  vertices and with an empty edge set. An *independent set* of a graph  $G$  is a subset of its vertex set which induces on  $G$  a graph isomorphic to an empty graph. The *path graph* on  $k$  vertices, denoted by  $P_k$  is the graph on vertex set  $V(P_k) = \{v_1, \dots, v_k\}$ , and with edge set  $\{v_i v_{i+1} : i = 1, \dots, k-1\}$ . The vertices  $v_1$  and  $v_k$  are called the *ends* of  $P_k$ . A *path* of a graph  $G$  is a subgraph of  $G$  which is isomorphic to a path graph. Finally, the *cycle graph* on  $k$  vertices, denoted by  $C_k$  is the graph on vertex set  $\{v_1, \dots, v_k\}$ , and with edge set  $\{v_i v_{i+1} : i = 1, \dots, k-1\} \cup \{v_1 v_k\}$ . A *cycle* of a graph  $G$  is a subgraph of  $G$  which is isomorphic to a cycle graph. Sometimes, we will use the term *k-cycle* to precise that the considered cycle has  $k$  vertices. A *hamiltonian* graph is a graph which admits a spanning cycle, an *acyclic graph* is a graph which contains no cycle, and a *chordal graph* is a graph with no induced cycle of size more than three. Finally, a *matching* in a graph is a set of pairwise disjoint edges of this graph.

Now, through these structures, we define some properties of graphs. First, a graph is *connected* if for every pair of vertices  $x$  and  $y$  of  $G$ , there exists a path in  $G$  with ends  $x$  and  $y$ . A *tree* is a connected graph without cycle. It is well known that a graph is connected if, and only if, it contains a spanning tree. A graph is *bipartite* if its vertex set admits a partition into two independent sets. The *complete bipartite*  $K_{p,q}$  is the graph on vertex set  $\{v_1, \dots, v_p, w_1, \dots, w_q\}$ , and with edge set  $\{v_i w_j : i = 1, \dots, p \text{ and } j = 1, \dots, q\}$ . Finally, we say that a graph  $G$  is *planar* if there exists a *plane representation* of  $G$ , i.e. a drawing of  $G$  in the plane such that its edges intersect only at their ends.

### 1.1.2 Directed graphs

A *directed graph* (or *digraph*)  $D$  is a pair  $(V(D), E(D))$  consisting of a finite set  $V(D)$ , also named the *vertex set* of  $D$ , and a subset  $A(D)$  of  $V(D) \times V(D)$ , named the *arc set* of  $D$ . For simplicity, we also denote by  $xy$  an arc  $(x, y)$  of  $D$ , but this time the order in the notation matters. We say that  $x$  is the *tail* of the arc  $xy$  and  $y$  its *head*. We obtain then a different notion of neighborhood than in the non-oriented case. For a vertex  $x$  of  $D$ , the *out-neighborhood* (resp. *in-neighborhood*) of  $x$ , denoted by  $N_D^+(x)$  (resp.  $N_D^-(x)$ ) is the set of all vertices  $y$  in  $V - x$  such that  $xy$  (resp.  $yx$ ) is an arc of  $A(D)$ . The *out-degree* (resp. *in-degree*) of a vertex  $x$ , denoted by  $d_D^+(x)$  (resp.  $d_D^-(x)$ ) is the cardinality of its out-neighborhood (resp. in-neighborhood). The notions of sub-digraph, induced and spanning sub-digraph, homomorphism and isomorphism for digraphs are similar to those from graphs. For digraphs, paths and cycles are always directed. Namely, the *(directed) path* on  $k$  vertices has vertex set  $\{v_1, \dots, v_k\}$ , and arc set  $\{v_i v_{i+1} : i = 1, \dots, k-1\}$ . The vertex  $v_1$  is the *beginning* of the path and  $v_k$  is its *end*. Similarly, the *(directed) cycle* (or *circuit*) on  $k$  vertices has vertex set  $\{v_1, \dots, v_k\}$ , and arc set  $\{v_i v_{i+1} : i = 1, \dots, k-1\} \cup \{v_k v_1\}$ . An *acyclic digraph* is a digraph which contains no (directed) cycle. A digraph  $D = (V, A)$  is *strongly connected* (or just *strong*) if there exists a path from  $x$  to  $y$  in  $D$  for every choice of distinct vertices  $x$  and  $y$  of  $D$ . A *feedback-vertex-set* (resp. *feedback-arc-set*) is a set  $X$  of vertices (resp. arcs) of  $D$  such that  $D \setminus X$  (resp.  $D - X$ ) is acyclic. The *underlying graph* of a

digraph  $D$ , denoted by  $UG(D)$ , is the (non-oriented) graph obtained from  $D$  by suppressing the orientation of each arc and deleting multiple edges.

To conclude on directed graph, a *tournament* is an orientation of a complete graph, that is a digraph  $D$  such that for every pair  $\{x, y\}$  of distinct vertices of  $D$  either  $xy \in A(D)$  or  $yx \in A(D)$ , but not both. Finally, the *complete digraph*, denoted by  $K_n^*$ , is the digraph on  $n$  vertices containing all the possible arcs, i.e. obtained from the complete graph by replacing each edge by a directed cycle of size two.

### 1.1.3 Some graphs invariants

The *independence number* of a graph (or a digraph)  $G$  is the size of a largest independent set of  $G$ . We denote it by  $\alpha(G)$ .

A *vertex-cut* in a graph  $G$  is a subset  $X$  of vertices of  $G$  such that  $G \setminus X$  is not connected. A graph  $G$  is *k-vertex-connected* if all its vertex-cut have size at least  $k$ , and the minimal size of one of its vertex-cut is the *vertex-connectivity* of  $G$  and it is denoted by  $\kappa(G)$ . This definition extends to *edge-cut*, which is a subset  $F$  of the edge-set of  $G$  such that  $G - F$  is not connected. The *edge-connectivity*, defined similarly than the vertex-connectivity, of a graph  $G$  is denoted by  $\lambda(G)$ . These definitions extend also to digraphs, where the notion of strong connectivity stands for connectivity. Remark that, given any two vertices  $x$  and  $y$  in a graph, if there exists  $p$  paths from  $x$  to  $y$  in  $G$ , vertex-disjoint except in their ends,  $x$  and  $y$ , it is not possible to find a vertex-cut of  $G$  with size less than  $p$  that separate  $x$  from  $y$ . In fact, Menger's Theorem states that this fact characterizes the vertex-connectivity of a graph.

**Theorem 1** ('Menger's Theorem', K. Menger, 1927, [103]). *Let  $G$  be graph. The vertex-connectivity of  $G$  is  $p$  if, and only if, for every pair of vertices  $x$  and  $y$  of  $G$ , there exists  $p$  paths from  $x$  to  $y$  in  $G$ , vertex-disjoint except in their ends.*

This theorem also holds for edge-connectivity and oriented case (where the paths are oriented paths).

Finally, we define the general notion of proper coloring of a graph. A *k-coloring* of a graph  $G$  is a mapping  $c$  from  $V(G)$  to the set  $\{1, \dots, k\}$  such that if  $xy$  is an edge of  $G$ , then the values  $c(x)$  and  $c(y)$  are distinct. Equivalently, a *k-coloring* of  $G$  is a vertex-partition of  $G$  into  $k$  independent sets. The *chromatic number* of  $G$ , denoted by  $\chi(G)$  is the minimum number  $k$  such that  $G$  admits a *k-coloring*. Coloring edges of graph leads to similar definitions. A *k-edge-coloring* of  $G$  is a mapping  $c'$  from  $E(G)$  to the set  $\{1, \dots, k\}$  such that if two edges  $e$  and  $f$  have a common extremity, then the values  $c'(e)$  and  $c'(f)$  are distinct. Equivalently, a *k-edge-coloring* of  $G$  is an edge-partition of  $G$  into  $k$  matchings. Similarly to the vertex case, the *edge-chromatic number* of  $G$ , denoted by  $\chi'(G)$  is the minimum number  $k$  such that  $G$  admits a *k-edge-coloring*.

### 1.1.4 Algorithmic basics

We refer to [45] for general definition on polynomially solvable problems and *NP*-complete problems. For what we need, we focus on algorithms on graphs only, and will give the definition of *FPT* algorithms in this context.

When a problem turns out to be *NP*-complete, a lot of algorithmic tools have been developed to find a acceptable solution to it, for instance, approximation algorithms, randomized algorithms, exact algorithms with exponential time... An *FPT* algorithm can be viewed as a member of this late class. The principle of such an algorithm is to contained the exponential explosion to a special parameter. Namely, a problem *parameterized* by some integer  $k$  (i.e. its input is a graph  $G$  and an integer  $k$ ) is said to be *fixed-parameter tractable* (*FPT* for short) whenever it can be solved in time  $f(k) \cdot n^c$  for some constant  $c > 0$ . As one of the most powerful technique to design fixed-parameter algorithms, *kernelization algorithms* have been extensively studied in the last decade (see [26] for a survey). A *kernelization algorithm* is a polynomial-time algorithm (called a series of *reduction rules*) that given an instance  $(G, k)$  of a parameterized problem  $P$  computes an instance  $(G', k')$  of  $P$  such that (i)  $(G, k)$  is a YES-instance if and only if  $(G', k')$  is a YES-instance and (ii)  $|G'| \leq h(k)$  for some computable function  $h()$  and  $k' \leq k$ . The instance  $(G', k')$  is called a *kernel* of  $P$ . Kernelization can be viewed as a sort of pre-processing algorithm that reduces the size of any instance. We say that  $(G', k')$  is a polynomial kernel if the function  $h()$  is a polynomial. It is well-known that a parameterized problem is *FPT* if and only if it has a kernelization algorithm [105]. But the proof



of this equivalence provides standard kernels of super-polynomial size (in the size of  $f(k)$ , precisely). So, to design efficient fixed-parameter algorithms, a kernel of small size - polynomial (or even linear) in  $k$  - is highly desirable. However, recent results give evidence that not every parameterized problem admits a polynomial kernel, unless something very unlikely happens in the polynomial hierarchy, see [27]. On the positive side, notable kernelization results include a  $2k$  kernel for VERTEX COVER [46], a  $4k^2$  kernel for FEEDBACK-VERTEX-SET [126] and a  $2k$  kernel for CLUSTER EDITING [41].

## 1.2 Problems on circuits in digraphs

In this section, we are concerned with digraphs and every concept discussed deals with directed graphs. We are interested in finding cycles in strongly connected digraphs. As trees is a cornerstone for connectivity in graphs, cycles have this central place for strong digraph. Cycles are the simplest oriented structure in which starting from any vertex it is possible to reach any other vertex.

In a strong digraph, every arc is contained in a cycle and so, there exists a family of cycles which union is the whole digraph. Problems considered here deal with finding a family of cycles in a strong digraph  $D$  which satisfies certain properties. For instance, if we want a minimum (in cardinality) family of cycles which union spans  $D$ , we obtain a *covering problem* (Section 1.2.1). One of the tools for this problem is *cyclic orders* for strong digraphs, which we develop with S. Thomassé (Section 1.2.2). Another important class of problems, the *packing problems*, deals with finding a maximum family of disjoint cycles in a digraph (Section 1.2.3). Some results which appear in the two first sections were already mentioned in my PhD Thesis. I decide to present them because they have led to some tools and problems on which I am still working.

### 1.2.1 Covering by directed cycles<sup>1</sup>

Let  $D = (V, E)$  be a strong digraph. We are mainly concerned here in finding a family  $\mathcal{F}$  of cycles of  $D$  which union covers all the vertices of  $D$ . The most natural problem in this context is to ask for a family  $\mathcal{F}$  with minimal cardinality. For long, this problem is known to be easy on tournaments, as stated by a result of P. Camion.

**Theorem 2** (*'Camion's Theorem'*, P. Camion, 1959, [35]). *Every strongly connected tournament is hamiltonian.*

For general digraphs, there is no hope to have such an exact value or even a polynomial time process to compute the minimum number of cycles needed in  $\mathcal{F}$ . Indeed, as this problem contains the HAMILTONIAN CYCLE PROBLEM for digraphs, it turns out to be *NP*-complete, and so we ask for upper bounds on the cardinality of  $\mathcal{F}$ .

Let us mention the closely related problem consisting in finding a family of paths, instead of cycles, which union covers all the vertices of a digraph. A well-known bound on the cardinality of such a family is given by a Theorem of T. Gallai and A. Milgram, which provides not only a covering by paths, but a partition of the digraph into paths.

**Theorem 3** (*'paths partition'*, T. Gallai and A. Milgram, 1960, [67]). *Every digraph  $D$  admits a vertex-partition into at most  $\alpha(D)$  paths.*

Thank to the above result on paths and Camion's theorem on tournament, T. Gallai conjectured in 1964 that the independence number could also be an upper bound for the minimum number of cycles which cover the vertices of a digraph [65]. During my PhD thesis, with S. Thomassé, we proved this conjecture and obtained the following.<sup>2</sup>

**Theorem 4** (S. Bessy, S. Thomassé, 2003, [22]). *Every strong digraph  $D$  contains a family of at most  $\alpha(D)$  cycles which union covers the vertices of  $D$ .*

In 1995, Gallai's conjecture was refined by A. Bondy [29] (stating explicitly a remark of C.C. Chen and

<sup>1</sup>This subsection and the next one are linked with the paper: S. Bessy and S. Thomassé, Spanning a strong digraph with alpha cycles: a conjecture of gallai. *Combinatorica*, 27(6):659–667, 2007, **annexed p.34**.

<sup>2</sup>I highlight the results appearing in the papers which form my habilitation.

P. Manalastas [40]) who asked for some control on the cycles. More precisely: a strong digraph  $D = (V, E)$  is a  $k$ -handle if  $k = |E| - |V| + 1$  (a 0-handle is simply a single vertex). A *handle* is a directed path  $H := x_1, \dots, x_l$  in which we allow  $x_1 = x_l$ . The vertices  $x_1$  and  $x_l$  are the *extremities* of the handle  $H$ , and its other vertices are its *internal* vertices. For a subdigraph  $H$  of  $D$ , an  $H$ -handle is a handle of  $D$  with its extremities in  $V(H)$  and its internal vertices disjoint from  $H$ . Finally, a *handle basis* of  $D$  (or *ear decomposition*, see [8]) is a sequence  $H_0, H_1, \dots, H_k$  of handles of  $D$  such that  $H_0$  is a single vertex,  $H_i$  is a  $(\cup\{H_j : j < i\})$ -handle for all  $i = 1, \dots, k$  and  $D = \cup\{H_i : i = 0, \dots, k\}$ . Clearly, a digraph has a handle basis  $H_0, \dots, H_k$  if and only if  $D$  is a  $k$ -handle. In this context the conjecture of A. Bondy is the following.

**Conjecture 5** (*'Bondy's Conjecture'*, A. Bondy, 1995, [29]). The vertices of every strong digraph  $D$  can be covered by the disjoint union of some  $k_i$ -handles, where  $k_i > 0$  for all  $i$ , and the sum of the  $k_i$  being at most  $\alpha(D)$ .

As it is possible to cover every  $k$ -handle by  $k$  cycles, Bondy's conjecture is stronger than the result of Theorem 4. For  $k = 1$ , Bondy's conjecture is true by Camion's Theorem. Furthermore, for  $k = 2$ , it has been solved by C.C. Chen and P. Manalastas [40], and by S. Thomassé for  $k = 3$  [125]. The techniques used in the proof of Theorem 4 seem to be useless to tackle Bondy's Conjecture. However, using a different approach, with S. Thomassé, we obtained a result closely related to this conjecture.

**Theorem 6** (S. Bessy, S. Thomassé, 2003, [21]). *Every strong digraph  $D$  is spanned by a  $k$ -handle, with  $k \leq 2\alpha(D) - 1$ .*

In other words, every strong digraph  $D$  admits a spanning strong subdigraph with at most  $n + 2\alpha(D) - 2$  arcs. The problem of finding such a subdigraph with a minimum number of arcs is a classical problem in graph theory, named the MSSS PROBLEM (for minimal strong spanning subdigraph). This problem is also NP-complete (it also contains the HAMILTONIAN CYCLE PROBLEM) and the best known approximation algorithm, found by A. Vetta in 2001 [131], achieves a  $\frac{3}{2}$  factor of approximation. A nice question is to look at what happens on some particular classes of digraphs, and particularly the following one.

**Problem 7.** Is there an approximation algorithm for the MSSS PROBLEM with a better factor than  $\frac{3}{2}$ ? What about if we restrict the instances to the class of planar strong digraphs?

### 1.2.2 Cyclic order of strong digraphs

In this section is presented a tool that we developed in order to prove Theorem 4. The notion of cyclic order allows, in some sense, cyclic statements of classical results on paths a digraphs.

Let  $D$  be a strong digraph on vertex set  $V$ . If  $E = v_1, \dots, v_n$  is an enumeration of  $V$ , for any  $k \in \{2, \dots, n\}$ , the enumeration  $v_k, \dots, v_n, v_1, \dots, v_{k-1}$  is obtained by *rolling*  $E$ . Two enumerations of  $V$  are *equivalent* if we can pass from one to the other by a sequence of the following operations: rolling and exchanging two consecutive but not adjacent vertices. The classes of this equivalence relation are called the *cyclic orders* of  $D$ . Roughly speaking, a cyclic order is a class of enumerations of the vertices on a circle, where one stay in the class while switching consecutive vertices which are not joined by an arc. We fix an enumeration  $E = v_1, \dots, v_n$  of  $V$ , the following definitions are understood with respect to  $E$ . An arc  $v_i v_j$  of  $D$  is a *forward arc* if  $i < j$ , otherwise it is a *backward arc*. With respect to  $E$ , the *index* of a cycle  $C$  of  $D$  is the number of backward arcs containing in  $C$ , we denote it by  $i_E(C)$ . This corresponds to the winding number of the cycle. Observe that  $i_E(C) = i_{E'}(C)$  if  $E$  and  $E'$  are equivalent. Consequently, the index of a cycle is invariant in a given cyclic order  $\mathcal{O}$  and we denote it by  $i_{\mathcal{O}}(C)$ . A circuit is *simple* if it has index one. A cyclic order  $\mathcal{O}$  is *coherent* if every arc of  $D$  is contained in a simple circuit.

The following lemma states the existence of coherent cyclic orders. It is proved by considering a cyclic order which minimizes the sum of the cyclic indices of all the cycles of  $D$ .

**Lemma 8** (D.E. Knuth, 1974, [91] / S. Bessy, S. Thomassé, 2007, [22]). *Every strong digraph has a coherent cyclic order.*

As mentioned by A. Bondy in [30], this result was found originally by D.E. Knuth in 1974 [91] (we ignored

it when we settled Lemma 8), in a different context, and no link with Gallai's conjecture have been done.

Now, we describe two min-max relations in context cyclic orders. As previously said, cyclic orders can be understood as a cyclic version of classical theorems on paths in digraphs. First, we give a cyclic version of Gallai-Milgram's paths partition Theorem. Given  $\mathcal{O}$  a cyclic order of a strong digraph  $D$ , we denote by  $\alpha(\mathcal{O})$  the size of a maximum *cyclic independent set* of  $\mathcal{O}$ , that is an independent set of  $D$  which is consecutive in an enumeration of  $\mathcal{O}$ . The following theorem provides a family of cycles which cover all the vertices of the considered digraph  $D$ .

**Theorem 9** (S. Bessy, S. Thomassé, 2007, [22]). *Let  $D$  be a strong digraph with a coherent cyclic order  $\mathcal{O}$ . The minimal  $\sum_{C \in \mathcal{R}} i_{\mathcal{O}}(C)$ , where  $\mathcal{R}$  is a spanning set of cycles of  $D$  is equal to  $\alpha(\mathcal{O})$ .*

The proof uses a very basic algorithmic process. Let us briefly explain it. We start by computing greedily a cyclic independent set  $X$  of  $\mathcal{O}$  and consider an enumeration  $E$  of  $\mathcal{O}$  where  $X$  stands at the beginning of  $E$ . Then, in a digraph build from the transitive closure of the acyclic digraph formed by the forward arcs of  $E$ , we apply Dilworth's Theorem [48] (a classical version of Gallai-Milgram's Theorem for orders). So, either we find a larger cyclic independent set than  $X$ , and we go on the process, or we find a set of  $|X|$  paths covering this digraph, and we stop and show that these paths can be turned into cycles covering  $D$ .

Remark that, as  $\alpha(\mathcal{O})$  is the size of an independent set of  $D$ , we find a set of at most  $\alpha(D)$  cycles which covers the vertices of  $D$ . This gives a proof of Gallai's Conjecture. However, we have no control on the number of arcs involved in this covering and then no bound on the sum of the  $k$ -handles needed to cover  $D$ , as asked by Bondy's Conjecture. On the other hand, in the proof of Theorem 9, as previously explained, we obtained a family  $\mathcal{F}$  of  $k$  cycles, with  $k \leq \alpha(D)$ , and a set  $X$  of  $k$  vertices such that the union of the cycles of  $\mathcal{F}$  minus  $X$  forms an acyclic digraph  $D'$ . In  $D'$ , the cycles of  $\mathcal{F}$  become paths. So, the main challenge to attempt resolving Bondy's Conjecture could be the following.

**Problem 10.** Is it possible to 'uncross' the  $k$  paths in  $D'$  in order to reduce their length and, then, the total number of arcs involved in the covering?

The second min-max theorem which we found in the field of cyclic order can be viewed as a cyclic version of the Gallai-Roy's Theorem.

**Theorem 11** ('The Gallai-Roy Theorem', T. Gallai, 1966, [66] and B. Roy, 1967, [114]). *Every digraph  $D$  contains a directed path on  $\chi(D)$  vertices.*

The cyclic version is the following. Given an enumeration  $E = v_1, \dots, v_n$  of the vertices of a digraph  $D$ , a *coloring* of  $E$  into  $r$  colors is a partition of  $V$  into  $r$  sets  $V_1, \dots, V_r$  such that for every  $j$ ,  $V_j$  are an independent set of  $D$  and  $V_j$  are consecutive on  $E$  (i.e. there exist integers  $i_0 = 0 < i_1 < \dots < i_r = n$  such that  $V_j = \{v_{i_{j-1}+1}, \dots, v_{i_j}\}$  for all  $j \in \{1, \dots, r\}$ ). The *chromatic number* of  $E$  is the minimum value  $r$  for which there exists a  $r$ -coloring of  $E$ . For a cyclic order  $\mathcal{O}$  of  $D$ , the *cyclic chromatic number* of  $\mathcal{O}$ , denoted by  $\chi(\mathcal{O})$  is the minimum value of the chromatic number of an enumeration belonging to  $\mathcal{O}$ . Finally, for a cycle  $C$  of  $D$  and a cyclic order  $\mathcal{O}$  of  $D$ , the *cyclic length* of  $C$  is the value  $|C|/i_{\mathcal{O}}(C)$ . We have the following min-max relation.

**Theorem 12** (S. Bessy, S. Thomassé, 2007, [22]). *Let  $\mathcal{O}$  be a coherent cyclic order of a strong digraph  $D$ . The maximal  $\lfloor l_{\mathcal{O}}(C) \rfloor$ , where  $C$  is a circuit of  $D$  is equal to  $\chi(\mathcal{O})$ .*

There even exists a fractional version of this theorem (see [22]). It is similar (but oriented) to the classical result on the circular chromatic number for non-oriented graphs (see the survey of X. Zhu [137], for instance). As a corollary of Theorem 12, we obtain a classical theorem of A. Bondy.

**Theorem 13** (A. Bondy, 1976, [28]). *Every strong digraph  $D$  contains a cycles on at least  $\chi(D)$  vertices.*

To conclude this section, remark that we established some results on cyclic orders using graph theoretical tools. However, there exists proofs of these results (and even others) obtained by techniques from linear programming or polyhedral combinatorial optimization (see the work of P. Charbit and A. Sebö [37] and

A. Sebö [117]). In particular, using these techniques, A. Sebö gave in [117] a cyclic version of Menger's Theorem (we missed it...). Let  $\mathcal{O}$  be a cyclic order of a strong digraph  $D$ . A *cyclic feedback-vertex-set* of  $\mathcal{O}$  is a set  $U$  of vertices of  $D$  such that for every cycle  $C$  of  $D$ , we have  $|V(C) \cap U| \geq i_{\mathcal{O}}(C)$ . A. Sebö establishes the following.

**Theorem 14** (A. Sebö, 2007, [117]). *Let  $\mathcal{O}$  be a cyclic order of a strong digraph  $D$ . The minimum cardinality of a cyclic feedback-vertex-set is equal to the maximum of  $\sum_{C \in \mathcal{R}} i_{\mathcal{O}}(C)$ , where  $\mathcal{R}$  is a set of vertex-disjoint cycles of  $D$ .*

A. Sebö gives also in [117] an 'arc version' of this result and weighted analogous of these two statements. Obviously, we can wonder if there exists other theorems on paths which can be turn into a cyclic form.

**Problem 15.** Is there other results on paths in digraphs which admits a cyclic equivalent?

### 1.2.3 Packing of directed cycles<sup>3</sup>

In this section, we are concerned about a converse problem (in a certain way) of the previous one. Given a digraph  $D$ , we denote by  $\nu_0(D)$  (resp.  $\nu_1(D)$ ) the maximum number of vertex-disjoint (resp. arc-disjoint) circuits in  $D$ . The problems dealing with *circuits packing* in digraphs consist in computing, or finding bounds on  $\nu_0$  and  $\nu_1$ .

In addition, we define  $\tau_0(D)$  (resp.  $\tau_1(D)$ ) to be the minimum size of a feedback-vertex-set of  $D$  (resp. feedback-arc-set of  $D$ ). It is clear that  $\tau_0$  (resp.  $\tau_1$ ) is a natural lower bound of for the packing number  $\nu_0$  (resp.  $\nu_1$ ). The converse is not true, but it is possible to bound above  $\nu_0$  (resp.  $\nu_1$ ) by a function of  $\tau_0$  (resp.  $\tau_1$ ). This statement was conjectured in 1973 by D.H Younger [136], and the first case of this conjecture 'is  $\tau_0$  bounded when  $\nu_0 = 1$ ?', was solved by W. McCuaig.

**Theorem 16** (W. McCuaig, 1991, [102]). *If  $D$  is a digraph with no two vertex-disjoint cycles, then there exists a set  $X$  of at most 3 vertices such that  $D \setminus X$  is acyclic.*

In fact, some years later, in 1996, Younger's Conjecture was settled by B. Reed, N. Robertson, P.D. Seymour and R. Thomas who proved the following.

**Theorem 17** ('Younger's Conjecture', B. Reed, N. Robertson, P.D. Seymour and R. Thomas, 1996, [112]). *There exist functions  $f_0, f_1 : \mathbb{N} \rightarrow \mathbb{N}$  such that for every digraph  $D$  we have  $\tau_0(D) \leq f_0(\nu_0(D))$  and  $\tau_1(D) \leq f_1(\nu_1(D))$ .*

More precisely, to prove Theorem 16, W. McCuaig gave a complete characterization of digraphs with no two disjoint cycles. To prove Theorem 17, the authors used Ramsey Theory, leading to exponential functions for  $f_0$  and  $f_1$ . We can ask if there is possible other ways to prove these two theorems. In particular, as cyclic orders have strong link with cycles in digraphs, they could be useful for that.

**Problem 18.** Is it possible to prove Theorem 16 or Theorem 17 using cyclic orders of digraphs, and then, obtaining simpler proofs or better bounds on  $f_0$  and  $f_1$ ?

For the end of this subsection, we focus on different lower bounds for the maximum number of vertex-disjoint cycles in a digraph (i.e.  $\nu_0$ ). By Theorem 17, we know that if  $\tau_0$  is large enough, then  $\nu_0$  will be large also. There is no hope to easily obtain an exact value for  $\tau_0$ , as it is known for long that computing  $\tau_0$  is an NP-hard problem (it has been proved by R.M. Karp in 1972 [85]). However, there is a natural and tractable lower bound for  $\tau_0$ . For a digraph  $D$ , the *minimum out-degree* (resp. *minimum in-degree*) of  $D$ , denoted by  $\delta^+(D)$  (resp.  $\delta^-(D)$ ), is the minimum value over the out-degrees (resp. in-degrees) of the vertices of  $D$ . As there exists a set  $X$  of  $\tau_0(D)$  vertices of  $D$  such that  $D \setminus X$  is acyclic, there is a vertex  $x$  with out-degree 0 in  $D \setminus X$  and thus, we have  $\delta^+(D) \leq d^+(x) \leq \tau_0(D)$ . So, using Theorem 17, we know that if  $\delta^+$  is large enough,  $\nu_0$  will be also large. This corollary of Theorem 17, was, in fact, directly proved by C. Thomassen in 1983 [127]. With J.C. Bermond, they raised a conjecture on the minimum value of  $\delta^+$  which insures  $\nu_0 \geq k$ .

<sup>3</sup>This subsection is linked with the papers: S. Bessy, N. Lichiardopol, and J.S. Sereni, Two proofs of the bermond-thomassen conjecture for tournaments with bounded minimum in-degree. *Discrete Mathematics*, 310:557–560, 2010, **annexed p.40** and J. Bang-Jensen, S. Bessy, and S. Thomassé, Disjoint 3-cycles in tournaments: a proof of the bermond-thomassen conjecture for tournaments, *submitted*, 2011, **annexed p.48**.

**Conjecture 19** (*'The Bermond Thomassen Conjecture'*, J.C. Bermond and C. Thomassen, 1981, [13]). If  $\delta^+(D) \geq 2k - 1$  then  $\nu_0(D) \geq k$ , what means that  $D$  contains at least  $k$  vertex-disjoint cycles.

Remark that the complete digraph (with all the possible arcs) is a sharp example for this statement. The conjecture is trivial for  $k = 1$  and it has been verified for general digraphs when  $k = 2$  by C. Thomassen [127] and  $k = 3$  by N. Lichiardopol, A. Pór and J.-S. Sereni [96]. Furthermore, N. Alon proved in 1996 that there exists a linear function of  $k$  that insure  $\nu_0 \geq k$ . Namely, using some probabilistic arguments, he proves the following.

**Theorem 20** (N. Alon, 1996, [3]). *If  $\delta^+(D) \geq 64k$  then  $\nu_0(D) \geq k$ .*

The status of the Bermond Thomassen Conjecture was even not known on tournaments. I have worked on this specific problem. In this case, as any cycle in tournament always contains a 3-cycle, we focus on disjoint 3-cycles. First, in 2005, with N. Lichiardopol and J.S. Sereni, we proved that the Bermond Thomassen Conjecture is true for regular tournaments. More precisely, we obtained the following result.

**Theorem 21** (S. Bessy, N. Lichiardopol and J.S. Sereni, 2005, [18]). *If  $T$  is a tournament with  $\delta^+(T) \geq 2k - 1$  and  $\delta^-(T) \geq 2k - 1$  then,  $\nu_0(T) \geq k$ .*

More precisely, we proved that, given a collection  $\mathcal{F}$  of  $t < k$  disjoint 3-cycles of  $T$ , it is always possible to find a 3-cycle  $C$  of  $\mathcal{F}$  such that  $T[(V(T) \setminus V(\mathcal{F})) \cup V(C)]$  contains two disjoint 3-cycles. Then, removing  $C$  from  $\mathcal{F}$  and adding these two 3-cycles, we obtain a collection of  $t + 1$  disjoint 3-cycles of  $T$ . Unfortunately, this scheme of proof does not work if we remove the condition  $\delta^-(T) \geq 2k - 1$ . However, some years later, in 2010, during a second attempt with J. Bang-Jensen and S. Thomassé, we finally proved the Bermond Thomassen Conjecture for tournaments.

**Theorem 22** (J. Bang-Jensen, S. Bessy and S. Thomassé, 2010, [7]). *Every tournament  $T$  with  $\delta^+(T) \geq 2k - 1$  has  $k$  disjoint cycles each of which have length 3.*

Here, the proof is also based on the possibility to extend a family of disjoint 3-cycles. More precisely, given a collection  $\mathcal{F}$  of  $t < k$  disjoint 3-cycles of  $T$ , we proved that is always possible to find a larger family of disjoint 3-cycles intersecting  $V(T) \setminus V(\mathcal{F})$  on a most four vertices. This method is similar to the one used in the proof of Theorem 21, but it allows more recombination possibilities on the 3-cycles when enlarging  $\mathcal{F}$ .

As previously mentioned, Bermond Thomassen Conjecture is sharp on complete digraphs. But for large  $k$ , we did not find any sharp example of this statement for tournaments, and then without 2-cycles. Indeed, such examples do not exist for tournament as we have shown by improving Theorem 22 for tournaments with large minimum out-degree. Roughly speaking, a tournament  $T$  with  $\delta^+(T) > \frac{3}{2}k$  and  $k$  large enough contains  $k$  disjoint cycles of length 3. More precisely, we proved the following.

**Theorem 23** (J. Bang-Jensen, S. Bessy and S. Thomassé, 2010, [7]). *For every value  $\alpha > \frac{3}{2}$ , there exists a constant  $k_\alpha$ , such that for every  $k \geq k_\alpha$ , every tournament  $T$  with  $\delta^+(T) \geq \alpha k$  has  $k$  disjoint 3-cycles.*

This statement is optimal for the value  $\frac{3}{2}$ , as shown by the family of regular tournament, i.e. tournament  $T$  that verify  $d^+(x) = d^-(x)$  for every vertices of  $T$ . However, we do not know what happens for tournaments with  $\delta^+ = \frac{3}{2}k$ , but we conjecture that they also contain  $k$  disjoint 3-cycles. As when we forbid small cycles (of length 2) we can asymptotically improve the statement of the Bermond Thomassen Conjecture on tournaments and we conjecture that this could be also true for general digraphs.

**Conjecture 24** (J. Bang-Jensen, S. Bessy and S. Thomassé, 2010, [7]). *If a digraph  $D$  has no cycles of length less than  $g$  and minimum out-degree at least  $k$ , with  $k$  large enough, then  $D$  contains at least  $\frac{g+1}{g} \cdot k$  disjoint cycles.*

To conclude this subsection, I just would like to mention two nice conjectures dealing with feedback-arc-set in digraphs. As previously mentioned, we know that  $\tau_1$  is *NP*-hard to compute for digraph, and even for tournaments [9]. However there is a class of digraphs where  $\tau_1$  is not hard to compute. Indeed, for planar

digraph, C.L. Lucchesi proved in 1976 [98] that is possible to compute a feedback-arc-set on a planar digraph in polynomial time. Furthermore, always for planar digraphs, the Lucchesi-Younger Theorem [99] asserts that  $\tau_1 = \nu_1$ . But even in the planar case, much remains unknown on feedback arc set, as shown by these two long-standing conjectures.

**Conjecture 25** ((*weak form*) D.R. Woodall, 1978, [134]). Every planar strong digraph with no 2-cycles admits three disjoint feedback-arc-sets.

**Conjecture 26** (V. Neumann-Lara, 1982, [104]). Every planar digraph  $D$  with no 2-cycles has a feedback-arc-set which forms a bipartite subdigraph of  $D$ .

## 1.3 Coloring and partitioning problems

During my research works, I also focused on some graph coloring problems or problems in colored graphs. These topics are structural and can be viewed as partitioning questions, as those from Section 1.2. However, they do not present such a unity and come from really different fields. The two first subsections deal with some questions arising in a modeling context. In Subsection 1.3.1, we use an arc-coloring model to obtain results on a function theory problem, and Subsection 1.3.2 deals with questions from a classical application of graph coloring theory: optimization in optical communication networks. Finally, in Subsection 1.3.3, we are interested in a problem from a slightly different context. Given a graph and a coloring of this graph, we look for a subgraph having some structural properties (e.g. a path, a cycle...) and some properties according to the coloring (e.g. monochromatic, bi-chromatic...).

### 1.3.1 Arc-coloring in digraphs<sup>4</sup>

Related to a function theory problem, with É. Birmelé and F. Havet, we have studied an extension of graph coloring to digraphs. The problem was initially proposed by A. El Sahili [58] and comes from a question arising in function theory. Namely, let  $f$  and  $g$  be two maps from a finite set  $A$  into a set  $B$ . Suppose that  $f$  and  $g$  are *nowhere coinciding*, that is for all  $a \in A$ ,  $f(a) \neq g(a)$ . A subset  $A'$  of  $A$  is  $(f, g)$ -*independent* if  $f(A') \cap g(A') = \emptyset$ . We are interested in finding the minimum number of  $(f, g)$ -independent subsets needed to partition  $A$  in the case where every element of  $B$  has a bounded number of antecedents by the functions  $f$  and  $g$ . As shown by El-Sahili [59], this can be translated into an arc-coloring problem.

We focus on a special type of arc-coloring for digraphs, introduced by S. Poljak and V. Rödl in 1981 [108]. Other classical arc-colorings exist, see [72] for instance, but this one model the previous problem. So, here, an *arc-coloring* of a digraph  $D$  is an application  $c$  from the arc-set  $A(D)$  into a set of colors  $S$  such that if the tail of an arc  $e$  is the head of an arc  $e'$  then  $c(e) \neq c(e')$ . In other words, the arcs from a same color class form a bipartite graph and are oriented from a part of the bipartition to the other one. The *arc-chromatic number* of  $D$ , denoted by  $\chi_a(D)$ , is the minimum number of colors used by an arc-coloring of  $D$ . Another way to define the notion of arc-coloring is the following: in an arc-coloring of  $D$ , for any arc  $xy$ , the set of colors appearing on the arcs with tail  $x$  must not be a subset of the set of colors appearing on the arcs with tail  $y$  (as it contains the color of  $xy$ ). We denote by  $\overline{H}_k$  the complementary of the hypercube of dimension  $k$ , i.e.  $\overline{H}_k$  is the digraph with vertex set all the subsets of  $\{1, \dots, k\}$  and with arc set  $\{XY : X \not\subseteq Y\}$ . With the previous remark, a digraph  $D$  has an arc-coloring with  $k$  colors if, and only if, it admits an homomorphism to  $\overline{H}_k$  (then, if an arc  $xy$  of  $D$  is mapped to an arc  $XY$  of  $\overline{H}_k$ , we color  $xy$  with an integer of  $X \setminus Y$ ). Using this and Sperner's Lemma [123] to find a homomorphism into a complete subdigraph of  $\overline{H}_k$ , S. Poljack and V. Rödl obtained the following theorem.

**Theorem 27** (S. Poljak and V. Rödl, 1981, [108]). *For every digraph  $D$ , we have  $\log(\chi(D)) \leq \chi_a(D) \leq \theta(\chi(D))$ , where  $\theta(k) = \min\{s : \binom{s}{\lfloor s/2 \rfloor} \geq k\}$ .*

Now, we come back to the function theory problem and the model proposed by A. El-Sahili in [59]. Let  $D_{f,g}$  be the digraph defined by:  $V(D_{f,g}) = B$  and  $(b, b') \in E(D_{f,g})$  if there exists an element  $a$  in  $A$  such that

<sup>4</sup>This subsection is linked with the paper: S. Bessy, E. Birmelé, and F. Havet, Arc-chromatic number of digraphs in which every vertex has bounded outdegree or bounded indegree, *Journal of Graph Theory*, 53(4):315–332, 2006, **annexed p.64**.

$g(a) = b$  and  $f(a) = b'$ . Then, a  $(f, g)$ -independent subset of  $A$  corresponds to a set of arcs of  $D_{f,g}$  which do not form paths of length more than one or cycles. And so, the minimum number of  $(f, g)$ -independent subsets needed to partition  $A$ , denoted by  $\phi(f, g)$  is exactly the arc-chromatic number of  $D_{f,g}$ . Furthermore, we want to take into consideration in the model the number of antecedents by  $f$  and  $g$  for the elements of  $B$ . Precisely, let  $\Phi(k)$  (resp.  $\Phi^\vee(k, l)$ ) be the maximum value of  $\phi(f, g)$  for two nowhere coinciding maps  $f$  and  $g$  from  $A$  into  $B$  such that for every  $z$  in  $B$ ,  $|g^{-1}(z)| \leq k$  (resp. either  $|g^{-1}(z)| \leq k$  or  $|f^{-1}(z)| \leq l$ ). The condition  $f^{-1}(z)$  (resp.  $g^{-1}(z)$ ) has at most  $k$  elements means that each vertex has in-degree (resp. out-degree) at most  $k$  in  $D_{f,g}$ . To turn these notions into digraphs context, A. El-Sahili [59] defines a  $k$ -digraph to be a digraph in which every vertex has out-degree at most  $k$ . Similarly, a  $(k \vee l)$ -digraph is a digraph in which every vertex has either out-degree at most  $k$  or in-degree at most  $l$ . Hence,  $\Phi(k)$  (resp.  $\Phi^\vee(k, l)$ ) is the maximum value of  $\chi_a(D)$  for  $D$  a  $k$ -digraph (resp. a  $(k \vee l)$ -digraph).

So, motivated by the previous interpretation in function theory and by the corresponding coloring problem, we studied the behavior of the functions  $\Phi$  and  $\Phi^\vee$ . The first results on these functions was given by A. El-Sahili, who proved the following.

**Theorem 28** (A. El-Sahili, 2003, [59]). *We have  $\Phi^\vee(k, k) \leq 2k + 1$ .*

Using Theorem 27, we improved this bound to the following.

**Theorem 29** (S. Bessy, É. Birmelé and F. Havet, 2006, [14]). *We have  $\Phi(k) \leq \theta(2k)$  if  $k \geq 2$ , and  $\Phi^\vee(k, l) \leq \theta(2k + 2l)$  if  $k + l \geq 3$ .*

As asymptotically the function  $\theta$  is equivalent to the function  $\log$ , we obtain quite better bounds than those from Theorem 28. Furthermore, we get examples proving that, up to constant multiplicative factor, the bounds given by Theorem 29 are optimal.

We have completed our work by finding some properties on the behavior of the functions  $\Phi$  and  $\Phi^\vee$ .

**Theorem 30** (S. Bessy, E. Birmelé and F. Havet, 2006, [14]). *For every  $k \geq 1$ , we have  $\Phi(k) \leq \Phi^\vee(k, 0) \leq \dots \leq \Phi^\vee(k, k) \leq \Phi(k) + 2$  and  $\Phi^\vee(k, 1) \leq \Phi(k) + 1$ .*

Moreover, we conjectured that the first inequality is not optimal, and that  $\Phi^\vee$  is closer to  $\Phi$ .

**Conjecture 31** (S. Bessy, E. Birmelé and F. Havet, 2006, [14]). *For every  $k \geq 1$ , we have  $\Phi^\vee(k, 1) = \Phi(k)$  and  $\Phi^\vee(k, k) \leq \Phi(k) + 1$ .*

Finally, we checked our conjecture for small values of  $k$  by computing exact values of  $\Phi^\vee(k, l)$  and  $\Phi(k)$  for  $k \leq 3$  and  $l \leq 3$ . In particular, as a remarkable result we obtain the following.

**Theorem 32** (S. Bessy, E. Birmelé and F. Havet, 2006, [14]). *We have  $\Phi^\vee(2, 2) = 4$ , that is, every digraph with the property that  $d^+(x) \leq 2$  or  $d^-(x) \leq 2$  for each of its vertex  $x$  admits an arc-coloring with at most four colors.*

This statement was settled by using more sharpened homomorphisms from  $(2 \vee 2)$ -digraph into  $\overline{H_4}$  than the usual mapping to a maximum complete subdigraph of  $\overline{H_4}$  (given by Sperner's Lemma [123]).

### 1.3.2 WDM<sup>5</sup>

This subsection presents another application of graph coloring, in the domain of network optimization and design. I worked in that field while I was in postdoc in 2004 in the *Mascotte Team* at Sophia Antipolis, a research team led by J.C. Bermond and working on algorithms, discrete mathematics and combinatorial optimization with motivations coming from communication networks. During this year, with C. Lepelletier, a Master's degree student, we were interested in a problem arising in the design of optical networks. This

<sup>5</sup>This subsection is linked with the paper: S. Bessy and C. Lepelletier, Optical index of fault tolerant routings in wdm networks, *Networks*, 56(2):95–102, 2010, **annexed p.81**.

topic has been of growing interest over the two last decades, using tools from graph theory and design theory (for instance, see [11], [75] or [10] for a background review of optical networks). The model considered here is valid for the so-called *wavelength division multiplexing* (or *WDM*) optical network. Such a network is modeled by a symmetric directed graph with arcs representing the fiber-optic links. A *request* in the network is an ordered pair of graph nodes, representing a possible communication in the network. A set of different requests is an *instance* in the network. For each request of the instance, we have to select a routing directed path to satisfy it, and the set of all selected paths forms a *routing set* according to the instance. To make the communications possible, a wavelength is allocated to each routing path, such that two paths sharing an arc do not carry the same wavelength; otherwise the corresponding communications could interfere. Given a routing set related to the wavelength assignment, we can define two classical invariants. The *arc-forwarding index* of the routing set is the maximum number of paths sharing the same arc. In the network, there is a general bound on the number of wavelengths which can transit at the same time in a fiber-optic link, corresponding to the admissible maximal arc-forwarding index. The other invariant, called the *optical index* of the routing set, is the minimum number of wavelengths to assign to the routing paths in order to ensure that there is no interference in the network. The main challenge is to provide, for a given instance, a routing set which minimizes the arc-forwarding index or the optical index, or both if possible.

Our work is a contribution to a variant of this problem, introduced by J. Mañuch and L. Stacho [101], in which we focus on possible breakdowns of nodes in the network. Precisely, for a given fixed integer  $f$ , we have to provide, for every request, not just one directed path to satisfy it, but rather a set of  $f + 1$  directed paths with the same starting and ending nodes (corresponding to the request) and which are pairwise internally disjoint. In this routing, if  $f$  nodes break down, every request between the remaining nodes could still be satisfied by a previously selected routing path which contains no failed component. Such a routing set of directed paths is called an  *$f$ -fault tolerant routing* or an  *$f$ -tolerant routing*.

Considering the problematics developed in [101], we focused on the very special cases of complete symmetric directed graphs and complete balanced bipartite symmetric directed graphs. Moreover, we only studied the case of *all-to-all* communication, i.e., where the instance of the problem is the set of all ordered pairs of nodes of the network. So, in a all-to-all context, for a digraph  $D$  and a fixed positive integer  $f$ , an  *$f$ -tolerant routing* in  $D$  is a set of paths  $\mathcal{R} = \{P_i(u, v) : u, v \in V, u \neq v, i = 0, \dots, f\}$  where, for each pair of distinct vertices  $u, v \in V(D)$ , the paths  $P_0(u, v), \dots, P_f(u, v)$  are internally vertex disjoint. Note that such a set of paths exists if and only if the connectivity of the directed graph is large enough (at least  $f + 1$ ), which will be the case in complete and complete bipartite networks for suitable  $f$ .

The basic parameters for WDM optical networks, the arc-forwarding index and the optical index, are generalized in  $f$ -tolerant routings. The *load* of an arc in  $\mathcal{R}$  is the number of directed paths of  $\mathcal{R}$  containing it. By extension, the maximum load over all the arcs of  $D$  is the *load* of the routing, which is also called the *arc-forwarding index of  $\mathcal{R}$*  and is denoted by  $\pi(\mathcal{R})$ . Finally, the *optical index of  $\mathcal{R}$* , denoted  $w(\mathcal{R})$ , is the minimum number of wavelengths to assign to paths of  $\mathcal{R}$  so that no two paths sharing an arc receive the same wavelength. In other words,  $w(\mathcal{R})$  is exactly the chromatic number of the graph with vertex set  $\mathcal{R}$  and where two paths of  $\mathcal{R}$  are linked if they share the same arc of  $D$  (known as the *path graph* of  $\mathcal{R}$ ). The goal, in that context, is to minimize  $\pi(\mathcal{R})$  and  $w(\mathcal{R})$ . So the  *$f$ -tolerant arc-forwarding index* of  $D$  and the  *$f$ -tolerant optical index* of  $D$  are respectively defined by:

$$\pi_f(D) = \min_{\mathcal{R}} \pi(\mathcal{R})$$

$$w_f(D) = \min_{\mathcal{R}} w(\mathcal{R})$$

where the minima span all the possible routing sets  $\mathcal{R}$ . A routing set achieving one of the bounds is said to be *optimal for the arc-forwarding index* or *optimal for the optical index*, respectively.

For a routing set  $\mathcal{R}$ , all paths sharing the same arc must receive different wavelengths in the computation of  $w(\mathcal{R})$ . In particular, we have  $\pi(\mathcal{R}) \leq w(\mathcal{R})$ . By considering a routing set which is optimal for the optical index, we obtain  $\pi_f(D) \leq w_f(D)$ . The equality was conjectured by J. Mañuch and L. Stacho [101].

**Conjecture 33** (J. Mañuch, L. Stacho, 2003, [101]). Let  $D$  be a symmetric directed  $k$ -vertex-connected graph. For any  $f$ ,  $0 \leq f < k$ , we have  $\pi_f(D) = w_f(D)$ .



For  $f = 0$  (without tolerating any faults), the conjecture was previously raised by B. Beauquier et al. [10] and plays a central role in the field of WDM networks.

Recall that we denote by  $K_n^*$  the *complete symmetric directed graph* on  $n$  vertices. In addition, the *complete balanced bipartite symmetric digraph*  $K_{n,n}^*$  is the directed graph on vertex set  $X \cup Y$  with  $X = \{x_1, \dots, x_n\}$  and  $Y = \{y_1, \dots, y_n\}$  and arc set  $\{xy, yx : x \in X, y \in Y\}$ . Thus, we have considered the problem of computing exactly  $w_f(K_n^*)$  and  $w_f(K_{n,n}^*)$ . It is easy to provide a lower bound for the arc-forwarding index of  $K_n^*$ . Indeed, any two vertices  $x$  and  $y$  of  $K_n^*$  have to be linked in an  $f$ -tolerant routing by  $f + 1$  internally disjoint paths. If one of these paths has length one (the arc  $xy$ ), all the others have length at least two, and at least  $2f + 1$  arcs are needed to ensure  $f$ -tolerant communication from  $x$  to  $y$ . So, by an average argument, one arc of  $K_n^*$  must have load at least  $2f + 1$ , providing  $\pi_f(K_n^*) \geq 2f + 1$ . Similarly, we obtain an easy lower bound on  $\pi_f(K_{n,n}^*)$ . In the case of  $K_n^*$ , in 2005, A. Gupta, J. Mañuch and L. Stacho proved in [75] that this lower bound gives exactly the value of the arc-forwarding index. Indeed, they construct  $f$ -tolerant routings through families of independent idempotent Latin squares which are optimal for the arc-forwarding index.

**Theorem 34** (A. Gupta, J. Mañuch, L. Stacho, 2005, [75]). *For every  $f$  with  $0 \leq f \leq n - 2$ , we have  $\pi_f(K_n^*) = 2f + 1$ .*

They also partially bound the optical index of their  $f$ -tolerant routings, proving that  $w_f(K_n^*) \leq 3f + 1$  for some values of  $f$ . This result was improved in 2006 by J.H. Dinitz, A.C.H. Ling and D.R. Stinson [49], who gave a better multiplicative factor for some infinite sets of values of  $n$  and the optimal index up to an additive constant for another infinite set of values of  $n$ . We have improved these results and fixed this computation by showing that every  $f$ -tolerant routing set of  $K_n^*$  which is optimal for the arc-forwarding index is also optimal for the optical index. We thus prove Conjecture 33 for the complete digraphs.

**Theorem 35** (S. Bessy, C. Lepelletier, 2007, [17]). *For every  $f$ ,  $0 \leq f \leq n - 2$ , and every  $f$ -tolerant routing set  $\mathcal{R}$  of  $K_n^*$  with  $\pi(\mathcal{R}) = \pi_f(K_n^*) = 2f + 1$ , we have  $w(\mathcal{R}) = 2f + 1$ . In particular, we have  $w_f(K_n^*) = \pi_f(K_n^*) = 2f + 1$ .*

We obtained this result using an edge coloring model. We define a graph with vertex set the arcs of  $K_n^*$  and where two arcs of  $K_n^*$  are linked by an edge if they belong to a same path of the considered routing. Theorem 35 is then simply obtained by applying Vizing's Theorem [132] to this special graph.

A remaining important issue concerning  $f$ -tolerant routings for  $K_n^*$  is the design of the routings.

**Problem 36.** Is there a simple way (without using idempotent Latin square, for instance) to design in  $K_n^*$  optimal  $f$ -tolerant routings for the arc-forwarding index?

Moreover, we have computed the exact optical index of  $K_{n,n}^*$  and thus proved Conjecture 33 also for this family of graphs. This improves the result of A. Gupta, J. Mañuch and L. Stacho [75], where the upper bound given on the optical index of  $K_{n,n}^*$  is 20% higher than the conjectured optimal value. For that, we described a family of routings and shown that they are all optimal for the arc-forwarding index and the optical index.

**Theorem 37** (S. Bessy, C. Lepelletier, 2007, [17]). *For any  $n \geq 1$  and any  $f$  with  $0 \leq f \leq n - 1$ , we have  $w_f(K_{n,n}^*) = \pi_f(K_{n,n}^*)$ .*

### 1.3.3 Substructures in colored graphs<sup>6</sup>

The problem studied in this section is not exactly a coloring problem, but concerns the existence of some structure in a colored graph. This problematic covers a broad range of problems, and with S. Thomassé, we focused on a conjecture of J. Lehel on the partition of a bi-colored complete graph into two monochromatic cycles. To be precise, we say that a colored graph has a *partition into  $p$  monochromatic cycles (or paths)* if

<sup>6</sup>This subsection is linked with the paper: S. Bessy and S. Thomassé, Partitionning a graph into a cycle and an anticyle, a proof of lehel's conjecture, *Journal of Combinatorial Theory, Serie B*, 100(2):176–180, 2010, **annexed p.105**.

it admits a vertex-partition into  $p$  subgraphs every one of which admits a spanning monochromatic cycle (or path).

Many questions deal with the existence of monochromatic paths and cycles in edge-colored complete graphs. For instance, in 1991, P. Erdős, A. Gyárfás and L. Pyber studied in [60] the minimal number of monochromatic cycles needed to partition the vertex set of the complete graph with edges colored with  $k$  colors. In 2006, A. Gyárfás, M. Ruszinkó, G.N. Sárközy and E. Szemerédi [77] proved that  $O(k \log k)$  such cycles suffice to partition the vertices. One case which received a particular attention was the case  $k = 2$ , where we would like to cover a complete graph whose edges are colored blue and red by two monochromatic cycles. A conjecture of Lehel, first cited in [6], asserts that a blue and a red cycle partition the vertices, where empty set, singletons and edges are allowed as cycles. This statement was proved for sufficiently large  $n$  by T. Łuczak, V. Rödl and E. Szemerédi [128], and more recently by P. Allen [2] with a better bound. Their proofs respectively use the Szemerédi Regularity Lemma and Ramsey's Theory to find useful partition of the vertex set of the colored complete graph. With S. Thomassé, we obtained a general proof of this statement.

**Theorem 38** (S. Bessy, S. Thomassé, 2010, [23]). *Every complete graph with red and blue edges has a vertex partition into a red cycle and a blue cycle.*

Our proof is based on induction, using as starting point the proof of A. Gyárfás of the existence of one red cycle and one blue cycle covering the vertices and intersecting on at most one vertex (see [76]). For this, he considered a longest path consisting of a red path followed by a blue path. The nice fact is that such a path  $P$  is hamiltonian. Indeed, if a vertex  $v$  is not covered, it must be joined in blue to the origin  $a$  of  $P$  and in red to the end  $b$  of  $P$ . But then, one can cover the vertices of  $P$  and  $v$  using the edge  $ab$ . Consequently, there exists a hamiltonian cycle consisting of two monochromatic paths. Hence, there exists a monochromatic cycle  $C$ , of size at least two, and a monochromatic path  $P$  with different colors partitioning the vertex set. The induction in our proof of Theorem 38 runs on the size of  $C$ : at each step, either we can find the two desired cycles, or we increase the length of  $C$ .

There exist many other interesting questions dealing with substructures in colored graphs. I list below some of them, which are either generalizations of Lehel's Conjecture, or famous questions raised in that field. The first natural extension to this problem is to increase the number of colors of the background structure. It was considered by P. Erdős, A. Gyárfás and L. Pyber in their seminal paper, where they raised the following conjecture, still open for  $k \geq 3$ .

**Conjecture 39** (P. Erdős, A. Gyárfás and L. Pyber, 1991, [60]). For every coloring of the edges of the complete graph  $K_n$  with  $k$  colors, there exists a partition of the vertex set of  $K_n$  into  $r$  monochromatic cycles.

Nothing is said on requirement for cycles with different colors. Maybe, some connectivity conditions could be asked for each graph induced by the edges with same color. In particular, the following question is interesting.

**Problem 40.** Let be a coloring of the edges of the complete graph  $K_n$  with 3 colors such that each graph induced by the edges with the same color has vertex-connectivity at least 2. Is it possible to partition the vertex set of  $K_n$  into three monochromatic cycles, one of each color?

Another kind of questions arises when we change the background graph. For instance, partitioning the complete balanced bipartite graph  $K_{n,n}$  with colored edges has been studied by P. Haxell [79]. She proved that for every  $k$ , there exists an integer  $c_k$  such that, for every coloring of the edges of  $K_{n,n}$  with  $k$  colors, there exists a partition of the vertex set of  $K_{n,n}$  into  $c_k$  monochromatic cycles. In the paper [79], there is no precise mention to the case  $k = 2$ . The bound computed on  $c_k$  gives  $c_2 \geq 64$ , but we can conjecture that the real value of  $c_2$  is really less than 64.

**Problem 41.** If  $n$  is large enough, for every coloring of the edges of  $K_{n,n}$  with 2 colors, does there exist a partition of the vertex set of  $K_{n,n}$  into two monochromatic cycles?

Remark that for  $n = 3$ , it is possible to color the edges of  $K_{3,3}$  with two colors such that each color class induced a forest of  $K_{3,3}$ . Then, in this case, there is no hope to have a partition of the vertex set of  $K_{3,3}$

into two monochromatic cycles. That is why we ask for  $n$  to be large enough, to have enough space in the edge set to form cycles.

A third generalization of Lehel's Conjecture could be obtained in considering objects of higher dimensions. On a ground set  $V$  of  $n$  points, we color into two colors, say red and blue, all the  $d$ -subsets of  $V$ . We call a *d-dimensional cycle* the set of all the facets (faces of dimension  $d$ ) of a polytope (a bounded intersection of half planes) of the  $d + 1$ -dimensional space which vertices are seen as elements of  $V$ . The natural extension to Lehel's question is the following.

**Conjecture 42** (S. Bessy, S. Thomassé, 2007). For every coloring of the  $d$ -subsets of a  $n$ -set  $V$ , there exists a partition of  $V$  into 2 parts, each of them being covered by a monochromatic  $d$ -dimensional cycle, one red and one blue.

For  $d = 2$ , we obtain the statement of Lehel's Conjecture, which is then true. However, for  $d = 3$ , the question remains open, and we obtained the following short statement.

**Conjecture 43** (S. Bessy, S. Thomassé, 2007). For every coloring of the triples of a  $n$ -set  $V$ , there exists a partition of  $V$  into 2 planar triangulations, one red and one blue.

Finally, the last kind of problems I am interested in that field concerns digraphs. If we want a direct translation of Lehel's problem on the complete digraph, we need to ensure that the considered colorings do not induce big acyclic part in each color. For instance, if we consider an enumeration of the vertices of the complete digraph and color all forward arcs in blue and all backward arcs in red, there is no hope to find a vertex partition into two monochromatic directed cycles. To raise a possible conjecture, we ask that for each color, the digraph induced by the arcs of this color must be strongly connected.

**Conjecture 44** (S. Bessy, S. Thomassé, 2007). For every coloring of the arcs of the complete digraph  $K_n^*$  into two colors such that each color induces a strongly connected digraph, there exists a vertex partition of  $K_n^*$  into two directed cycles, one of each color.

Remark that there exists an oriented version of the starting point of the proof of Theorem 38: the existence of a hamiltonian cycle consisting of two monochromatic paths in every edge colored complete graph. Indeed, in 1973, H. Raynaud [111] proved that every arc coloring of the complete digraph contains a hamiltonian (oriented) cycle consisting of two monochromatic (oriented) paths. This should be worth to exploit this result in order to tackle Conjecture 44.

To conclude, let us mention a long standing open problem, initially stated by P. Erdős (cited in [116]) and concerning arc-colored tournaments. This problem is not related to monochromatic cycles but it is quite natural in that field and deals with unavoidable structures in colored directed graphs. In a colored digraph  $D$ , a set of vertices  $S$  is a *set of monochromatic sources* if from every vertex  $x$  of  $D$ , there exists a monochromatic path from a vertex of  $S$  to  $x$ . This problem first appeared in print in a paper of B. Sands, N. Sauer, and R. Woodrow [116], where they proved that every tournament with arcs colored with two colors has a vertex which is a monochromatic source.

**Problem 45** (P. Erdős, 1982). For every  $k$ , is there an integer  $f(k)$  such that every tournament with arcs colored with  $k$  colors has set of monochromatic sources with cardinality at most  $f(k)$ .

So, according to this formalism, B. Sands, N. Sauer, and R. Woodrow proved that  $f(2) = 1$ . However, for  $k \geq 3$ , the value  $f(k)$  is not known and even not known to exist. More precisely, for  $k = 3$  the last authors conjectured the following

**Problem 46** (B. Sands, N. Sauer, and R. Woodrow, 1982, [116]). We have  $f(3) = 3$ , that is, every tournament with arcs colored with 3 colors has set of monochromatic sources with cardinality at most 3.

## 1.4 Algorithmic problems on graphs

This last section deals with some algorithmic problems on graphs and exact solutions for these problems. The first subsection is an overview of the work I have done with A. Perez (and other co-authors) while he

prepared his PhD Thesis which I co-supervised with C. Paul. We looked at some parameterized problems and presented kernelization algorithms for these problems. The second subsection presents a joint work with F. Havet on a problem consisting in counting the number of edge-colorings of a regular graph. We gave upper bound for this number and yielded to exponential algorithms, with lowest as possible exponential basis, to enumerate all this colorings.

### 1.4.1 Kernelization for some editing problems

With A. Perez, we worked on some modification problems for graphs and digraphs. Given a class  $\Pi$  of graphs (or digraphs), generally defined by some properties or a set of forbidden induced subgraphs, the generic modification problem is the following.

PARAMETERIZED  $\Pi$ -MODIFICATION PROBLEM

**Input:** A graph (or a directed graph)  $G = (V, E)$ .

**Parameter:** An integer  $k \geq 0$ .

**Question:** Is there a subset  $F \subseteq V \times V$  with  $|F| \leq k$  such that the graph  $G + F = (V, E \Delta F)$  belongs to the class  $\Pi$ ?

Graph modification problems cover a broad range of *NP*-Complete problems and have been extensively studied in the literature [100, 119, 120]. Well-known examples include the VERTEX COVER [46], FEEDBACK-VERTEX-SET [126], or CLUSTER EDITING [41] problems. These problems find applications in various domains, such as computational biology [83, 120], image processing [119] or relational databases [124]. Precisely, for a given graph  $G = (V, E)$ , in a *completion problem*, the set  $F$  of modified edges is constrained to be disjoint from  $E$ , whereas in an *edge deletion problem*  $F$  has to be a subset of  $E$ . If no restriction applies to  $F$ , then we obtain an *edition problem*. Though most of the edge-modification problems turn out to be *NP*-hard problems, in some cases, efficient algorithms can be obtained to solve the natural parameterized version of some of them. The goal is to obtain a classification in the context of parameterized complexity (polynomial kernel, *FPT* without polynomial kernel or not *FPT*, for instance) of the PARAMETERIZED  $\Pi$ -MODIFICATION PROBLEMS according to the class  $\Pi$ . Very few general results are known in this problematic. For instance, a graph modification problem is *FPT* whenever  $\Pi$  can be characterized by a finite set of forbidden induced subgraphs [34]. But, even in this simple case, the existence of a polynomial kernel is not ensured. We will discuss later this question more precisely, but it motivated our work on graph modification problems. Thus, in order to find polynomial kernel for some PARAMETERIZED  $\Pi$ -MODIFICATION PROBLEMS, we focused on very structured class of graphs  $\Pi$  (classes having a tree-like decomposition and tournaments). More precisely, we found three polynomial kernelizations: for the 3-LEAF POWER EDITING PROBLEM (joint work with C. Paul and A. Perez), for the PROPER INTERVAL COMPLETION PROBLEM (joint work with A. Perez) and for the FEEDBACK-ARC-SET IN TOURNAMENT PROBLEM (joint work with F.V. Fomin, S. Gaspers, C. Paul, A. Perez, S. Saurabh and S. Thomassé). We used a very similar approach for the two first problems, which are closed. For the third problem, which is on tournaments, the techniques are quite different but the general scheme of the algorithm design is also similar.

Very classically, given a PARAMETERIZED  $\Pi$ -MODIFICATION PROBLEM, the general process used to find a small kernel is the following. For an instance  $(G, k)$  of this problem, we apply on  $G$  a set of *rules* to obtain a graph  $G'$  equivalent to  $G$ , and we show that if  $G$  is a positive instance of the considered problem, then, the size of  $G'$  is bounded by a polynomial in  $k$ . Some of the rules we used are quite generic. I list them below.

- If a connected component of  $G$  is already a graph belonging to  $\Pi$ , then we can remove it, under the condition that  $\Pi$  is closed under disjoint union. Provided that the class  $\Pi$  has a polynomial algorithm of recognition, this rule can be applied in polynomial time.
- If  $G$  has a big set of vertices which have the same behavior, then we can edit this set in a same way. Precisely, we proved in [19] the following. If  $\Pi$  is closed under true twin addition and induced subgraphs then, from every set  $T$  of true twins in  $G$  with  $|T| > k$ , we can remove  $|T| - (k + 1)$  arbitrary vertices from  $T$ . Moreover, this rule can be applied in polynomial time using a modular decomposition algorithm or more easily, partition refinement (see [78] for example).

- If  $G$  has an edge or a non-edge  $e$  contained in more than  $k$  obstructions of the class  $\Pi$ , which are elsewhere disjoint, then we can edit  $e$  and reduce the parameter consequently. Classically, this rule is called a *sunflower rule*, and is usually applied for the finite obstructions of the class  $\Pi$ . Thus, it can be computed in polynomial time.

After that, we tried to generalize the first of these rules by reducing parts of the graph  $G$  which are already 'clean', but not necessarily form a connected component. We call such a part of  $G$ , a *branch* of  $G$ . The exact definition has to be adapted for each singular case, but broadly speaking, a subgraph  $H$  of  $G$  is a branch if  $H$  belongs to  $\Pi$  and has some special properties of adjacency with the remaining of  $G$ . The '*branches rule*' consists then in localizing big branches in the graph and reducing them ('cutting the branches'). It is possible if we can show that for the considered PARAMETERIZED  $\Pi$ -MODIFICATION PROBLEM, the relevant information contained in a branch lies in its border. This 'concept' of branch is a natural idea, and as been used before for kernelization algorithms (see [83], for instance). Finally, we have to prove that if  $G$  is a positive instance of the problem, then, after applying these rules, the size of the obtained graph  $G'$  will be small, i.e. polynomial in  $k$ . This will be possible, as if  $G$  is a positive instance, up to few edges, it looks like a graph of  $\Pi$ , and then has big parts, branches, behaving like subgraphs of a graph of  $\Pi$ , and will be reduced by the '*branches rule*'.

### A polynomial kernel for the 3-LEAF POWER EDITING PROBLEM<sup>7</sup>

In this subsection, I present a joint work with C. Paul and A. Perez, which consists in finding a polynomial kernel for a PARAMETERIZED  $\Pi$ -MODIFICATION PROBLEM, where  $\Pi$  is the class of *3-leaf powers*, graphs arising from a phylogenetic reconstruction context [86, 87, 106]. Briefly these graphs come from the following problem. We want to extract, from a threshold graph  $G$  on a set  $S$  of species, a tree  $T$ , whose leaf set is  $S$  and such that the distance between two species is at most  $p$  in  $T$  if, and only if, they are adjacent in  $G$  ( $p$  being the value used to extract  $G$  from dissimilarity information). If such a tree  $T$  exists, then  $G$  is a *p-leaf power* and  $T$  is its *p-leaf root*. Here, we are dealing with 3-leaf power which have several nice characterizations (see [32] and [52]). The *critical graph* of a graph  $G$  is obtained by contracting all the set of pairwise true twins of the graphs  $G$ . Then, a graph is a 3-leaf power if its critical graph is a tree. Equivalently, 3-leaf powers are the chordal graphs without induced *bull* (a 3-cycle with two pending vertices), *dart* (build from a path of length 2 and an isolated vertex, both dominated by a fifth vertex) and *gem* (a path of length three with a dominating vertex).

Following theoretical motivations, we looked for a polynomial kernel for the 3-LEAF POWER EDITING PROBLEM, answering to an open question of M. Dom, J. Guo, F. Hüffner and R. Niedermeier [53, 51].

Concerning algorithmic on  $p$ -leaf power, the following is known. For  $p \leq 5$ , the  $p$ -leaf power recognition is polynomial time solvable [33, 36], whereas the question is still open for  $p$  strictly larger than 5. Parameterized  $p$ -leaf power edge modification problems have been studied so far for  $p \leq 4$ . The edition problem for  $p = 2$  is known as the classical CLUSTER EDITING problem for which the kernel size bound has been successively improved in a series of papers [62, 71, 109, 74], culminating in 2010 [41] with a kernel with  $2k$  vertices. For larger values of  $p$ , the edition problem is known as the CLOSEST  $p$ -LEAF POWER problem. For  $p = 3$  and 4, the CLOSEST  $p$ -LEAF POWER problem is known to be *FPT* [52, 51], while its fixed-parameterized tractability is still open for larger values of  $p$ . However, the existence of a polynomial kernel for  $p > 2$  remained an open question [50, 53]. Moreover, though the completion and edge-deletion problems also are *FPT* for  $p \leq 4$  [51, 53], no polynomial kernel was known for  $p \neq 2$  [74]. In this context, we focused on the case  $p = 3$  and obtained the following.

**Theorem 47** (S. Bessy, C. Paul, A. Perez, 2010, [19]). *The CLOSEST 3-LEAF POWER, the 3-LEAF POWER COMPLETION and the 3-LEAF POWER EDGE-DELETION admit a kernel with  $O(k^3)$  vertices.*

To obtain this kernel, we followed the general scheme explained previously. In this context, for an instance graph  $G$ , a branch of  $G$  is a subgraph which forms a sub-tree of the critical graph of  $G$  containing at most

<sup>7</sup>This subsection is linked with the paper: S. Bessy, C. Paul, and A. Perez, Polynomial kernels for 3-leaf power graph modification problems, *Discrete Applied Mathematics*, 158(16):1732–1744, 2010, **annexed p.110**.

two vertices with neighbors outside of this sub-tree in the critical graph of  $G$ . Then, the analysis of the rules yielded to the desired kernel.

#### A polynomial kernel for the PROPER INTERVAL COMPLETION PROBLEM<sup>8</sup>

In this second subsection, I present a joint work with A. Perez, related to a PARAMETERIZED  $\Pi$ -MODIFICATION PROBLEM, where the allowed modifications are only edge completions and the class  $\Pi$  is the class of *proper interval graphs*, which are the intersection graphs of finite sets of unit length intervals on a line. Thus, we studied the PROPER INTERVAL COMPLETION PROBLEM and found a kernelization algorithm for this problem. The class of proper interval graphs is a well-studied class of graphs, and several characterizations are known to exist. In particular, there exists a set of forbidden induced subgraphs that characterizes proper interval graphs [133]: all the  $k$ -cycles with  $k \geq 4$ , the *claw*, which is the complete bipartite graph  $K_{1,3}$ , the *net*, a 3-cycle with three pending vertices, and the *3-sun*, which is the complementary of the net. The proper interval graphs are also characterized by having an *umbrella ordering* [97]. An umbrella ordering of a graph  $G$  is an ordering  $v_1, \dots, v_n$  of its vertices such that for every edge  $v_i v_j$  of  $G$  with  $i < j$ , the set  $\{v_i, \dots, v_j\}$  is a clique of  $G$  (it corresponds to the order of the first extremity of each interval in an interval representation of  $G$ ).

Interval completion problems find applications in molecular biology and genomic research [80, 83], and in particular in *physical mapping* of DNA. This motivation was cited in the first papers dealing with PROPER INTERVAL COMPLETION PROBLEM (see [83] for instance). This problem is known to be NP-Complete for a long time [70], but fixed-parameter tractable due to a result of H. Kaplan, R. Shamir and R.E. Tarjan in FOCS '94 [83, 84]. Nevertheless, it was not known whether this problem admits a polynomial kernel or not. We settled this question by proving the following.

**Theorem 48** (S. Bessy, A. Perez, 2011, [20]). *The PROPER INTERVAL COMPLETION problem admits a kernel with at most  $O(k^3)$  vertices.*

Remark that this problem is quite similar to the 3-LEAF POWER COMPLETION: this is an edge completion problem to a class of chordal graph defined by a finite set of obstructions. The proof also follows the previous general scheme, but this time the proof for the 'branches rules' is really more technical. A branch for this problem is a subgraph  $H$  of our instance graph  $G$  which induces a proper interval graph and such that the edges standing across the partition  $(H, G \setminus H)$  form at most two *generalized join*. Such a join is a set of edges which contains no induced  $2K_2$  (two disjoint edges). It corresponds to the edges across a partition  $(\{v_1, \dots, v_k\}, \{v_{k+1}, \dots, v_n\})$  of an umbrella ordering of a proper interval graph. Detecting and reducing the branches produced the cubic kernel.

Moreover, we applied our techniques to the so-called BIPARTITE CHAIN DELETION problem, closely related to the PROPER INTERVAL COMPLETION problem where one is given a graph  $G = (V, E)$  and seeks a set of at most  $k$  edges whose deletion from  $E$  result in a bipartite chain graph (a graph that can be partitioned into two independent sets connected by a generalized join). For this problem, we obtained a quadratic kernel.

**Theorem 49** (S. Bessy, A. Perez, 2011, [20]). *The problem BIPARTITE CHAIN DELETION admits a kernel with at most  $O(k^2)$  vertices.*

This result completes a previous result of Guo [74] who proved that the BIPARTITE CHAIN DELETION WITH FIXED BIPARTITION problem admits a kernel with  $O(k^2)$  vertices.

To conclude these two parts, we return on a more general framework for the PARAMETERIZED  $\Pi$ -MODIFICATION PROBLEMS. As previously mentioned, it is known that a graph modification problem is FPT whenever  $\Pi$  can be characterized by a finite set of forbidden induced subgraphs [34]. However, recent results proved that several graph modification problems do not admit a polynomial kernel even for such classes  $\Pi$  [73, 94]. For instance, an impressive result is that if  $\Pi$  is the class of the graphs without induced

<sup>8</sup>This subsection is linked with the paper: S. Bessy and A. Perez. Polynomial kernels for proper interval completion and related problems. In *FCT*, volume 6914 of *LNCIS*, pages 1732–1744, 2011, **annexed p.134**.

$2K_2$ , then the **PARAMETERIZED  $\Pi$ -COMPLETION PROBLEM** has no polynomial kernel (personal communication of F. Havet, C. Paul, A. Perez and S. Guillemot on a work in progress). So, in this field, the following question is a central one.

**Problem 50.** Is it possible to characterize the class  $\Pi$  of graphs such that the **PARAMETERIZED  $\Pi$ -MODIFICATION PROBLEM** is *FPT* or admits a polynomial kernel?

More precisely, with A. Perez, focusing on completion problems, we tried to generalize the notion of branches and apply it to the **PARAMETERIZED  $\Pi$ -COMPLETION PROBLEM**. In the two examples presented, the fact that the classes of graphs (3-leaf power and proper interval) are chordal seems very useful to obtain polynomial kernels. So, we asked the following question.

**Conjecture 51** (S. Bessy, A. Perez, 2011 [20]). If  $\Pi$  is a class of chordal graphs defined by a finite set of obstructions, then the **PARAMETERIZED  $\Pi$ -COMPLETION PROBLEM** admits a polynomial kernel.

As mentioned in [20], it is easy to see that such problems are *FPT*. Moreover, another clue for this conjecture is that when  $\Pi$  is simply the class of chordal graphs, H. Kaplan, R. Shamir and R.E. Tarjan have shown in 1994 [83] that the **PARAMETERIZED  $\Pi$ -COMPLETION PROBLEM** (also called the **MINIMUM FILL-IN PROBLEM**) admits a cubic kernel.

### A polynomial kernel for the **FEEDBACK-ARC-SET IN TOURNAMENT PROBLEM**<sup>9</sup>

The last problem I looked at in the context of parameterized complexity deals with feedback-arc-set in tournaments. It is (very) collective work done in 2009 and join with F.V. Fomin, S. Gaspers, C. Paul, A. Perez, S. Saurabh and S. Thomassé and it is not very far from the problems presented in the two first subsection. Given a directed graph  $G = (V, A)$  on  $n$  vertices and an integer parameter  $k$ , the **FEEDBACK-ARC-SET** problem asks whether the given digraph has a set of  $k$  arcs whose removal results in an acyclic directed graph. It is a **PARAMETERIZED  $\Pi$ -ARC-DELETION PROBLEM**, where  $\Pi$  stands for the class of acyclic digraphs. We considered this problem in the class of tournaments. More precisely, the problem is the following.

**FEEDBACK-ARC-SET IN TOURNAMENTS (FAST):**

**Input:** A tournament  $T = (V, A)$  and a positive integer  $k$ .

**Parameter:**  $k$ .

**Question:** Is there a subset  $F \subseteq A$  of at most  $k$  arcs whose removal makes  $T$  acyclic?

Feedback-arc-sets in tournaments are well studied from the combinatorial [61, 82, 113, 118, 122, 135], statistical [121] and algorithmic [1, 4, 44, 90, 129, 130] points of view. The problem **FAST** has some nice applications, as for instance, in *rank aggregation*, where we are given several rankings of a set of objects, and we wish to produce a single ranking that on average is as consistent as possible with the given ones, according to some chosen measure of consistency. This problem has been studied in the context of voting [31, 39, 43], machine learning [42], and search engine ranking [56, 57]. A natural consistency measure for rank aggregation is the number of pairs that occur in a different order in the two rankings. This leads to *Kemeny rank aggregation* [88, 89], a special case of a weighted version of **FAST**.

However, we were mainly motivated by theoretical aspects of the problem, on which the following is known. The **FAST** problem is *NP*-complete by recent results of N. Alon [4] and P. Charbit et al. [38]. From an approximation perspective, **FAST** admits an approximation algorithm, found in 2007 by C. Kenyon-Mathieu and W. Schudy [90] and which is used in our kernelization process. The problem is also well studied in parameterized complexity. V. Raman and S. Saurabh [110] showed that **FAST** is *FPT* and a kernel on  $O(k^2)$  vertices is known for this problem, a result from N. Alon et al. [5] and M. Dom et al. [54]. We improved these last results by providing a *linear vertex* kernel for **FAST**.

**Theorem 52** (S. Bessy, F.V. Fomin, S. Gaspers, C. Paul, A. Perez, S. Saurabh and S. Thomassé., 2009, [15]). *Given any fixed  $\epsilon > 0$ , **FAST** admits a kernel with at most  $(2 + \epsilon)k$  vertices.*

For that, given an instance  $(T, k)$  of **FAST**, we start by computing a feedback-arc-set with at most

<sup>9</sup>This subsection is linked with the paper: S. Bessy, F. V. Fomin, S. Gaspers, C. Paul, A. Perez, S. Saurabh, and S. Thomassé, *Kernels for feedback arc set in tournaments*, In *FSTTCS*, volume 4 of *LIPIcs*, pages 37–47, 2009, **annexed p.161**.

$(1 + \frac{\epsilon}{2})k$  arcs by using the approximation algorithm of C. Kenyon-Mathieu and W. Schudy [90] (if we do not success, then we answer 'NO' for the instance  $(T, k)$ ). Then, if  $T$  is large enough, with more than  $(2 + \epsilon)k$  vertices, we find a partition of  $T$  with few arcs going across the partition. Using the following useful lemma, we can remove these arcs and decrease  $k$  accordingly.

**Lemma 53** (S. Bessy, F.V. Fomin, S. Gaspers, C. Paul, A. Perez, S. Saurabh and S. Thomassé., 2009, [15]). *Let  $E = v_1, \dots, v_n$  be an enumeration of a tournament  $T$  with  $p$  backward arcs (i.e. arcs  $v_i v_j$  with  $i > j$ ). If every interval  $v_i, \dots, v_j$  of  $E$  with  $i < j$  contains at most  $\frac{j-i}{2}$  backward arcs of  $T$ , then,  $T$  contains exactly  $p$  arc-disjoint 3-cycles.*

After that, we remove from  $T$  the vertices not contained in any cycle and repeat the process until we obtain the kernel with the desired size (or answer 'NO' if we have to reduce  $k$  below 0).

Remark that the complexity needed to compute the kernel depends on the complexity of the feedback-arc-set approximation of [90], and thus is in time  $O(n^{O(\epsilon^{-12})})$ . However, C. Paul, A. Perez and S. Thomassé gave in 2011 [107] a simpler kernelization process for FAST, providing a  $4k$  kernel in quadratic time.

To conclude, there is an interesting open question related to the FAST problem. Indeed, if we are interested now in computing a feedback-vertex-set of size less than  $k$  in a tournament, there exists a kernel on  $O(k^2)$  vertices for the parametrized version of this problem, see [126] for instance. But the existence of a linear kernel for this problem is still open.

**Problem 54.** Is there a kernel on  $O(k)$  vertices for the PARAMETRIZED FEEDBACK-VERTEX-SET IN TOURNAMENT problem?

### 1.4.2 Counting edge-colorings of regular graphs<sup>10</sup>

In this last subsection, I present a work done with F. Havet on the number of edge-colorings of regular graphs. This work was initially motivated by results and questions from P.A. Golovach, D. Kratsch and J.F. Couturier. Indeed, in [69] there were interested in enumerating all the edge-colorings of a regular graph and provided exponential exact algorithms for this problem. However, they asked if the exponential bases for their algorithms could be improved, specially in the case of edge-coloring of cubic graphs. By using some structural tools to enumerate the edge-colorings, we settled the question and improved their results.

Algorithmic for graph coloring is a very large field of research and a lot of results are known in this area. Very classically, every kind of usual chromatic number (related to vertex-coloring, edge-coloring...) is NP-complete to compute (see [68, 81, 115] for instance). So, many exact algorithms with exponential time concerning these problems have been published in the last decade. One of the major results is the  $O^*(2^n)$  inclusion-exclusion algorithm to compute the chromatic number of a graph found independently by A. Björklund, T. Husfeldt [24] and M. Koivisto [92]. This approach may also be used to establish a  $O^*(2^n)$  algorithm to count the  $k$ -colorings and to compute the chromatic polynomial of a graph. It also implies a  $O^*(2^m)$  algorithm to count the  $k$ -edge-colorings. Since edge-coloring is a particular case of vertex-coloring, a natural question is to ask if faster algorithms than the general one may be designed in these cases. For instance, very recently A. Björklund et al. [25] showed how to detect whether a  $k$ -regular graph admits a  $k$ -edge-coloring in time  $O^*(2^{(k-1)n/2})$ .

The existential problem, asking whether a graph has a coloring with a fixed and small number  $k$  of colors, has also attracted a lot of attention. For vertex-colorability the fastest algorithm for  $k = 3$  has running time  $O^*(1.3289^n)$  and was proposed by R. Beigel and D. Eppstein [12], and the fastest algorithm for  $k = 4$  has running time  $O^*(1.7272^n)$  and was given by F. Fomin et al. [64]. They also established algorithms for counting  $k$ -colorings for  $k = 3$  and 4. The existence problem for a 3-edge-coloring is considered in [12, 93, 69]. L. Kowalik [93] gave an algorithm deciding if a graph is 3-edge-colorable in time  $O^*(1.344^n)$  and polynomial space and P.A. Golovach et al. [69] presented an algorithm counting the number of 3-edge-colorings of a graph in time  $O^*(3^{n/6}) = O^*(1.201^n)$  and exponential space. They also showed a branching algorithm to

<sup>10</sup>This subsection is linked with the paper: S. Bessy and F. Havet. Enumerating the edge-colourings and total colourings of a regular graph. *accepted in Journal of Combinatorial Optimization*, 2011, **annexed p.172**.



enumerate all the 3-edge-colorings of a connected cubic graph of running time  $O^*(25^{n/8}) = O^*(1.5423^n)$  using polynomial space. In particular, this implies that every connected cubic graph of order  $n$  has at most  $O(1.5423^n)$  3-edge-colorings. Moreover, they gave an example of a connected cubic graph of order  $n$  having  $\Omega(1.2820^n)$  3-edge-colorings.

We filled the gap between these two bounds and improved their results by proving the following.

**Theorem 55** (S. Bessy, F. Havet, 2011, [16]). *In every connected cubic multi-graph of order  $n$ , the number of 3-edge-colorings is at most  $3 \cdot 2^{n/2}$ . Furthermore, they can be all enumerated in time  $O^*(2^{n/2}) = O^*(1.4143^n)$  using polynomial space by a branching algorithm.*

Moreover, we gave an example of connected cubic multi-graph achieving this bound. To compute efficiently the number of edge-colorings of a cubic graph  $G$ , we used an special enumeration of  $G$ , defined by A. Lempel et al. in 1967 and called a *st*-ordering of  $G$  [95]. In such an ordering, every vertex of  $G$  (excepted from the first and the last) has degree at least one on its left and also degree at least one on its right. More precisely, if we orient every edge of  $G$  from the left to the right in an *st*-ordering, we can see that the first vertex has out-degree 3, the last vertex as out-degree 0 and  $\frac{n}{2} - 1$  vertices of  $G$  have out-degree 1 and  $\frac{n}{2} - 1$  vertices of  $G$  have out-degree 2. So, given an *st*-ordering  $E$  of  $G$ , we sort the edges of  $G$  according to their left extremity in  $E$ . Then, we greedily enumerate all the edge-colorings of  $G$  by coloring as many ways as possible the edges of  $G$  according to this sorting. As, the only choices for colorings appear when a vertex has two neighbors in its right in  $E$ , we obtained the announced bound.

For simple graphs, we tried to sharpen the previous bound and showed the following.

**Theorem 56** (S. Bessy, F. Havet, 2011, [16]). *In every connected cubic simple graph of order  $n$ , the number of 3-edge-colorings is at most  $\frac{9}{4} \cdot 2^{n/2}$ .*

However, we did not find an example of graph having this number of 3-edge-colorings, and we believe that  $\frac{9}{4}$  is not the optimal value in the previous statement. Precisely, guided by some examples with high number of 3-edge-colorings, we conjectured the following.

**Conjecture 57** (S. Bessy, F. Havet, 2011, [16]). Up to an additive constant, in every connected cubic simple graph of order  $n$  the number of 3-edge-colorings is at most  $2^{n/2}$ .

To conclude, let me mention that we extended our results to  $k$ -regular connected multi-graph and obtained the following statement.

**Theorem 58** (S. Bessy, F. Havet, 2011, [16]). *In every connected  $k$ -regular multi-graph of order  $n$ , the number of  $k$ -edge-colorings is at most  $k \cdot ((k-1)!)^{n/2}$ . Furthermore, they can be all enumerated in time  $O^*(((k-1)!)^{n/2})$  using polynomial space by a branching algorithm.*

# Bibliography

- [1] N. Ailon, M. Charikar, and A. Newman. Aggregating inconsistent information: ranking and clustering. In *ACM Symposium on Theory of Computing (STOC)*, pages 684–693, 2005.
- [2] P. Allen. Covering two-edge-coloured complete graphs with two disjoint monochromatic cycles. *Combinatorics, Probability and Computing*, 17:471–486, 2008.
- [3] N. Alon. Disjoint directed cycles. *Journal of Combinatorial Theory, Series B*, 68(2):167–178, 1996.
- [4] N. Alon. Ranking tournaments. *SIAM Journal of Discrete Mathematics*, 20(1):137–142, 2006.
- [5] N. Alon, D. Lokshtanov, and S. Saurabh. Fast FAST. In *ICALP*, volume 5555 of *LNCS*, pages 49–58. Springer, 2009.
- [6] J. Ayel. *Sur l’existence de deux cycles supplémentaires unicolores, disjoints et de couleurs différentes dans un graphe complet bicolore*. PhD thesis, Université de Grenoble, 1979.
- [7] J. Bang-Jensen, S. Bessy, and S. Thomassé. Disjoint 3-cycles in tournaments: a proof of the bermond-thomassen conjecture for tournaments. *submitted*, 2011.
- [8] J. Bang-Jensen and G. Gutin. *Digraphs. Theory, algorithms and applications*. Springer Monographs in Mathematics. Springer-Verlag London, 2nd edition edition, 2009.
- [9] J. Bang-Jensen and C. Thomassen. A polynomial algorithm for the 2-path problem for semicomplete digraphs. *SIAM Journal of Discrete Mathematics*, 5:366–376, 1992.
- [10] B. Beauquier, J.-C. Bermond, L. Gargano, P. Hell, S. Pérennes, and U. Vaccaro. Graph problems arising from wavelength-routing in all-optical networks. In *the Second IEEE Workshop on Optics and Computer Science*, 1997.
- [11] B. Beauquier, P. Hell, and S. Pérennes. Optimal wavelength-routed multicasting. *Discrete Applied Mathematics*, 84:15–20, 1998.
- [12] R. Beigel and D. Eppstein. 3-coloring in time  $o(1.3289^n)$ . *Journal of Algorithms*, 54:168–204, 2005.
- [13] J.-C. Bermond and C. Thomassen. Cycles in digraphs - a survey. *Journal of Graph Theory*, 5(1):1–43, 1981.
- [14] S. Bessy, E. Birmelé, and F. Havet. Arc-chromatic number of digraphs in which every vertex has bounded outdegree or bounded indegree. *Journal of Graph Theory*, 53(4):315–332, 2006.
- [15] S. Bessy, F. V. Fomin, S. Gaspers, C. Paul, A. Perez, S. Saurabh, and S. Thomassé. Kernels for feedback arc set in tournaments. In *FSTTCS*, volume 4 of *LIPIcs*, pages 37–47, 2009.
- [16] S. Bessy and F. Havet. Enumerating the edge-colourings and total colourings of a regular graph. *accepted in Journal of Combinatorial Optimization*, 2011.
- [17] S. Bessy and C. Lepelletier. Optical index of fault tolerant routings in wdm networks. *Networks*, 56(2):95–102, 2010.

- [18] S. Bessy, N. Lichiardopol, and J.S. Sereni. Two proofs of the bermond-thomassen conjecture for tournaments with bounded minimum in-degree. *Discrete Mathematics*, 310:557–560, 2010.
- [19] S. Bessy, C. Paul, and A. Perez. Polynomial kernels for 3-leaf power graph modification problems. *Discrete Applied Mathematics*, 158(16):1732–1744, 2010.
- [20] S. Bessy and A. Perez. Polynomial kernels for proper interval completion and related problems. In *FCT*, volume 6914 of *LNCS*, pages 1732–1744, 2011.
- [21] S. Bessy and S. Thomassé. Every strong digraph has a spanning strong subgraph with at most  $n+2\alpha-2$  arcs. *Journal of Combinatorial Theory, Series B*, 87:289–299, 2003.
- [22] S. Bessy and S. Thomassé. Spanning a strong digraph with  $\alpha$  cycles: a conjecture of gallai. *Combinatorica*, 27(6):659–667, 2007.
- [23] S. Bessy and S. Thomassé. Partitionning a graph into a cycle and an anticycle, a proof of lehel’s conjecture. *Journal of Combinatorial Theory, Serie B*, 100(2):176–180, 2010.
- [24] A. Björklund and T. Husfeldt. Inclusion-exclusion algorithms for counting set partitions. In *In Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, pages 575–582. IEEE, 2006.
- [25] A. Björklund, T. Husfeldt, P. Kaski, and M. Koivisto. Narrow sieves for parameterized paths and packings. *arXiv:1007.1161v1*, 2010.
- [26] H.L. Bodlaender. Kernelization: New upper and lower bound techniques. In *International Workshop on Parameterized and Exact Computation (IWPEC)*, Lecture Notes in Computer Science, pages 17–37, 2009.
- [27] H.L. Bodlaender, R.G. Downey, M.R. Fellows, and D. Hermelin. On problems without polynomial kernels. *Journal of Computer and System Sciences*, 75(8):423–434, 2009.
- [28] J.A. Bondy. Disconnected orientations and a conjecture of las vergnas. *Journal of the London Mathematical Society*, 14(2):277 – 282, 1976.
- [29] J.A. Bondy. A short proof of the chen-manalastas theorem. *Discrete Mathematics*, 146(1-3):289 – 292, 1995.
- [30] J.A. Bondy and U.S.R. Murty. *Graph Theory*, volume 244 of *Graduate Texts in Mathematics*. Springer-Verlag London, 3rd corrected printing edition, 2008.
- [31] J. Borda. Mémoire sur les élections au scrutin. *Histoire de l’Académie Royale des Sciences*, 1781.
- [32] A. Brandstädt and V.B. Le. Structure and linear time recognition of 3-leaf powers. *Information Processing Letters*, 98(4):133–138, 2006.
- [33] A. Brandstädt, V.B. Le, and R. Sritharan. Structure and linear time recognition of 4-leaf powers. *ACM Transactions on Algorithms (TALG)*, 5(1):1–22, 2008.
- [34] L. Cai. Fixed-parameter tractability of graph modification problems for hereditary properties. *Information Processing Letters*, 58(4):171–176, 1996.
- [35] P. Camion. Chemins et circuits hamiltoniens des graphes complets. *Comptes Rendus de l’Académie des Sciences*, 249:2151 – 2152, 1959.
- [36] M.-S. Chang and M.-T. Ko. The 3-steiner root problem. In *International Workshop on Graph Theoretical Concepts in Computer Science (WG)*, Lecture Notes in Computer Science, pages 109–120, 2007.
- [37] P. Charbit and A. Sebö. Cyclic orders: equivalence and duality. *Combinatorica*, 28(2):131–143, 2008.

- [38] P. Charbit, S. Thomassé, and A. Yeo. The minimum feedback arc set problem is NP-hard for tournaments. *Combinatorics, Probability and Computing*, 16(1):1–4, 2007.
- [39] I. Charon and O. Hudry. A survey on the linear ordering problem for weighted or unweighted tournaments. *4OR*, 5(1):5–60, 2007.
- [40] C.C. Chen and P. Manalastas. Every finite strongly connected digraph of stability 2 has a hamiltonian path. *Discrete Mathematics*, 44:243 – 250, 1983.
- [41] J. Chen and J. Meng. A  $2k$  kernel for the cluster editing problem. In *In Lin, G. (ed.) COCOON 2007*, volume 6196 of *Lecture Notes in Computer Science*, pages 459–468, 2010.
- [42] W.W. Cohen, R.E. Schapire, and Y. Singer. Learning to order things. In *Advances in neural information processing systems (NIPS)*, pages 451–457, 1997.
- [43] M. Condorcet. Essai sur l’application de l’analyse à la probabilité des décisions rendues à la pluralité des voix, 1785.
- [44] D. Coppersmith, L. Fleischer, and A. Rudra. Ordering by weighted number of wins gives a good ranking for weighted tournaments. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 776–782, 2006.
- [45] T.H. Cormen, Charles E. Leiserson, R.L. Rivest, and C. Stein. *Introduction to Algorithms, Third Edition*. The MIT Press and McGraw-Hill Book Company, 2009.
- [46] F.K.H.A. Dehne, M.R. Fellows, F.A. Rosamond, and P. Shaw. Greedy localization, iterative compression, modeled crown reductions: New FPT techniques, an improved algorithm for set splitting, and a novel  $2k$  kernelization for vertex cover. In *International Workshop on Parameterized and Exact Computation (IWPEC)*, volume 3162 of *Lecture Notes in Computer Science*, pages 271–280, 2004.
- [47] R. Diestel. *Graph Theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag Heidelberg, 4th edition edition, 2010.
- [48] R.P. Dilworth. A decomposition theorem for partially ordered sets. *Annals of Mathematics*, 51:161 – 166, 1950.
- [49] J.H. Dinitz, A.C.H. Ling, and D.R. Stinson. Fault tolerant routings with minimum optical index. *Networks*, 48:47–54, 2006.
- [50] M. Dom, J. Guo, F. Hüffner, and R. Niedermeier. Error compensation in leaf root problems. In *International Symposium on Algorithms and Computation (ISAAC)*, number 3341 in *Lecture Notes in Computer Science*, pages 389–401, 2004.
- [51] M. Dom, J. Guo, F. Hüffner, and R. Niedermeier. Extending the tractability border for closest leaf powers. In *International Workshop on Graph Theoretical Concepts in Computer Science (WG)*, number 3787 in *Lecture Notes in Computer Science*, pages 397–408, 2005.
- [52] M. Dom, J. Guo, F. Hüffner, and R. Niedermeier. Error compensation in leaf root problems. *Algorithmica*, 44:363–381, 2006.
- [53] M. Dom, J. Guo, F. Hüffner, and R. Niedermeier. Closest 4-leaf power is fixed-parameter tractable. *Discrete Applied Mathematics*, 156(18):3345–3361, 2008.
- [54] M. Dom, J. Guo, F. Hüffner, R. Niedermeier, and A. Truß. Fixed-parameter tractability results for feedback set problems in tournaments. In *Conference on Algorithms and Complexity (CIAC)*, volume 3998 of *LNCS*, pages 320–331, 2006.
- [55] R.G. Downey and M.R. Fellows. *Parameterized complexity*. Springer, 1999.
- [56] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation revisited. Technical report.

- [57] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *World Wide Web Conference (WWW)*, 2001.
- [58] A. El-Sahili. Fonctions de graphes disjoints. *Comptes Rendus de l'Académie des Sciences*, 319(Série I):519–521, 1994.
- [59] A. El-Sahili. Functions and line digraphs. *Journal of Graph Theory*, 44(4):296–303, 2003.
- [60] P. Erdős, A. Gyárfás, and L. Pyber. Vertex coverings by monochromatic cycles and trees. *Journal of Combinatorial Theory, Series B*, 51:90–95, 1991.
- [61] P. Erdős and J. W. Moon. On sets on consistent arcs in tournaments. *Canad. Math. Bull.*, 8:269–271, 1965.
- [62] M.R. Fellows, M. Langston, F. Rosamond, and P. Shaw. Efficient parameterized preprocessing for cluster editing. In *International Symposium on Fundamentals of Computation Theory (FCT)*, number 4639 in *Lecture Notes in Computer Science*, pages 312–321, 2007.
- [63] J. Flum and M. Grohe. *Parameterized complexity theory*. Texts in Theoretical Computer Science. Springer, 2006.
- [64] F.V. Fomin, S. Gaspers, and S. Saurabh. Improved exact algorithms for counting 3- and 4-colorings. In *In Lin, G. (ed.) COCOON 2007*, volume 4598 of *Lecture Notes in Computer Science*, pages 65–74, 2007.
- [65] T. Gallai. *Problem 15*. MIT Press, Cambridge, MA, USA, 1964.
- [66] T. Gallai. On directed paths and circuits. In *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, pages 115–118, New York, 1968. Academic Press.
- [67] T. Gallai and A.N. Milgram. Verallgemeinerung eines graphentheoretischen satzes von rédei. *Acta Sci. Math. Szeged*, 21:181 – 186, 1960.
- [68] M.R. Garey and D.S. Johnson. *Computers and intractability A guide to the theory of NP-completeness*. A Series of Books in the Mathematical Sciences. W.H. Freeman and Co., 1979.
- [69] P.A. Golovach, D. Kratsch, and J.F. Couturier. Colorings with few colors: counting, enumeration and combinatorial bounds. In *International Workshop on Graph Theoretical Concepts in Computer Science (WG)*, volume 6410 of *Lecture Notes in Computer Science*, pages 39–50, 2010.
- [70] M.C. Golumbic, H. Kaplan, and R. Shamir. On the complexity of DNA physical mapping. *ADVAM: Advances in Applied Mathematics*, 15, 1994.
- [71] J. Gramm, J. Guo, F. Hüffner, and R. Niedermeier. Graph-modeled data clustering: Exact algorithms for clique generation. *Theory of Computing Systems*, 38(4):373–392, 2005.
- [72] B. Guiduli. On incidence coloring and star arboricity of graphs. *Discrete Mathematics*, 163:275–278, 1997.
- [73] S. Guillemot, C. Paul, and A. Perez. On the (non-)existence of polynomial kernels for  $p_l$ -free edge modification problems. In *IPEC*, volume 6478 of *LNCS*, pages 147–157, 2010.
- [74] J. Guo. Problem kernels for NP-complete edge deletion problems: Split and related graphs. In *ISAAC*, volume 4835 of *LNCS*, pages 915–926, 2007.
- [75] A. Gupta, J. Mañuch, and L. Stacho. Fault tolerant forwarding and optical indexes: a design theory approach. *Journal of Combinatorial Designs*, 14:25–40, 2006.
- [76] A. Gyárfás. Vertex coverings by monochromatic paths and cycles. *Journal of Graph Theory*, 7:131–135, 1983.

- [77] A. Gyárfás, M. Ruszinkó, G.N. Sárközy, and E. Szemerédi. An improved bound for the monochromatic cycle partition number. *Journal of Combinatorial Theory, Series B*, 96:855–873, 2006.
- [78] M. Habib and C. Paul. A survey on algorithmic aspects of modular decomposition. *Computer Science Review*, 4(1):41–59, 2010.
- [79] P.E. Haxell. Partitioning complete bipartite graphs by monochromatic cycles. *Journal of Combinatorial Theory, Series B*, 69:210–218, 1997.
- [80] P. Hell, R. Shamir, and R. Sharan. A fully dynamic algorithm for recognizing and representing proper interval graphs. *SIAM Journal of Computing*, 31(1):289–305, 2001.
- [81] I. Holyer. The np-completeness of edge-colouring. *SIAM Journal of Computing*, 2:225–231, 1981.
- [82] H.A. Jung. On subgraphs without cycles in tournaments. *Combinatorial Theory and its Applications II*, pages 675–677, 1970.
- [83] H. Kaplan, R. Shamir, and R.E. Tarjan. Tractability of parameterized completion problems on chordal and interval graphs: Minimum fill-in and physical mapping. In *In Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1994)*, pages 780–791. IEEE, 1994.
- [84] H. Kaplan, R. Shamir, and R.E. Tarjan. Tractability of parameterized completion problems on chordal, strongly chordal, and proper interval graphs. *SIAM Journal of Computing*, 28(5):1906–1922, 1999.
- [85] R.M. Karp. Reducibility among combinatorial problems. In *Complexity of computer computations (Proc. Symp., IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y., 1972)*, pages 85–103. Plenum, 1972.
- [86] P. Kearney and D. Corneil. Tree powers. *Journal of Algorithms*, 29(1):111–131, 1998.
- [87] P.E. Kearney, G.H. Lin, and T. Jiang. Phylogenetic k-root and steiner k-root. In *International Symposium on Algorithms and Computation (ISAAC)*, number 1969 in *Lecture Notes in Computer Science*, pages 539–551, 2000.
- [88] J. Kemeny. Mathematics without numbers. *Daedalus*, 88:571–591, 1959.
- [89] J. Kemeny and J. Snell. *Mathematical models in the social sciences*. Blaisdell, 1962.
- [90] C. Kenyon-Mathieu and W. Schudy. How to rank with few errors. In *ACM Symposium on Theory of Computing (STOC)*, pages 95–103, 2007.
- [91] D.E. Knuth. Wheels within wheels. *Journal of Combinatorial Theory, Series B*, 16(1):42 – 46, 1974.
- [92] M. Koivisto. An  $o(2^n)$  algorithm for graph coloring and other partitioning problems via inclusion-exclusion. In *In Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, pages 583–590. IEEE, 2006.
- [93] L. Kowalik. Improved edge-coloring with three colors. *Theoretical Computer Science*, 410:3733–3742, 2009.
- [94] S. Kratsch and M. Wahlström. Two edge modification problems without polynomial kernels. In *International Workshop on Parameterized and Exact Computation (IWPEC)*, volume 5917 of *Lecture Notes in Computer Science*, pages 264–275. Springer, 2009.
- [95] A. Lempel, S. Even, and I. Cederbaum. An algorithm for planarity testing of graphs. In *In P. Rosenstiehl, editor, Proceedings of the International Symposium on the Theory of Graphs (Rome, July 1966)*, pages 215–232. Gordon and Breach, 1967.
- [96] N. Lichiardopol, A. Pór, and J.-S. Sereni. A step towards the Bermond-Thomassen conjecture about disjoint cycles in digraphs. *SIAM Journal of Discrete Mathematics*, 23:979–992, 2009.

- [97] P.J. Looges and S. Olariu. Optimal greedy algorithms for indifference graphs. *Computers & Mathematics with Applications*, 25(7):15 – 25, 1993.
- [98] C.L. Lucchesi. *A minimax equality for directed graphs*. PhD thesis, University of Waterloo, Ontario, Canada, 1976.
- [99] C.L. Lucchesi and D.H. Younger. A minimax theorem for directed graphs. *Journal of the London Mathematical Society*, 17(3):369–374, 1978.
- [100] F. Mancini. *Graph modification problems related to graph classes*. PhD thesis, University of Bergen, Norway, 2008.
- [101] J. Maňuch and L. Stacho. On  $f$ -wise arc forwarding index and wavelength allocations in faulty all-optical hypercubes. *Theoretical Informatics and Applications*, 37:255–270, 2003.
- [102] W. McCuaig. Intercyclic digraphs. In *Graph Structure Theory*, pages 203–246, 1991.
- [103] K. Menger. Zur allgemeinen kurventheorie. *Fundamenta Mathematicae*, 10:96–115, 1927.
- [104] V. Neumann-Lara. The dichromatic number of a digraph. *Journal Combinatorial Theory, Series B*, 33(3):265–270, 1982.
- [105] R. Niedermeier. *Invitation to fixed parameter algorithms*, volume 31 of *Oxford Lectures Series in Mathematics and its Applications*. Oxford University Press, 2006.
- [106] N. Nishimura, P. Ragde, and D. Thilikos. On graph powers for leaf-labeled trees. *Journal of Algorithms*, 42(1):69–108, 2002.
- [107] C. Paul, A. Perez, and S. Thomassé. Conflict packing yields linear vertex-kernels for  $k$ -fast,  $k$ -dense rti and a related problem. *Lecture Notes in Computer Science, MFCS2011*, 6907:497–507, 2011.
- [108] S. Poljak and V. Rödl. On the arc-chromatic number of a digraph. *Journal of Combinatorial Theory, Series B*, 31(2):190–198, 1981.
- [109] F. Protti, M. Dantas da Silva, and J.L. Szwarcfiter. Applying modular decomposition to parameterized cluster editing problems. *Theory of Computing Systems*, 2007.
- [110] V. Raman and S. Saurabh. Parameterized algorithms for feedback set problems and their duals in tournaments. *Theoretical Computer Science*, 351(3):446–458, 2006.
- [111] H. Raynaud. Sur le circuit hamiltonien bi-coloré dans les graphes orientés. *Periodica Mathematica Hungarica*, 3(3–4):289–297, 1973.
- [112] B.A. Reed, N. Robertson, P.D. Seymour, and R. Thomas. Packing directed circuits. *Combinatorica*, 16(4):535–554, 1996.
- [113] K.D. Reid and E.T. Parker. Disproof of a conjecture of Erdős and Moser on tournaments. *Journal of Combinatorial Theory*, 9:225–238, 1970.
- [114] B. Roy. Nombre chromatique et plus longs chemins d’un graphe. *Rev. Française Informat. Recherche Opérationnelle*, 1:129–132, 1967.
- [115] A. Sánchez-Arroyo. Determining the total colouring number is np-hard. *Discrete Mathematics*, 78(3):315–319, 1989.
- [116] B. Sands, N. Sauer, and R. Woodrow. On monochromatic paths in edge-coloured digraphs. *Journal of Combinatorial Theory, Series B*, 33(3):271–275, 1982.
- [117] A. Sebö. Minmax relations for cyclically ordered digraphs. *Journal of Combinatorial Theory, Series B*, 97(4):518–552, 2007.

- [118] S. Seshu and M.B. Reed. *Linear Graphs and Electrical Networks*. Addison-Wesley, 1961.
- [119] R. Shamir, R. Sharan, and D. Tsur. Cluster graph modification problems. *Discrete Applied Mathematics*, 144(1-2):173–182, 2004.
- [120] R. Sharan. *Graph modification problems and their applications to genomic research*. PhD thesis, Tel-Aviv University, 2002.
- [121] P. Slater. Inconsistencies in a schedule of paired comparisons. *Biometrika*, 48:303–312, 1961.
- [122] J. Spencer. Optimal ranking of tournaments. *Networks*, 1:135–138, 1971.
- [123] E. Sperner. Ein satz über untermengen einer endlichen menge. *Mathematische Zeitschrift*, 27, 1928.
- [124] R.E. Tarjan and M. Yannakakis. Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM Journal of Computing*, 13(3):566–579, 1984.
- [125] S. Thomassé. Spanning a strong digraph by  $\alpha$  cycles, the case  $\alpha = 3$  of gallai’s conjecture. *preprint*, 2000.
- [126] S. Thomassé. A  $4k^2$  kernel for feedback vertex set. *ACM Transactions on Algorithms*, 6(2), 2010.
- [127] C. Thomassen. Disjoint cycles in digraphs. *Combinatorica*, 3(3-4):393–396, 1983.
- [128] T. Łuczak, V. Rödl, and E. Szemerédi. Partitioning two-colored complete graphs into two monochromatic cycles. *Combinatorics, Probability and Computing*, 7:423–436, 1998.
- [129] A. van Zuylen. Deterministic approximation algorithms for ranking and clusterings. Technical Report 1431, Cornell ORIE, 2005.
- [130] A. van Zuylen, R. Hegde, K. Jain, and D.P. Williamson. Deterministic pivoting algorithms for constrained ranking and clustering problems. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 405–414, 2007.
- [131] A. Vetta. Approximating the minimum strongly connected subgraph via a matching lower bound. In *SODA*, pages 417–426, 2001.
- [132] V.G. Vizing. On an estimate of the chromatic class of a  $p$ -graph. *Diskret. Analiz*, 3:25–30, 1964.
- [133] G. Wegner. *Eigenschaften der nerven homologische-einfactor familien in  $R^n$* . PhD thesis, Universität Göttingen, Göttingen, Germany, 1967.
- [134] D.R. Woodall. Menger and König systems. In *Theory and applications of graphs (Proc. Int. Conf., Western Mich. Univ., Kalamazoo, 1976)*, volume 642 of *Lect. Notes Math.*, pages 620–635. Springer, 1978.
- [135] D.H. Younger. Minimum feedback arc sets for a directed graph. *IEEE Trans. Circuit Theory*, 10:238–245, 1963.
- [136] D.H. Younger. Graphs with interlinked directed circuits. In *Proc. Midwest symposium on circuits theory 2*, pages XVI 2.1– XVI 2.7, 1973.
- [137] X. Zhu. Circular chromatic number: a survey. *Discrete Mathematics*, 229:371 – 410, 2001.





## Chapter 2

# Materials

### Part I, problems on cycle in digraphs:

1. S. Bessy, S. Thomassé, Spanning a strong digraph with alpha cycles: a conjecture of Gallai. *Combinatorica*, 27 (6), 2007, and *Lecture Notes in Computer Science*, Vol. 3064, Springer-Verlag, Actes de IPCO X 2004.
2. S. Bessy, N. Lichiardopol and J.-S. Sereni, Two proofs of Bermond-Thomassen conjecture for regular tournaments. *Discrete Mathematics*, 310 (3), 2010, 557–560, and 6th Czech-Slovak International Symposium on Combinatorics (2006).
3. J. Bang-Jensen, S. Bessy and S. Thomassé, Disjoint 3-cycles in tournaments: a proof of the Bermond-Thomassen conjecture for tournaments, submitted.

### Part II, colouring and partitionning problems :

1. S. B., C. Lepelletier, Optical index of fault tolerant routings in WDM networks. *Networks*, 56 (2), 2010, 95–102.
2. S. Bessy, S. Thomassé, Partitionning a graph into a cycle and an anticycle, a proof of Lehel’s conjecture. *Journal of Combinatorial Theory Serie B*, 100 (2), 2010, 176–180.
3. S. Bessy, E. Birmelé, F. Havet, Arc-chromatic number of digraphs in which every vertex has bounded outdegree or bounded indegree. *Journal of Graph Theory*, 53 (4), 2006, 315–332.

### Part III, algorithmic problems on graphs :

1. S. Bessy, F.V. Fomin, S. Gaspers, C. Paul, A. Perez, S. Saurabh, S. Thomassé, Kernels for Feedback Arc Set In Tournaments. *FSTTCS 2009*: 37-47, et accepté à *Journal of Computer and System Sciences*.
2. S. Bessy, C. Paul, A. Perez, Polynomial kernels for 3-leaf power graph modification problems. *IWOCA 2009*: 72–82, et *Discrete Applied Mathematics*, 158 (2010) pp. 1732-1744 .
3. S. Bessy and A. Perez, Polynomial kernels for Proper Interval Completion and related problems, *FCT*, volume 6914 of *LNCS*, pages 1732-1744, 2011.
4. S. Bessy and F. Havet, Enumerating the edge-colourings and total colourings of a regular graph, accepted to *Journal of Combinatorial Optimization*, 2011.

## 2.1 Spanning a strong digraph with alpha cycles

### Spanning a strong digraph by $\alpha$ circuits: A proof of Gallai's conjecture.

Stéphane Bessy  
and  
Stéphan Thomassé

Laboratoire LaPCS, U.F.R. de Mathématiques,  
Université Claude Bernard - Lyon 1, 50, avenue Tony Garnier,  
Bâtiment RECHERCHE [B], Domaine de Gerland  
69367 Lyon Cedex 07, France  
email : bessy@univ-lyon1.fr  
thomasse@univ-lyon1.fr

#### Abstract

In 1963, Tibor Gallai [9] asked whether every strongly connected directed graph  $D$  is spanned by  $\alpha$  directed circuits, where  $\alpha$  is the stability of  $D$ . We give a proof of this conjecture.

## 1 Coherent cyclic orders.

In this paper, circuits of length two are allowed. Since loops and multiple arcs play no role in this topic, we will simply assume that our digraphs are loopless and simple. A directed graph (digraph) is *strongly connected*, or simply *strong*, if for all vertices  $x, y$ , there exists a directed path from  $x$  to  $y$ . A *stable set* of a directed graph  $D$  is a subset of vertices which are not pairwise joined by arcs. The *stability* of  $D$ , denoted by  $\alpha(D)$ , is the number of vertices of a maximum stable set of  $D$ . It is well-known, by the Gallai-Milgram theorem [10] (see also [1] p. 234 and [3] p. 44), that  $D$  admits a vertex-partition into  $\alpha(D)$  disjoint paths. We shall use in our proof a particular case of this result, known as Dilworth's theorem [8]: a partial order  $P$  admits a vertex-partition into  $\alpha(P)$  chains (linear orders). Here  $\alpha(P)$  is the size of a maximal antichain. In [9], Gallai raised the problem, when  $D$  is strongly connected, of spanning  $D$  by a union of circuits. Precisely, he made the following conjecture (also formulated in [1] p. 330, [2] and [3] p. 45):

**Conjecture 1** *Every strong digraph with stability  $\alpha$  is spanned by the union of  $\alpha$  circuits.*

The case  $\alpha = 1$  is Camion's theorem [6]: Every strong tournament has a hamilton circuit. The case  $\alpha = 2$  is a corollary of a result of Chen and Manalastas [7] (see also Bondy [4]): Every strong digraph with stability two is spanned by two circuits intersecting each other on a (possibly empty) path. In [11] was proved the case  $\alpha = 3$ . In the next section of this paper, we will give a proof of Gallai's conjecture for every  $\alpha$ .

Let  $D$  be a strong digraph on vertex set  $V$ . An enumeration  $E = v_1, \dots, v_n$  of  $V$  is *elementary equivalent* to  $E' = v_n, v_1, \dots, v_{n-1}$ , or  $E' = v_2, v_1, v_3, \dots, v_n$  if neither

$v_1v_2$  nor  $v_2v_1$  is an arc of  $D$ . Two enumerations  $E, E'$  of  $V$  are *equivalent* if there is a sequence  $E = E_1, \dots, E_k = E'$  such that  $E_i$  and  $E_{i+1}$  are elementary equivalent, for  $i = 1, \dots, k-1$ . The classes of this equivalence relation are called the *cyclic orders* of  $D$ . Roughly speaking, a cyclic order is a class of enumerations of the vertices on the integers modulo  $n$ , where one stay in the class while switching consecutive vertices which are not joined by an arc. We fix an enumeration  $E = v_1, \dots, v_n$  of  $V$ , the following definitions are understood with respect to  $E$ . An arc  $v_iv_j$  of  $D$  is a *forward arc* if  $i < j$ , otherwise it is a *backward arc*. A directed path of  $D$  is a *forward path* if it only contains forward arcs. The *index* of a directed circuit  $C$  of  $D$  is the number of backward arcs of  $C$ , we denote it by  $i_E(C)$ . This correspond to the winding number of the circuit. Observe that  $i_E(C) = i_{E'}(C)$  if  $E'$  is elementary equivalent to  $E$ . Consequently, the index of a circuit is invariant in a given cyclic order  $\mathcal{C}$ , we denote it by  $i_{\mathcal{C}}(C)$ . By extension, the index  $i_{\mathcal{C}}(S)$  of a set of circuits  $S$  is the sum of the indices of the circuits of  $S$ . A circuit is *simple* if it has index one. A cyclic order  $\mathcal{C}$  is *coherent* if every arc of  $D$  is contained in a simple circuit, or, equivalently, if for every enumeration  $E$  of  $\mathcal{C}$  and every backward arc  $v_jv_i$  of  $E$ , there exists a forward path from  $v_i$  to  $v_j$ . We denote by  $\text{cir}(D)$  the set of all directed circuits of  $D$ .

**Lemma 1** *Every strong digraph has a coherent cyclic order.*

**Proof.** Let us consider a cyclic order  $\mathcal{C}$  which is minimum with respect to  $i_{\mathcal{C}}(\text{cir}(D))$ . We suppose for contradiction that  $\mathcal{C}$  is not coherent. There exists an enumeration  $E = v_1, \dots, v_n$  and a backward arc  $a = v_jv_i$  which is not in a simple circuit. Assume moreover that  $E$  and  $a$  are chosen in order to minimize  $j-i$ . Let  $k$  be the largest integer  $i \leq k < j$  such that there exists a forward path from  $v_i$  to  $v_k$ . Observe that  $v_k$  has no out-neighbour in  $]v_k, v_j]$ . If  $k \neq i$ , by the minimality of  $j-i$ ,  $v_k$  has no in-neighbour in  $]v_k, v_j]$ . In particular the enumeration  $E' = v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_j, v_k, v_{j+1}, \dots, v_n$  is equivalent to  $E$ , and contradicts the minimality of  $j-i$ . Thus  $k = i$ , and by the minimality of  $j-i$ , there is no in-neighbour of  $v_i$  in  $]v_i, v_j[$ . In particular the enumeration  $E' = v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_{j-1}, v_i, v_j, \dots, v_n$  is equivalent to  $E$ . Observe now that in  $E'' = v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_{j-1}, v_j, v_i, \dots, v_n$ , every circuit  $C$  satisfies  $i_{E''}(C) \leq i_{E'}(C)$ , and the inequality is strict if the arc  $a$  belongs to  $C$ , a contradiction.  $\square$

A direct corollary of Lemma 1 is that every strong tournament has a hamilton circuit, just consider for this any coherent cyclic order.

## 2 Cyclic stability versus spanning circuits.

The *cyclic stability* of a coherent cyclic order  $\mathcal{C}$  is the maximum  $k$  for which there exists an enumeration  $v_1, \dots, v_n$  of  $\mathcal{C}$  such that  $\{v_1, \dots, v_k\}$  is a stable set of  $D$ . We denote it by  $\alpha(\mathcal{C})$ , observe that we clearly have  $\alpha(\mathcal{C}) \leq \alpha(D)$ .

**Lemma 2** *Let  $D$  be a strong digraph and  $v_1, \dots, v_n$  be an enumeration of a coherent cyclic order  $\mathcal{C}$  of  $D$ . Let  $X$  be a subset of vertices of  $D$  such that there is no forward path between two distinct vertices of  $X$ . Then  $|X| \leq \alpha(\mathcal{C})$ .*

**Proof.** We consider an enumeration  $E = v_1, \dots, v_n$  of  $\mathcal{C}$  such that there is no forward path between two distinct vertices of  $X$ , and chosen in such a way that  $j-i$  is minimum, where  $v_i$  is the first element of  $X$  in the enumeration, and  $v_j$  is the last element of  $X$  in the enumeration. Suppose for contradiction that  $X \neq \{v_i, \dots, v_j\}$ . There exists  $v_k \notin X$  for some  $i < k < j$ . There cannot exist both a forward path from  $X \cap \{v_i, \dots, v_{k-1}\}$  to  $v_k$  and a forward path from  $v_k$  to  $X \cap \{v_{k+1}, \dots, v_j\}$ . Without loss of generality, we assume that there is no forward path from  $X \cap \{v_i, \dots, v_{k-1}\}$  to  $v_k$ . Suppose moreover that  $v_k$  is chosen with minimum index  $k$ . Clearly,  $v_k$  has no in-neighbour in  $\{v_i, \dots, v_{k-1}\}$ , and since  $\mathcal{C}$  is coherent,  $v_k$  has no out-neighbour in  $\{v_i, \dots, v_{k-1}\}$ . Thus the enumeration  $v_1, \dots, v_{i-1}, v_k, v_i, \dots, v_{k-1}, v_{k+1}, \dots, v_n$  belongs to  $\mathcal{C}$ , contradicting the minimality of  $j-i$ . Consequently,  $X = \{v_i, \dots, v_j\}$ , and there is no

forward arcs, and then no backward arcs, between the vertices of  $X$ . Considering now the enumeration  $v_i, \dots, v_n, v_1, \dots, v_{i-1}$ , we conclude that  $|X| \leq \alpha(\mathcal{C})$ .  $\square$

Let  $P = x_1, \dots, x_k$  be a directed path, we call  $x_1$  the *head* of  $P$  and  $x_k$  the *tail* of  $P$ . We denote the restriction of  $P$  to  $\{x_i, \dots, x_j\}$  by  $P[x_i, x_j]$ .

**Theorem 1** *Let  $D$  be a strong digraph with a coherent cyclic order  $\mathcal{C}$ . The minimal  $i_{\mathcal{C}}(S)$ , where  $S$  is a spanning set of circuits of  $D$  is equal to  $\alpha(\mathcal{C})$ .*

**Proof.** We consider a coherent cyclic order  $\mathcal{C}$  of  $D$  with cyclic stability  $k := \alpha(\mathcal{C})$ . Let  $E = v_1, \dots, v_n$  be an enumeration of  $\mathcal{C}$  such that  $S = \{v_1, \dots, v_k\}$  is a stable set of  $D$ . Clearly, if a circuit  $C$  contains  $q$  vertices of  $S$ , the index of  $C$  is at least  $q$ . In particular the inequality  $i_{\mathcal{C}}(S) \geq k$  is satisfied for every spanning set of circuits of  $D$ . To prove that equality holds, we consider an auxiliary acyclic digraph  $D'$  on vertex set  $V \cup \{v'_1, \dots, v'_k\}$  which arc set consists of every forward arc of  $E$  and every arc  $v_i v'_j$  for which  $v_i v_j$  is an arc of  $D$ . We call  $T'$  the transitive closure of  $D'$ . Let us prove that the size of a maximal antichain in the partial order  $T'$  is exactly  $k$ . Consider such an antichain  $A$ , and set  $A_1 := A \cap \{v_1, \dots, v_k\}$ ,  $A_2 := A \cap \{v_{k+1}, \dots, v_n\}$  and  $A_3 := A \cap \{v'_1, \dots, v'_k\}$ . Since one can arbitrarily permute the vertices of  $S$  in the enumeration  $E$  and still remain in  $\mathcal{C}$ , we may assume that  $A_3 = \{v'_1, \dots, v'_j\}$  for some  $0 \leq j \leq k$ . Since every vertex is in a simple circuit, there is a directed path in  $D'$  from  $v_i$  to  $v'_i$ , and consequently we cannot both have  $v_i \in A$  and  $v'_i \in A$ . Clearly, the enumeration  $E' = v_{j+1}, \dots, v_n, v_1, \dots, v_j$  belongs to  $\mathcal{C}$ . By the fact that  $A$  is an antichain of  $T'$ , there is no forward path joining two elements of  $(A \cap V) \cup \{v'_1, \dots, v'_j\}$  in  $E'$ , and thus, by Lemma 2,  $|A| = |(A \cap V) \cup \{v'_1, \dots, v'_j\}| \leq k$ . Observe also that  $\{v_1, \dots, v_k\}$  are the sources of  $T'$  and  $\{v'_1, \dots, v'_k\}$  are the sinks of  $T'$ , and both are maximal antichains of  $T'$ . We apply Dilworth's theorem in order to partition  $T'$  into  $k$  chains (thus starting in the set  $\{v_1, \dots, v_k\}$  and ending in the set  $\{v'_1, \dots, v'_k\}$ ), and by this, there exists a spanning set  $P_1, \dots, P_k$  of directed paths of  $D'$  with heads in  $\{v_1, \dots, v_k\}$  and tails in  $\{v'_1, \dots, v'_k\}$ . We can assume without loss of generality that the head of  $P_i$  is exactly  $v_i$ , for all  $i = 1, \dots, k$ . Let us now denote by  $\sigma$  the permutation of  $\{1, \dots, k\}$  such that  $v'_{\sigma(i)}$  is the tail of  $P_i$ , for all  $i$ . Assume that among all spanning sets of paths, we have chosen  $P_1, \dots, P_k$  (with respective heads  $v_1, \dots, v_k$ ) in such a way that the permutation  $\sigma$  has a maximum number of cycles. We claim that if  $(i_1, \dots, i_p)$  is a cycle of  $\sigma$  (meaning that  $\sigma(i_j) = i_{j+1}$  and  $\sigma(i_p) = i_1$ ), then the paths  $P_{i_1}, \dots, P_{i_p}$  are pairwise vertex-disjoint. If not, suppose that  $v$  is a common vertex of  $P_{i_1}$  and  $P_{i_m}$ , and replace  $P_{i_1}$  by  $P_{i_1}[v_{i_1}, v] \cup P_{i_m}[v, v_{\sigma(i_m)}]$  and  $P_{i_m}$  by  $P_{i_m}[v_{i_m}, v] \cup P_{i_1}[v, v_{\sigma(i_1)}]$ . This is a contradiction to the maximality of the number of cycles of  $\sigma$ . Now, in the set of paths  $P_1, \dots, P_k$ , contract all the pairs  $\{v_i, v'_i\}$ , for  $i = 1, \dots, k$ . This gives a spanning set  $S$  of circuits of  $D$  which satisfies  $i_{\mathcal{C}}(S) = k$ .  $\square$

**Corollary 1.1** *Every strong digraph  $D$  is spanned by  $\alpha(D)$  circuits.*

**Proof.** By Lemma 1,  $D$  has a coherent cyclic order  $\mathcal{C}$ . By Theorem 1,  $D$  is spanned by a set  $\mathcal{S}$  of circuits such that  $|\mathcal{S}| \leq i_{\mathcal{C}}(S) = \alpha(\mathcal{C}) \leq \alpha(D)$ .  $\square$

We now establish the arc-cover analogue of Theorem 1. Again, a minimax result holds.

### 3 Cyclic feedback arc set versus arc cover.

Let  $\mathcal{C}$  be a cyclic order of a strong digraph  $D$ . We denote by  $\beta(\mathcal{C})$  the maximum  $k$  for which there exists an enumeration of  $\mathcal{C}$  with  $k$  backward arcs. We call  $k$  the *maximal feedback arc set* of  $\mathcal{C}$ . Since every vertex of  $D$  has indegree at least one, we clearly have  $\alpha(\mathcal{C}) \leq \beta(\mathcal{C})$ .

**Theorem 2** *Let  $D = (V, A)$  be a strong digraph with a coherent cyclic order  $\mathcal{C}$ . The minimal  $i_{\mathcal{C}}(S)$ , where  $S$  is a set of circuits which covers the arc set of  $D$ , is equal to  $\beta(\mathcal{C})$ .*

**Proof.** If  $\mathcal{S}$  is a set of circuits which spans the arcs of  $D$ , every backward arc in any enumeration of  $\mathcal{C}$  must be in a circuit of  $\mathcal{S}$ . In particular the inequality  $i_{\mathcal{C}}(\mathcal{S}) \geq \beta(\mathcal{C})$  clearly holds. Let  $D'$  be the subdivision of  $D$ , i.e. the digraph with vertex set  $V \cup A$  and arc set  $\{(v, e) : v \text{ is the head of } e \text{ and } e \in A\} \cup \{(e, v) : v \text{ is the tail of } e \text{ and } e \in A\}$ . There is a one-to-one correspondence  $\phi$  between the circuits of  $D'$  and the circuits of  $D$ . Let  $E = v_1, \dots, v_n$  be an enumeration of  $\mathcal{C}$  with backward arc set  $\{e_1, \dots, e_k\}$ , where  $k = \beta(\mathcal{C})$ . Consider the enumeration  $E'$  of  $D'$  given by

$$E' := e_1, \dots, e_k, v_1, f_1^1, \dots, f_{n_1}^1, v_2, f_1^2, \dots, f_{n_2}^2, \dots, v_n$$

where  $\{f_1^i, \dots, f_{n_i}^i\}$  is the set of forward arcs in  $E$  with head  $v_i$ . Let  $\mathcal{C}'$  be the cyclic order of  $E'$ . The index in  $\mathcal{C}'$  of a circuit  $C$  of  $D'$  is equal to the index in  $\mathcal{C}$  of  $\phi(C)$ , thus  $\mathcal{C}'$  is coherent. Let  $F'$  be any enumeration of  $\mathcal{C}'$ . We denote by  $F$  the enumeration induced by  $F'$  on  $V$ . Since  $\mathcal{C}'$  is coherent, if  $e := xy$  is a forward arc of  $F$ ,  $xy$  is a forward path of  $F'$ . In particular,  $F$  belongs to  $\mathcal{C}$  (since one cannot switch  $x$  and  $y$  in  $F'$ ). Moreover, if  $e := xy$  is a backward arc of  $F$ , exactly one of  $xe$  or  $ey$  is a backward arc of  $F'$ . Thus, we have  $\beta(\mathcal{C}') \leq \beta(\mathcal{C})$ . By Theorem 1, the vertex set of  $D'$  is spanned by a set of circuits  $\mathcal{S}'$  with  $i_{\mathcal{C}'}(\mathcal{S}') \leq \alpha(\mathcal{C}') \leq \beta(\mathcal{C}') \leq \beta(\mathcal{C})$ . To conclude, observe that  $\mathcal{S} := \{\phi(C), C \in \mathcal{S}'\}$  is a set of circuits which covers the arc set of  $D$  and verify  $i_{\mathcal{C}}(\mathcal{S}) = i_{\mathcal{C}'}(\mathcal{S}') \leq \beta(\mathcal{C})$ .  $\square$

#### 4 Longest circuit versus minimum cyclic coloration.

In this section, we present a third min-max theorem which consists of a fractional version of a theorem of J.A. Bondy ([5]). Our proof is similar to the classical proof on the circular chromatic number in the non-oriented case, see X. Zhu ([12]) for a survey.

The *cyclic chromatic number* of a coherent cyclic order  $\mathcal{C}$  of a strong digraph  $D$ , denoted by  $\chi(\mathcal{C})$ , is the minimum  $k$  for which there exists an enumeration  $E = v_1, \dots, v_{i_1}, v_{i_1+1}, \dots, v_{i_2}, v_{i_2+1}, \dots, v_{i_k}$  of  $\mathcal{C}$  for which  $v_{i_j+1}, \dots, v_{i_{j+1}}$  is a stable set for all  $j = 0, \dots, k-1$  (with  $i_0 := 0$ ).

Under the same hypothesis, the *circular chromatic number* of  $\mathcal{C}$ , denoted by  $\chi_c(\mathcal{C})$  is the infimum of the numbers  $r \geq 1$  for which  $\mathcal{C}$  admits an  $r$ -circular coloration. A mapping  $f : V \rightarrow [0, r[$  is called an  *$r$ -circular coloration* if  $f$  verifies:

- 1) If  $x$  and  $y$  are linked in  $D$ , then  $1 \leq |f(x) - f(y)| \leq r - 1$ .
- 2) If  $0 \leq f(v_1) \leq f(v_2) \leq \dots \leq f(v_n) < r$ , then  $v_1, \dots, v_n$  must be an enumeration of  $\mathcal{C}$ . Such an enumeration is called *related* to  $f$ .

As usual, it is convenient to represent such an application as a mapping from  $V$  into a circle of the euclidean plane with circumference  $r$ . Condition 1) asserts then that two linked vertices have distance at least 1 on this circle. And by condition 2), the vertices of  $D$  are placed on the circle according to an enumeration of the cyclic order  $\mathcal{C}$ . By compactness of this representation, the infimum used in the definition of  $\chi_c$  is a minimum, that is to say that there exists a  $\chi_c(\mathcal{C})$ -circular coloration of  $\mathcal{C}$ .

Note that the enumeration given by 2) is possibly not unique. Indeed, two vertices of  $V$  may have the same image by  $f$ . In this case, these two vertices are not linked in  $D$  (because of 1)) and so, the two enumerations are equivalent. Moreover, two enumerations related to  $f$  have same sets of forward arcs and backward arcs.

**Lemma 3** *For  $D$  a strong digraph and  $\mathcal{C}$  a coherent cyclic order of  $D$ , we have  $\lceil \chi_c(\mathcal{C}) \rceil = \chi(\mathcal{C})$ .*

**Proof.** First, if  $E = v_1, \dots, v_{i_1}, v_{i_1+1}, \dots, v_{i_2}, v_{i_2+1}, \dots, v_{i_k}$  is an enumeration of  $\mathcal{C}$  which realizes  $\chi(\mathcal{C})$ , we can easily check that  $f : V \rightarrow [0, k[$  defined by  $f(v_p) = j$  if  $i_j + 1 \leq p \leq i_{j+1}$  for  $0 \leq j \leq k-1$  with  $i_0 = 0$  is a  $k$ -circular coloration of  $\mathcal{C}$ . So, we get the inequality  $\chi_c(\mathcal{C}) \leq \chi(\mathcal{C})$ . Conversely, the existence of an enumeration of  $\mathcal{C}$  which realizes  $\chi_c(\mathcal{C})$  will achieve the bound. Indeed, fix

$f$  an  $r$ -circular coloration of  $\mathcal{C}$  with  $r = \chi_c(\mathcal{C})$  and  $E$  an enumeration of  $\mathcal{C}$  related to  $f$ . By definition of  $f$ , for every integer  $0 \leq j \leq \lceil r \rceil - 1$ , the set  $\{v \in V : f(v) \in [j, j+1]\}$  is a stable set of  $D$  and forms an interval on  $E$ , so we get  $\chi(\mathcal{C}) \leq \lceil \chi_c(\mathcal{C}) \rceil$ .  $\square$

The following Lemma gives a criterion to decide whether an  $r$ -circular coloration  $f$  is best possible or not. We define an auxiliary digraph  $D_f$  with vertex set  $V$  and arc set  $\{xy \in E(D) : f(y) - f(x) = 1 \text{ or } f(x) - f(y) = r - 1\}$ . Observe that the arcs  $xy$  of  $D_f$  with  $f(y) - f(x) = 1$  (resp.  $f(x) - f(y) = r - 1$ ) are forward (resp. backward) in any enumeration related to  $f$ .

**Lemma 4** *If  $f$  is an  $r$ -circular coloration of  $\mathcal{C}$  with  $r > 2$  for which  $D_f$  is an acyclic digraph, then we can provide a real number  $r' < r$  such that  $\mathcal{C}$  admits an  $r'$ -circular coloration.*

**Proof.** First of all, if a vertex  $x$  of  $D$  has an out-neighbour  $y$  with  $f(x) - f(y) = 1$ , by property 1) of  $f$  and coherence of  $\mathcal{C}$ , the arc  $yx$  must be also in  $D$ , and similarly if  $x$  has an out-neighbour  $y$  with  $f(y) - f(x) = r - 1$ , the arc  $yx$  must be in  $D$ . So, a vertex  $x$  with in-degree 0 (resp. out-degree 0) in  $D_f$  has no neighbour  $z$  with  $f(x) - f(z) = 1$  modulo  $r$  (resp.  $f(z) - f(x) = 1$  modulo  $r$ ). Then, if  $E(D_f) = \emptyset$ , it is easy to provide an  $r'$ -circular coloration  $f'$  of  $\mathcal{C}$  with  $r' < r$ . Just multiply  $f$  by a factor  $1 - \epsilon$  with  $\epsilon > 0$  and  $\epsilon$  small enough.

Now, amongst the  $r$ -circular colorations  $f$  of  $\mathcal{C}$  for which  $D_f$  is acyclic, choose one with minimal number of arcs for  $D_f$ . Assume that  $E(D_f) \neq \emptyset$ . We can choose a vertex  $x$  of  $D_f$  with in-degree 0 and out-degree at least 1. Denote by  $y_1, \dots, y_p$  the out-neighbours of  $x$  in  $D_f$ , we have for all  $i$ ,  $f(y_i) = f(x) + 1$  modulo  $r$ . By definition of  $f$ ,  $x$  has no neighbour  $z$  such that  $f(x) - f(z) < 1$  or  $f(z) - f(x) > r - 1$  and, moreover, since  $x$  has in-degree 0 in  $D_f$ , by the previous remark,  $x$  has no neighbour  $z$  with  $f(x) - f(z) = 1$  or  $f(z) - f(x) = r - 1$ . Observe that none of the  $y_i$  verifies this, because  $r > 2$ . So, we can provide an  $r$ -circular coloration  $f'$  derived from  $f$  just by changing the value of  $f(x)$ : choose  $f'(x) = f(x) - \epsilon$  modulo  $r$  with  $\epsilon > 0$  and such that no neighbour of  $x$  has an image by  $f$  in  $[f'(x) - 1, f'(x)] \cup [r - 1 + f'(x), r]$ . We check that  $E(D_{f'}) = E(D_f) \setminus \{xy_i : i = 1, \dots, p\}$ , which contradicts the choice of  $f$ . So,  $E(D_f) = \emptyset$  and we provide an  $r'$ -circular coloration of  $\mathcal{C}$  with  $r' < r$  as previously.  $\square$

Finally, we can state a third min-max theorem about cyclic orders. For this, we define, for a fixed cyclic order  $\mathcal{C}$ , the *cyclic length* of a circuit  $C$  of  $D$ , denoted by  $l_{\mathcal{C}}(C)$ , as the number of vertices of  $C$  in  $D$ , divided by the index of  $C$  in  $\mathcal{C}$ .

**Theorem 3** *Let  $\mathcal{C}$  be a coherent cyclic order of a strong digraph  $D$ . The maximal  $l_{\mathcal{C}}(C)$ , where  $C$  is a circuit of  $D$  is equal to  $\chi_c(\mathcal{C})$ .*

**Proof.** Consider an  $r$ -circular coloration  $f$  of  $\mathcal{C}$  with  $r = \chi_c(\mathcal{C})$  and  $E = v_1, \dots, v_n$  an enumeration of  $\mathcal{C}$  related to  $f$ . For a circuit  $C$  in  $D$ , we compute the length  $l$  of the image of  $C$  by  $f$ :

$$l := \sum_{\substack{xy \in E(C) \\ xy \text{ forward in } E}} (f(y) - f(x)) + \sum_{\substack{xy \in E(C) \\ xy \text{ backward in } E}} (r + f(y) - f(x))$$

A straightforward simplification of the sum gives  $l = r \cdot i_{\mathcal{C}}(C)$ . Furthermore, condition 1) of the definition of  $f$  implies that  $f(y) - f(x) \geq 1$  if  $xy \in E(C)$  and  $xy$  is forward in  $E$  (i.e.  $f(x) < f(y)$ ) and  $r + f(y) - f(x) \geq 1$  if  $xy \in E(C)$  and  $xy$  is backward in  $E$  (i.e.  $f(y) < f(x)$ ). So, we have  $c = r \cdot i_{\mathcal{C}}(C) \geq l(C)$ , hence  $r \geq l_{\mathcal{C}}(C)$ , and the inequality  $\chi_c(\mathcal{C}) \geq \max\{l_{\mathcal{C}}(C) : C \text{ circuit of } D\}$  holds.

To get the equality, we have to find a circuit  $C$  of  $D$  such that  $l_{\mathcal{C}}(C) = r$ .

First of all, since  $\mathcal{C}$  is coherent, it has a circuit of index 1 and, then the cyclic length of this circuit is greater or equal to 2. Thus, the previous inequality gives  $r \geq 2$  and so, states the case  $r = 2$ . From now on, assume that  $r > 2$ , Lemma 4 asserts that there exists a circuit  $C$  in the digraph  $D_f$ . So,

now, the inequalities provided in the direct sens of the proof are equalities: the image of every arc of  $C$  by  $f$  has length 1 if the arc is forward or  $r - 1$  if the arc is backward. So, we have  $l = l(C)$ , and  $l_C(C) = l(C)/i_C(C) = r$ , which achieves the bound.  $\square$

A corollary of Theorem 3 is an earlier result of J.A. Bondy, known since 1976. The *chromatic number* of a digraph  $D$ , denoted by  $\chi(D)$ , is the minimal number  $k$  such that the vertices of  $D$  admit a partition into  $k$  stable sets of  $D$ . Clearly,  $\chi(D) \leq \chi_c(C)$ .

**Corollary 3.1** (Bondy [5]) *Every strong digraph  $D$  has a circuit with length at least  $\chi(D)$ .*

**Proof.** Consider a coherent cyclic order  $\mathcal{C}$  for  $D$  and apply Theorem 3 to provide a circuit  $C$  with  $l_C(C) = \chi_c(\mathcal{C})$ . Since by Lemma 3  $\lceil l_C(C) \rceil = \lceil \chi_c(\mathcal{C}) \rceil = \chi(\mathcal{C})$ , we get  $\chi(D) \leq \chi(\mathcal{C}) = \lceil l_C(C) \rceil \leq l(C)$ .  $\square$

We gratefully thank J.A. Bondy who told us that a link could exist between [5] and Gallai's problem.

## References

- [1] J. Bang-Jensen and G. Gutin, Digraphs. Theory, algorithms and applications, *Springer Monographs in Mathematics*, Springer-Verlag London, 2001.
- [2] C. Berge, Path partitions in directed graphs, *Ann. Discrete Math.*, **17** (1983), 59–63.
- [3] J.A. Bondy, Basic graph theory: paths and circuits. *Handbook of combinatorics*, Vol. 1, 2, Elsevier, Amsterdam, 1995.
- [4] J.A. Bondy, A short proof of the Chen-Manalastas theorem, *Discrete Maths*, **146** (1995), 289–292.
- [5] J.A. Bondy, Disconnected orientations and a conjecture of Las Vergnas, *J. London Math. Soc.* (2), **14** (1976), 277–282.
- [6] P. Camion, Chemins et circuits hamiltoniens des graphes complets, *C. R. Acad. Sci.*, **249** (1959), 2151–2152.
- [7] C.C. Chen and P. Manalastas, Every finite strongly connected digraph of stability 2 has a Hamiltonian path, *Discrete Mathematics*, **44** (1983), 243–250.
- [8] R.P. Dilworth, A decomposition theorem for partially ordered sets, *Annals of Mathematics*, **51** (1950), 161–166.
- [9] T. Gallai, Problem 15, in: M. Fiedler, ed., *Theory of Graphs and its Applications* (Czech. Acad. Sci. Prague, 1964), 161.
- [10] T. Gallai and A.N. Milgram, Verallgemeinerung eines graphentheoretischen Satzes von Rédei, *Acta Sci. Math. Szeged*, **21** (1960), 181–186.
- [11] S. Thomassé, Spanning a strong digraph by  $\alpha$  cycles, the case  $\alpha = 3$  of Gallai's conjecture, preprint.
- [12] X. Zhu, Circular chromatic number: a survey, *Discrete Mathematics*, **229** (2001), 371–410.



## 2.2 Bermond-Thomassen conjecture for regular tournaments

### Two proofs of the Bermond-Thomassen conjecture for almost regular tournaments

Stéphane Bessy

LIRMM  
*Montpellier, France.*

Nicolas Lichiardopol

I3S-EPU  
*Sophia Antipolis, France.*

Jean-Sébastien Sereni

MASCOTTE I3S (CNRS-UNSA) – INRIA  
*Sophia Antipolis, France.*

---

#### Abstract

The Bermond-Thomassen conjecture states that, for any positive integer  $r$ , a digraph of minimum out-degree at least  $2r - 1$  contains at least  $r$  vertex-disjoint directed cycles. Thomassen proved that it is true when  $r = 2$ , and very recently the conjecture was proved for the case where  $r = 3$ . It is still open for larger values of  $r$ , even when restricted to (regular) tournaments. In this paper, we present two proofs of this conjecture for tournaments with minimum out-degree and minimum in-degree at least  $2r - 1$ . In particular, this shows that the conjecture is true for almost regular tournament. In the first proof, we prove auxiliary results about union of sets contained in other union of sets, that might be of independent interest. The second one uses a more graph-theoretical approach, by studying the properties of a maximum set of vertex-disjoint directed triangles.

*Key words:* digraph, circuit, tournament, the Bermond-Thomassen conjecture, cycle

*1991 MSC:* 05C20

---



---

*Email addresses:* `Stephane.Bessy@lirmm.fr` (Stéphane Bessy),  
`lichiar@club-internet.fr` (Nicolas Lichiardopol), `sereni@sophia.inria.fr`  
(Jean-Sébastien Sereni).

## 1 Introduction

In 1981, Bermond and Thomassen [1] conjectured that for any positive integer  $r$ , any digraph of minimum out-degree at least  $2r - 1$  contains at least  $r$  vertex-disjoint directed cycles. It is trivially true when  $r$  is one, and it was proved by Thomassen [2] when  $r$  is two in 1983. Very recently, the conjecture was also proved in the case where  $r$  is three [3]. It is still open for larger values of  $r$ . We prove, in two different ways, that the restriction of this conjecture to almost regular tournaments is true.

Chen, Gould and Li [4] proved that a  $k$ -strongly-connected tournament of order at least  $5k - 3$ , contains  $k$  vertex-disjoint directed cycles. Given a tournament  $T$ , let  $q(T)$  be the maximum order of a transitive subtournament of  $T$ . Li and Shu [5] showed that any strong tournament  $T$  of order  $n$  with  $q(T) \leq \frac{n-5k+8}{2}$  can be vertex-partitioned into  $k$  cycles. However, these results do not prove the Bermond-Thomassen conjecture for regular tournaments.

The following definitions are those of the monograph by Bang-Jensen and Gutin [6]. A *tournament* is a digraph  $T$  such that for any two distinct vertices  $x$  and  $y$ , exactly one of the couples  $(x, y)$  and  $(y, x)$  is an arc of  $T$ . The vertex set of  $T$  is  $V(T)$ , and its cardinality is the *order* of  $T$ . The set of arcs of  $T$  is  $\mathcal{A}(T)$ . A vertex  $y$  is a *successor* of a vertex  $x$  if  $(x, y)$  is an arc of  $T$ . A vertex  $y$  is a *predecessor* of a vertex  $x$  if  $x$  is a successor of  $y$ . The number of successors of  $x$  is the *out-degree*  $\delta^+(x)$  of  $x$ , and the number of predecessors of  $x$  is the *in-degree*  $\delta^-(x)$  of  $x$ . Let  $\delta^+(T) := \min\{\delta^+(x) : x \in V(T)\}$ ,  $\delta^-(T) := \min\{\delta^-(x) : x \in V(T)\}$  and  $\delta(T) := \min\{\delta^+(T), \delta^-(T)\}$ .

Given a tournament  $T$ , its *reversing tournament* is the tournament  $T' = (V(T), \mathcal{A}')$ , where  $\mathcal{A}' := \{(x, y) : (y, x) \in \mathcal{A}(T)\}$ . A tournament is *regular of degree  $d$*  if  $\delta^+(x) = \delta^-(x) = d$  for every vertex  $x$ . Necessarily, the order of such a tournament is  $2d + 1$ . It is *almost regular* if  $|\delta^+(x) - \delta^-(x)| \leq 1$  for every vertex  $x$ . An almost regular tournament of odd order is regular, and an almost regular tournament  $T$  of even order  $v$  is characterised by  $\delta^+(T) = \delta^-(T) = \frac{v}{2} - 1$ .

For any subset  $A$  of  $V(T)$ , we let  $T(A)$  be the sub-tournament induced by the vertices of  $A$ . By a path or a cycle of a tournament  $T$ , we mean a directed path or a directed cycle of  $T$ , respectively. By disjoint cycles, we mean vertex-disjoint cycles. A cycle of length three is a *triangle*.

A tournament is *acyclic*, or *transitive*, if it does not contain cycles, i.e. if its vertices can be ranged into a unique Hamiltonian path  $x_1, \dots, x_n$  such that  $(x_i, x_j)$  is an arc if and only if  $i < j$ . As is well-known, and straightforward to prove, a non-acyclic tournament contains a triangle. In particular, note that if a tournament contains  $k$  disjoint cycles, then it contains  $k$  disjoint triangles.

## 2 Preliminary results

Let  $(x, y)$  be an arc of a tournament  $T$ . We set

$$\begin{aligned} A(x, y) &:= \{z \in V(T) : (z, x) \in \mathcal{A}(T) \text{ and } (z, y) \in \mathcal{A}(T)\}, \\ B(x, y) &:= \{z \in V(T) : (x, z) \in \mathcal{A}(T) \text{ and } (y, z) \in \mathcal{A}(T)\}, \\ E(x, y) &:= \{z \in V(T) : (z, x) \in \mathcal{A}(T) \text{ and } (y, z) \in \mathcal{A}(T)\}, \text{ and} \\ F(x, y) &:= \{z \in V(T) : (x, z) \in \mathcal{A}(T) \text{ and } (z, y) \in \mathcal{A}(T)\}. \end{aligned}$$

Note that  $E(x, y)$  is the set of vertices  $z$  such that  $x, y$  and  $z$  form a triangle. We let  $a(x, y), b(x, y), e(x, y)$  and  $f(x, y)$  be the respective cardinalities of these four sets. The proof of the following proposition is straightforward, and can be found in [7], so we omit it.

**Proposition 1** *If  $(x, y)$  is an arc of a tournament, then  $e(x, y) = f(x, y) + \delta^+(y) - \delta^+(x) + 1$ .*

A set of cardinality  $m$  is an  $m$ -set. We give now three new results, which may be of independent interest. The first one is essential in our first proof of the Bermond-Thomassen conjecture for almost regular tournaments.

**Theorem 2** *Fix two integers  $m \geq 3$  and  $r \geq 1$ . Let  $n \in \{1, 2, \dots, r\}$  and  $s = \left\lceil \frac{r+m-1}{2} \right\rceil$ . For every  $i \in \{1, 2, \dots, n\}$ , let  $B_i$  be an  $m$ -set, and for every  $j \in \{1, 2, \dots, s\}$ , fix a set  $A_j \subseteq \bigcup_{1 \leq i \leq n} B_i$  of cardinality at least  $r + m + 1 - 2j$ . Then, there exist  $i \in \{1, 2, \dots, n\}$  and distinct elements  $j$  and  $k$  of  $\{1, 2, \dots, s\}$  such that  $B_i$  has distinct elements  $x$  and  $y$  with  $x \in A_j$  and  $y \in A_k$ .*

**Proof.** If  $n < r$ , then proving the result for the sets  $B'_1, B'_2, \dots, B'_r$  with  $B'_i = B_i$  if  $i \leq n$  and  $B'_i = B_n$  if  $i > n$  will yield the desired conclusion. So, we suppose now that  $n = r$ , and we use induction on  $r$ .

Observe that it is sufficient to prove that there exist  $i \in \{1, 2, \dots, n\}$  and distinct integers  $j, k \in \{1, 2, \dots, s\}$  such that  $|A_j \cap B_i| \geq 1$  and  $|A_k \cap B_i| \geq 2$ .

The assertion is true when  $r = 1$ . Indeed, in this case,  $s = \left\lceil \frac{1+m-1}{2} \right\rceil = \left\lceil \frac{m}{2} \right\rceil \geq 2$ ,  $|A_1| \geq m \geq 3$ ,  $|A_2| \geq m - 2 \geq 1$  and  $B_1$  is an  $m$ -set such that  $A_i \subseteq B_1$  for  $i \in \{1, 2, \dots, s\}$ . Therefore,  $|A_1 \cap B_1| \geq 3$  and  $|A_2 \cap B_1| \geq 1$ , which yields the desired conclusion.

The assertion is true also for  $r = 2$ . Indeed, in this case,  $s = \left\lceil \frac{2+m-1}{2} \right\rceil = \left\lceil \frac{m+1}{2} \right\rceil \geq 2$ ,  $|A_1| \geq m + 1 \geq 4$ ,  $|A_2| \geq m - 1 \geq 2$  and  $A_1 \cup A_2 \subseteq B_1 \cup B_2$ . Clearly,  $A_1 \cap B_1 \neq \emptyset$  — otherwise  $B_2$  would contain  $A_1$ , which has at least  $m+1$  elements — and similarly,  $A_1 \cap B_2 \neq \emptyset$ . If  $|A_1 \cap B_1| \geq 2$  and  $|A_1 \cap B_2| \geq 2$ ,

then the result holds. Otherwise, we have, say,  $|A_1 \cap B_1| = 1$  and hence  $|A_1 \cap B_2| = m$ . Now, either  $|A_2 \cap B_1| \geq 2$  or  $|A_2 \cap B_2| \geq 1$ , so the result holds.

Suppose now that the assertion is true for every  $k < r$ , for some integer  $r \geq 3$ , and let us prove it for  $r$ . Then,  $s = \lceil \frac{r+m-1}{2} \rceil \geq 3$ ,  $|A_1| \geq r + m - 1$  and  $|A_2| \geq r + m - 3 \geq r$ . Without loss of generality, we assume that  $|B_1 \cap A_1| \geq |B_2 \cap A_1| \geq \dots \geq |B_r \cap A_1|$ .

Suppose first that  $|B_2 \cap A_1| \leq 1$ . Then,  $B_2 \cup \dots \cup B_r$  contains at most  $r - 1$  elements of  $A_1$  and  $B_1 \cup B_2 \cup \dots \cup B_r$  contains at least  $r + m - 1$  elements of  $A_1$ . So, we deduce that  $|B_1 \cap A_1| = m$  and  $|B_i \cap A_1| = 1$  for every  $i \in \{2, 3, \dots, r\}$ . The assertion of the theorem holds if  $|B_1 \cap A_2| \geq 1$ . If  $|B_1 \cap A_2| = 0$ , then there exists  $i \in \{2, 3, \dots, r\}$ , such that  $|B_i \cap A_2| \geq 2$  — otherwise we would have  $|(B_1 \cup B_2 \cup \dots \cup B_r) \cap A_2| \leq r - 1$ , a contradiction. Clearly,  $B_i$  contains distinct elements  $x$  and  $y$  with  $x \in A_1$  and  $y \in A_2$ .

Suppose now that  $|B_2 \cap A_1| \geq 2$ . In this case,  $|B_1 \cap A_1| \geq 2$ ,  $|B_2 \cap A_1| \geq 2$  and the desired conclusion holds if  $B_1 \cup B_2$  contains an element of  $A_2 \cup \dots \cup A_s$ . If  $B_1 \cup B_2$  does not contain an element of  $A_2 \cup \dots \cup A_s$ , let  $A'_i := A_{i+1}$  for  $i \in \{1, 2, \dots, s - 1\}$ . We have  $s - 1 = \lceil \frac{r-2+m-1}{2} \rceil$ ,  $|A'_i| \geq r - 2 + m + 1 - 2i$  and  $A'_i \subseteq \bigcup_{3 \leq j \leq r} B_j$  for  $i \in \{1, 2, \dots, s - 1\}$ . Therefore, by induction hypothesis there exist  $i \in \{3, \dots, r\}$  and distinct elements  $j$  and  $k$  of  $\{2, \dots, s\}$  such that  $B_i$  contains distinct elements  $x$  and  $y$  with  $x \in A_j$  and  $y \in A_k$ , which concludes the proof.  $\square$

The second and third results can be proved analogously, and we omit their proofs.

**Theorem 3** *Fix two integers  $m \geq 3$  and  $r \geq 2$ . Let  $n \in \{1, 2, \dots, r\}$ , and for every  $i \in \{1, 2, \dots, n\}$ , denote by  $B_i$  an  $m$ -set. For every  $j \in \{1, 2, \dots, r\}$ , let  $A_j \subseteq \bigcup_{1 \leq i \leq n} B_i$  with  $|A_j| \geq r + m + 1 - 2j$ . Then, there exist  $i \in \{1, \dots, n\}$  and distinct elements  $j$  and  $k$  of  $\{1, \dots, r\}$  such that  $B_i$  has distinct elements  $x$  and  $y$  with  $x \in A_j$  and  $y \in A_k$ .*

The best result is a combination of the first two.

**Theorem 4** *Fix two integers  $m \geq 3$  and  $r \geq 2$ . Let  $n \in \{1, 2, \dots, r\}$  and set  $s = \min \left( \lceil \frac{r+m-1}{2} \rceil, r \right)$ . For  $i \in \{1, 2, \dots, n\}$ , denote by  $B_i$  an  $m$ -set, and for every  $j \in \{1, 2, \dots, s\}$ , let  $A_j \subseteq \bigcup_{1 \leq i \leq n} B_i$  with  $|A_j| \geq r + m + 1 - 2j$ . Then, there exist  $i \in \{1, \dots, n\}$  and distinct elements  $j$  and  $k$  of  $\{1, \dots, s\}$  such that  $B_i$  has distinct elements  $x$  and  $y$  with  $x \in A_j$  and  $y \in A_k$ .*

### 3 Disjoint cycles in tournaments $T$ with $\delta(T) \geq 2r - 1$

In this section, we give two different proofs of the following result.

**Theorem 5** *For any  $r \geq 1$ , every tournament  $T$  with  $\delta(T) \geq 2r - 1$  contains  $r$  disjoint cycles.*

**Proof.** The case  $r = 1$  being a simple observation, we assume that  $r \geq 2$ . Let  $v$  be the order of  $T$ , and let  $n$  be the maximum number of disjoint cycles of  $T$ . Thus,  $n$  is also the maximum number of disjoint triangles: let  $T_i$ ,  $i \in \{1, 2, \dots, n\}$  be  $n$  disjoint triangles. Let  $V' := V(T) \setminus \bigcup_{1 \leq j \leq n} V(T_j)$  and  $p := |V'|$ . Suppose that  $n \leq r - 1$ . Thus,  $p \geq v - 3(r - 1)$ , that is  $p \geq r + 2$ , since  $v \geq 4r - 1$ . The subtournament  $T(V')$  is acyclic — otherwise, we would have an extra cycle — and, consequently, its vertices can be ranged into a Hamiltonian path  $x_1, \dots, x_p$  such that  $(x_i, x_j)$  is an arc of  $T(V')$  if and only if  $i < j$ , see Figure 1.

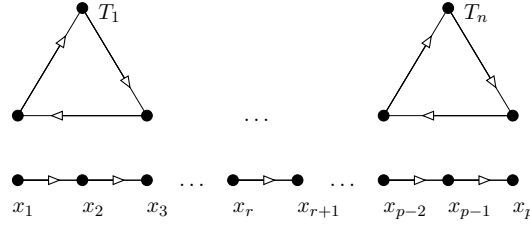


Fig. 1. Disjoint triangles and Hamiltonian path of  $T(V')$

For  $i \in \{1, 2, \dots, \lceil \frac{r+1}{2} \rceil\}$ , consider the arc  $(x_i, x_{p+1-i})$ : each vertex  $x_j$  with  $j \in \{i+1, i+2, \dots, r+2-i\}$  belongs to  $F(x_i, x_{p+1-i})$ . Therefore,

$$f(x_i, x_{p+1-i}) \geq p - 2i \geq v - 3n - 2i.$$

By Proposition 1,

$$e(x_i, x_{p+1-i}) \geq p - 2i + \delta^+(x_{p+1-i}) - \delta^+(x_i) + 1.$$

Since  $2r - 1 \leq \delta^+(x) \leq v - 2r$  for every vertex  $x$ , we deduce that

$$e(x_i, x_{p+1-i}) \geq v - 3n - 2i + 2r - 1 - (v - 2r) + 1 \geq (r - 1) + 3 + 1 - 2i,$$

as  $n \leq r - 1$ .

Observe now that every vertex of  $E(x_i, x_{p+1-i})$  forms a triangle with the vertices  $x_i$  and  $x_{p+1-i}$ . Moreover, as  $T(V')$  is acyclic, we have  $E(x_i, x_{p+1-i}) \subseteq \bigcup_{1 \leq j \leq n} V(T_j)$  for  $i \in \{1, 2, \dots, \lceil \frac{r+1}{2} \rceil\}$ . Hence, the conditions of Theorem 2

are fulfilled — the  $r$  of the theorem being  $r - 1$ ,  $m$  being three,  $s = \lceil \frac{r+1}{2} \rceil$ ,  $A_i = E(x_i, x_{p+1-i})$  and  $B_j = V(T_j)$ . Consequently, with  $s = \lceil \frac{r+1}{2} \rceil$ , there exist  $i \in \{1, \dots, n\}$  and distinct elements  $j$  and  $k$  of  $\{1, \dots, s\}$  such that  $V(T_i)$  contains distinct vertices  $x$  and  $y$  with  $x \in E(x_j, x_{p+1-j})$  and  $y \in E(x_k, x_{p+1-k})$ . Each  $T_q$ , for  $q \in \{1, 2, \dots, n\} \setminus \{i\}$ , and the tournaments induced by  $x_j, x_{p+1-j}, x$  and by  $x_k, x_{p+1-k}, y$  are  $n + 1$  disjoint triangles, which contradicts the definition of  $n$ . Therefore,  $T$  contains at least  $r$  disjoint cycles, as desired.  $\square$

**Second proof.** As mentioned in the Introduction, Thomassen [2] proved the conjecture in the general case for  $r \leq 2$ , and the general case for  $r = 3$  was recently proved [3]. Thus, we assume in this proof that  $r \geq 4$ .

Suppose that  $V'$  is a subset of at least 6 vertices such that  $T(V')$  is acyclic. Let  $\{x_1, x_2, \dots, x_p\}$  be the vertices of  $V'$ , indexed such that  $(x_i, x_j)$  is an arc if and only if  $i < j$ . We set  $A_{V'} := \{x_1, x_2, x_3\}$  and  $B_{V'} := \{x_{p-2}, x_{p-1}, x_p\}$ . For a vertex  $x$ , let  $s_{V'}^-(x)$  be the *in-score of  $x$  with respect to  $V'$* , that is the number of predecessors of  $x$  in  $B_{V'}$ . Analogously,  $s_{V'}^+(x)$  is the *out-score of  $x$  with respect to  $V'$* , that is the number of successors of  $x$  in  $A_{V'}$ . Given a subgraph  $H$  of  $T$ , the *in-score of  $H$  with respect to  $V'$*  is

$$s_{V'}^-(H) := \sum_{x \in V(H)} s_{V'}^-(x).$$

We define  $s_{V'}^+(H)$ , the *out-score of  $H$  with respect to  $V'$* , analogously regarding the out-scores of the vertices of  $H$ . Last, the *score of  $H$  with respect to  $V'$*  is  $s_{V'}(H) = s_{V'}^-(H) + s_{V'}^+(H)$ . In all these notations, we may omit the subscript if the context is clear.

As in the first proof, let  $n$  be the maximum number of disjoint triangles, and consider a family  $T_i$ ,  $i \in \{1, 2, \dots, n\}$ , of  $n$  disjoint triangles. We set  $V' := V(T) \setminus \bigcup_{1 \leq j \leq n} V(T_j)$  and  $p := |V'|$ . Again, we consider the Hamiltonian path  $x_1, \dots, x_p$  of the acyclic tournament  $T(V')$  such that  $(x_i, x_j)$  is an arc of  $T(V')$  if and only if  $i < j$ .

Suppose that  $n \leq r - 1$ . Then, we obtain that  $p \geq 4r - 1 - 3(r - 1)$ , that is  $p \geq r + 2$ , and hence  $p \geq 6$  since  $r \geq 4$ .

For each triangle  $T_i$ , we have  $s^-(T_i) \leq 9$  and  $s^+(T_i) \leq 9$ . If  $s^-(T_i) \geq 7$  and  $s^+(T_i) \geq 4$ , then there exists a matching of size three from  $B_{V'}$  to  $T_i$ , and a matching of size two from  $T_i$  to  $A_{V'}$ . Therefore,  $T(A_{V'} \cup B_{V'} \cup V(T_i))$  contains two disjoint triangles, which contradicts the maximality of  $n$ . Thus, either  $s^-(T_i) \leq 6$  or  $s^+(T_i) \leq 3$ . Similarly, either  $s^+(T_i) \leq 6$  or  $s^-(T_i) \leq 3$ .

We assert that  $s(T_i) \leq 12$  for each triangle  $T_i$ : indeed, if  $s^-(T_i) > 6$ , then  $s^+(T_i) \leq 3$ , and since  $s^-(T_i) \leq 9$ , we infer that  $s(T_i) \leq 12$ . In the same way, if  $s^+(T_i) > 6$ , one can deduce that  $s(T_i) \leq 12$ . Finally, if  $s^-(T_i) \leq 6$  and  $s^+(T_i) \leq 6$ , we also have  $s(T_i) \leq 12$ . Hence, the sum  $s$  of the scores of the  $n$  triangles is at most  $12n$ .

Observe that the vertices  $x_p, x_{p-1}$  and  $x_{p-2}$  have  $\delta^+(x_p), \delta^+(x_{p-1}) - 1$  and  $\delta^+(x_{p-2}) - 2$  successors in  $\bigcup_{1 \leq j \leq n} V(T_j)$ , respectively. Moreover, the vertices  $x_1, x_2$  and  $x_3$  have respectively  $\delta^-(x_1), \delta^-(x_2) - 1$  and  $\delta^-(x_3) - 2$  predecessors in  $\bigcup_{1 \leq j \leq n} V(T_j)$ . It follows that

$$s = \delta^+(x_p) + \delta^+(x_{p-1}) + \delta^+(x_{p-2}) + \delta^-(x_1) + \delta^-(x_2) + \delta^-(x_3) - 6.$$

Therefore, it holds that

$$\delta^+(x_p) + \delta^+(x_{p-1}) + \delta^+(x_{p-2}) + \delta^-(x_1) + \delta^-(x_2) + \delta^-(x_3) - 6 \leq 12n \leq 12r - 12.$$

Recall that  $\delta^+(x) \geq 2r - 1$  and  $\delta^-(x) \geq 2r - 1$  for every vertex  $x$ . Thus, we infer that  $\delta^+(x_p) = \delta^+(x_{p-1}) = \delta^+(x_{p-2}) = \delta^-(x_1) = \delta^-(x_2) = \delta^-(x_3) = 2r - 1$ ,  $n = r - 1$  and  $s(T_i) = 12$  for every triangle  $T_i$ . Note that this assertion holds for any set on  $n$  disjoint triangles — their score being with respect to the remaining vertices.

For each integer  $i \in \{4, 5, \dots, p-3\}$ , the vertex  $x_i$  belongs to  $F(x_3, x_{p-2})$ , and hence  $f(x_3, x_{p-2}) \geq p - 6$ . Therefore, by Proposition 1,

$$\begin{aligned} e(x_3, x_{p-2}) &\geq p - 6 + \delta^+(x_{p-2}) - \delta^+(x_3) + 1 \\ &\geq v - 3(r - 1) - 6 + (2r - 1) - (v - 1 - 2r + 1) + 1 \\ &\geq r - 3 \geq 1. \end{aligned}$$

Consequently, there exists a vertex  $x$  of some triangle  $T_j$  such that the vertices  $x_3, x_{p-2}, x$  induce a triangle  $T'$ . Let  $y$  and  $z$  be the vertices of  $T_j$  different from  $x$ . The triangles  $T'$  and  $T_i$  for  $i \neq j$  form a new collection of  $n$  disjoint triangles, and  $V'' := (V' \setminus \{x_3, x_{p-2}\}) \cup \{y, z\}$  is the set of the remaining vertices. Consider now the set  $A_{V''}$ : observe that  $x_3$  has at most two successors in  $A_{V''}$ , and it can have two only if both  $y$  and  $z$  belong to  $A_{V''}$ . Furthermore, the predecessors of  $x_3$  in  $B_{V''}$  can only be  $y$  and  $z$ . Therefore, it follows that  $s_{V''}^-(x_3) + s_{V''}^+(x_3) \leq 3$  with equality only if both  $y$  and  $z$  belong to  $B_{V''}$ . Similarly,  $s_{V''}^-(x_{p-2}) + s_{V''}^+(x_{p-2}) \leq 3$  with equality only if both  $y$  and  $z$  belong to  $A_{V''}$ . Thus, the score of the triangle  $T'$  with respect to  $V''$  is at most 11, a contradiction. This concludes the proof.  $\square$

### References

- [1] J.-C. Bermond, C. Thomassen, Cycles in digraphs—a survey, *J. Graph Theory* 5 (1) (1981) 1–43.
- [2] C. Thomassen, Disjoint cycles in digraphs, *Combinatorica* 3 (3-4) (1983) 393–396.
- [3] N. Lichiardopol, A. Pór, J.-S. Sereni, A step towards the Bermond-Thomassen conjecture about disjoint cycles in digraphs, submitted to *Combinatorica. ITI Series* 2007-339 (2007).
- [4] G. Chen, R. J. Gould, H. Li, Partitioning vertices of a tournament into independent cycles, *J. Combin. Theory Ser. B* 83 (2) (2001) 213–220.
- [5] H. Li, J. Shu, The partition of a strong tournament, *Discrete Math.* 290 (2-3) (2005) 211–220.
- [6] J. Bang-Jensen, G. Gutin, *Digraphs*, Springer-Verlag London Ltd., London, 2001, theory, algorithms and applications.
- [7] N. Lichiardopol, Dominated and dominating arcs in tournaments, characterization of regular tournaments, vertex strong connectivity of a Paley tournament, submitted to *J. Graph Theory*.



## 2.3 A proof of Bermond-Thomassen conjecture for tournaments

### Disjoint 3-cycles in tournaments: a proof of the Bermond-Thomassen conjecture for tournaments\*

Jørgen Bang-Jensen<sup>†</sup>   Stéphane Bessy<sup>‡</sup>   Stéphan Thomassé<sup>§</sup>

July 22, 2011

#### Abstract

We prove that every tournament with minimum out-degree at least  $2k - 1$  contains  $k$  disjoint 3-cycles. This provides additional support for the conjecture by Bermond and Thomassen that every digraph  $D$  of minimum out-degree  $2k - 1$  contains  $k$  vertex disjoint cycles. We also prove that for every  $\epsilon > 0$ , when  $k$  is large enough, every tournament with minimum out-degree at least  $(1.5 + \epsilon)k$  contains  $k$  disjoint cycles. The linear factor 1.5 is best possible as shown by the regular tournaments.

**Keywords:** Disjoint cycles, tournaments.

## 1 Introduction

Notation not given below is consistent with [3]. Paths and cycles are always directed unless otherwise specified. In a digraph  $D = (V, A)$ , a  $k$ -cycle is a cycle of length  $k$ , and for  $k \geq 3$ , we denote by  $x_1x_2 \dots x_k$  the  $k$ -cycle on  $\{x_1, \dots, x_k\}$  with arc set  $\{x_1x_2, x_2x_3, \dots, x_{k-1}x_k, x_kx_1\}$ . The minimum length of a cycle in  $D$  is called the **girth** of  $D$ . The **underlying graph** of a digraph  $D$ , denoted  $UG(D)$ , is obtained from  $D$  by suppressing the orientation of each arc and deleting multiple edges. For a set  $X \subseteq V$ , we use the notation  $D\langle X \rangle$  to denote the subdigraph of  $D$  **induced** by the vertices in  $X$ . For two disjoint sets  $X$  and  $Y$  of vertices of  $D$ , we say that  $X$  **dominates**  $Y$  if  $xy$  is an arc of  $D$  for every  $x \in X$  and every  $y \in Y$ . In the digraph  $D$ , if  $X$  and  $Y$  are two disjoint subsets of vertices of  $D$  or subdigraphs of  $D$ , we say that there is a  $k$ -**matching** from  $X$  to  $Y$  if the set of arcs from  $X$  to  $Y$  contains a matching (in  $UG(D)$ ) of size at least  $k$ . A **tournament** is an orientation of a complete graph, that is a digraph  $D$  such that for every pair  $\{x, y\}$  of distinct vertices of  $D$  either  $xy \in A(D)$  or  $yx \in A(D)$ , but not both. Finally, an **out-neighbour** (resp. **in-neighbour**) of a vertex  $x$  of  $D$  is a vertex  $y$  with  $xy \in A(D)$  (resp.  $yx \in A(D)$ ). The **out-degree** (resp. **in-degree**)  $d_D^+(x)$  (resp.  $d_D^-(x)$ ) of a vertex  $x \in V$  is the number of out-neighbours (resp. in-neighbours) of  $x$ . We denote by  $\delta^+(D)$  the minimum out-degree of a vertex in  $D$ .

The following conjecture, due to J.C. Bermond and C. Thomassen, gives a relationship between  $\delta^+$  and the maximum number of vertex disjoint cycles in a digraph.

**Conjecture 1.1 (Bermond and Thomassen, 1981)** [4] *If  $\delta^+(D) \geq 2k - 1$  then  $D$  contains  $k$  vertex disjoint cycles.*

Remark that the complete digraph (with all the possible arcs) shows that this statement is best possible. The conjecture is trivial for  $k = 1$  and it has been verified for general digraphs when  $k = 2$

\*Part of this work was done while the first author was on sabbatical at AIGCo, LIRMM, Université Montpellier 2, France whose hospitality is gratefully acknowledged. Financial support from the Danish National Science research council (FNU) (under grant no. 09-066741) is gratefully acknowledged.

<sup>†</sup>Department of Mathematics and Computer Science, University of Southern Denmark, Odense DK-5230, Denmark (email: jbj@imada.sdu.dk).

<sup>‡</sup>AIGCo, LIRMM, Université Montpellier 2, France (email: bessy@lirmm.fr). Financial support: ANR GRATOS NANR-09-JCJC-0041-01

<sup>§</sup>Laboratoire LIP (U. Lyon, CNRS, ENS Lyon, INRIA, UCBL), Lyon, France (email: stephan.thomasse@ens-lyon.fr)

in [8] and for  $k = 3$  in [7]. N. Alon proved in [1] that a lower bound of  $64k$  on the minimum outdegree gives  $k$  disjoint cycles.

It was shown in [5] that every tournament with both minimum out-degree and minimum in-degree at least  $2k - 1$  has  $k$  disjoint cycles each of which have length 3. Very recently Lichiardopol [6] obtained a generalization of this result to the existence of  $k$  disjoint cycles of prescribed length  $q$  in a tournament with sufficiently high minimum degree.

In this paper we will prove Conjecture 1.1 for tournaments. Recall that by Moon's Theorem [3, Theorem 1.5.1], a tournament has  $k$  disjoint cycles if and only if it has  $k$  disjoint 3-cycles.

**Theorem 1.2** *Every tournament  $T$  with  $\delta^+(T) \geq 2k - 1$  has  $k$  disjoint cycles each of which have length 3.*

We also show how to improve this result for tournaments with large minimum out-degree.

**Theorem 1.3** *For every value  $\alpha > 1.5$ , there exists a constant  $k_\alpha$ , such that for every  $k \geq k_\alpha$ , every tournament  $T$  with  $\delta^+(T) \geq \alpha k$  has  $k$  disjoint 3-cycles.*

Remark that the constant 1.5 is best possible in the previous statement. Indeed, a family of sharp examples is provided by the rotative tournaments  $TR_{2p+1}$  on  $2p + 1$  vertices  $\{x_1, \dots, x_{2p+1}\}$  with arc set  $\{x_i x_j : j - i \bmod 2p + 1 \in \{1, \dots, p\}\}$ . For  $2p + 1 = 0 \bmod 3$ , we denote  $2p + 1/3$  by  $k$ . Then, we have  $\delta^+(TR_{2p+1}) = \lfloor 1.5k \rfloor$  and  $TR_{2p+1}$  admits a partition into  $k$  vertex disjoint 3-cycles and no more.

Theorem 1.3 does not give any result both for small values of  $k$  and for tournaments with  $\delta^+ \geq 1.5k$ , even asymptotically. We conjecture that we could still have  $k$  disjoint 3-cycles in these cases. Furthermore, in the light of the sharp examples to Conjecture 1.1 and Theorem 1.3, we extend these questions to digraphs with no short cycles. Namely, we conjecture the following.

**Conjecture 1.4** *For every integer  $g > 1$ , every digraph  $D$  with girth at least  $g$  and with  $\delta^+(D) \geq \frac{g}{g-1}k$  contains  $k$  disjoint cycles.*

Once again, the constant  $\frac{g}{g-1}$  is best possible. Indeed, for every integers  $p$  and  $g$ , we define the digraph  $D_{g,p}$  on  $n = p(g - 1) + 1$  vertices with vertex set  $\{x_1, \dots, x_n\}$  and arc set  $\{x_i x_j : j - i \bmod n \in \{1, \dots, p\}\}$ . The digraph  $D_{g,p}$  has girth  $g$  and we have  $\delta^+(D_{g,p}) = p = \lfloor \frac{g}{g-1}k \rfloor$ . Moreover, for  $n = 0 \bmod g$ , the digraph  $D_{g,p}$  admits a partition into  $k$  vertex disjoint 3-cycles and no more.

Even a proof of Conjecture 1.4 for large values of  $k$  or  $g$  (or both) would be of interest by itself. On the other hand, for  $g = 3$ , the first case of our conjecture which differs from Conjecture 1.1 and which is not already known corresponds to the following question: does every digraph  $D$  without 2-cycles and  $\delta^+(D) \geq 6$  admit four vertex disjoint cycles?

In Section 2 and Section 3, we respectively prove Theorem 1.2 and Theorem 1.3. Before starting these, we precise notations that will be used in both next sections. Let  $T$  be a tournament and  $\mathcal{F}$  a maximal collection of 3-cycles of  $T$ . The 3-cycles of  $\mathcal{F}$  are denoted by  $C_1, \dots, C_p$  and their ground set  $V(C_1) \cup \dots \cup V(C_p)$  is denoted by  $W$ . The remaining part of  $T$ ,  $T \setminus W$  is denoted by  $U$ . By the choice of  $\mathcal{F}$ ,  $U$  induces an acyclic tournament on  $T$ , and we denote its vertices by  $\{u_p, u_{p-1}, \dots, u_2, u_1\}$ , such that the arc  $u_i u_j$  exists if and only if  $i > j$ .

## 2 Proof of Conjecture 1.1 for tournaments

In this section, we prove Conjecture 1.1 for tournaments. In fact, we strengthen a little bit the statement and prove the following:

**Theorem 2.1** *For every tournament  $T$  with  $\delta^+(T) \geq 2k - 1$  and every collection  $\mathcal{F} = \{C_1, \dots, C_{k-1}\}$  of  $k - 1$  disjoint 3-cycles of  $T$ , there exists a collection of  $k$  disjoint 3-cycles of  $T$  which intersects  $T - V(C_1) \cup \dots \cup V(C_{k-1})$  on at most 4 vertices.*

This result implies Theorem 1.2. Indeed, for a tournament  $T$  with  $\delta^+(T) \geq 2k_0 - 1$ , we apply Theorem 2.1  $k_0$  times, with  $k = 1$  to obtain a family  $\mathcal{F}$  of one 3-cycle, and then with this family  $\mathcal{F}$  and  $k = 2$  to obtain a new family  $\mathcal{F}$  of two 3-cycles, and so on.

To prove Theorem 2.1, we consider a counter-example  $T$  and a family  $\mathcal{F}$  of  $k - 1$  disjoint 3-cycles with  $k$  minimum. The chosen family  $\mathcal{F}$  is then maximal. So, from now on we use the notation stated in the first Section.

We will say that  $i$  3-cycles of  $\mathcal{F}$ , with  $i = 1$  or  $i = 2$  can be **extended** if we can make  $i + 1$  3-cycles using the vertices of the initial  $i$  3-cycles and at most four vertices of  $U$ . If there one or two 3-cycles in  $\mathcal{F}$  can be extended, we say that we could extend  $\mathcal{F}$ . If this happens, it would contradict the choice of  $T$  and  $\mathcal{F}$ . The following definition will be very useful in all this section. For an arc  $xy$  with  $x, y \in W$ , we say that a vertex  $z$  of  $U$  is a **breaker** of  $xy$  if  $xyz$  forms a 3-cycle. By extension, a vertex  $z$  of  $U$  is a **breaker** of a 3-cycle  $C_i$  of  $\mathcal{F}$  if it is a breaker of one of the arcs of  $C_i$ .

The following claim is fundamental, and we will use it later several times without explicit mention.

**Claim 1** *Every 3-cycle  $C$  of  $\mathcal{F}$  has breakers for at most two of its three arcs, and every arc of  $C$  has at most three breakers. As a consequence,  $C$  has at most six breakers.*

**Proof:** Consider a 3-cycle  $C_i = xyz$  of  $\mathcal{F}$ . Assume that  $C_i$  has a breaker for each of its arcs. We denote by  $v_e$  a breaker of the arc  $e$ , for  $e \in \{xy, yz, zx\}$ . If  $v_{yz}$  dominates  $v_{zx}$  then we form the 3-cycles  $xyv_{xy}$  and  $zv_{yz}v_{zx}$ , which intersect  $U$  on three vertices and we extend  $\mathcal{F}$ . So, by symmetry, we obtain that  $v_{zx}v_{yz}v_{xy}$  forms a 3-cycle. This contradicts that  $T\langle U \rangle$  is acyclic. Now, if an arc  $xy$  of  $C_i$  has four breakers  $v_1, v_2, v_3, v_4$  in  $U$ , then in  $T \setminus \{x, y\}$  every vertex has out-degree at least  $2(k - 1) - 1$ , and  $\mathcal{F} \setminus C_i$  forms a collection of  $k - 2$  3-cycles. So, by the choice of  $T$ , there exists a collection  $\mathcal{F}'$  of  $k - 1$  3-cycles of  $T \setminus \{x, y\}$  which intersect  $U \cup z$  in at most four vertices. Then  $\mathcal{F}'$  does not contain one of the vertices  $v_1, v_2, v_3, v_4, z$ . If  $z \notin V(\mathcal{F}')$ , we complete  $\mathcal{F}'$  with the 3-cycle  $xyz$ , and obtain a collection of  $k$  3-cycles which has the same intersection with  $U$  than  $\mathcal{F}'$ . If  $z \in V(\mathcal{F}')$ , then one of the  $v_i$ , say  $v_1$  does not belong to  $V(\mathcal{F}')$  and  $\mathcal{F}'$  intersect  $U$  on at most three vertices. Then, we complete  $\mathcal{F}'$  with the 3-cycle  $xyv_1$ , and obtain a collection of  $k$  3-cycles which intersect  $U$  on at most four vertices.  $\diamond$

Observe that if a 3-cycle  $xyz$  of  $\mathcal{F}$  has a breaker for two of its arcs, then these breakers are disjoint. Indeed, if  $x'$  and  $y'$  are respectively breaker of  $xy$  and  $yz$  then  $yx'$  and  $y'y$  are arcs of  $T$ . As  $T$  has no 2-cycle,  $x'$  and  $y'$  have to be distinct.

Informally, Claim 1 gives that every 3-cycle  $C$  of  $\mathcal{F}$  can be extended or can be inserted in the transitive tournament  $T\langle U \rangle$ , that is, there exists a partition  $(U_2, U_1)$  of  $U$  such that there is no arc from  $U_1$  to  $U_2$ , there is few arcs from  $U_1$  to  $C$  and few from  $C$  to  $U_2$  (otherwise, too roughly many breakers appear). This will be settled at Claim 2. The condition on the minimum out-degree of  $T$  will then allow one or two 3-cycles of  $\mathcal{F}$  to be extended. Fixing precisely the computation will show, in the following subsection, that  $k$  cannot be too large ( $k \leq 6$ ). Then, we treat the small cases in the last subsection.

## 2.1 A bound on $k$

For any partition  $(U_1, U_2)$  of  $U$  with no arc from  $U_1$  to  $U_2$ , we have the following.

**Claim 2** *For every 3-cycle  $C = xyz$  of  $\mathcal{F}$ , we have:*

1. *If  $C$  receives at least four arcs from  $U_1$  then there exists a 2-matching from  $U_1$  to  $C$ .*
2. *If  $C$  receives at least eight arcs from  $U_1$  then either there exists a 3-matching from  $U_1$  to  $C$  or, up to permutation on  $x, y, z$ ,  $yz$  has three breakers,  $xy$  has at least two breakers and  $x$  has in-degree at least five in  $U_1$ . Furthermore,  $x$  is dominated by  $U_2$  and both  $y$  and  $z$  have each at most one out-neighbour in  $U_2$ .*

3. Consequently, if  $C$  receives at least eight arcs from  $U_1$  then, there is no 2-matching from  $C$  to  $U_2$  and, in particular,  $C$  sends at most three arcs to  $U_2$ .

Symmetrically, the same statements hold if we exchange the role of  $U_1$  and  $U_2$ , and the bounds on in- and out-neighbours for every vertex.

**Proof:** 1. Assume that there is no 2-matching from  $U_1$  to  $C$  then one vertex  $x$  of  $U_1 \cup C$  belongs to all the arcs from  $U_1$  to  $C$ . It is clear that  $x \in C$ . Hence if  $y$  is the successor of  $x$  in  $C$ , then four in-neighbours of  $x$  in  $U_1$  form four breakers for the arc  $xy$ , which is not possible.

2. If there is no 3-matching from  $U_1$  to  $C$ , then two vertices  $\{x, y\}$  in  $U_1 \cup C$  belongs to all arcs from  $U_1$  to  $C$ . If  $x \in U_1$  and  $y \in C$ , then there exists at least four in-neighbours of  $y$  different of  $x$  which form four breakers for the arc  $yz$ , where  $z$  is the successor of  $y$  in  $C$ , which is forbidden. As the case  $\{x, y\} \subset U_1$  is not possible, we have  $\{x, y\} \subset C$ . Assume that  $x$  dominates  $y$  and call  $z$  the third vertex of  $C$ . If  $d_{U_1}^-(y) \leq 2$ , then  $d_{U_1}^-(x) \geq 6$  and  $xy$  has four breakers, which is not possible. If  $d_{U_1}^-(y) \geq 4$ ,  $yz$  has four breakers. So  $d_{U_1}^-(y) = 3$  and  $d_{U_1}^-(x) \geq 5$  which means that  $yz$  has three breakers and that  $x$  has at least two in-neighbours in  $U_1$  which are not in-neighbours of  $y$ , and so, are breakers of  $xy$ . If  $x$  has an out-neighbour  $x'$  in  $U_2$ , we extend  $C$  using the 3-cycles  $xx'x_1$  and  $yzx_1$  where  $x_1$  and  $y_1$  are breakers of respectively  $xy$  and  $yz$ . So  $U_2 \Rightarrow x$  must hold (that is, there is no arc from  $x$  to  $U_2$ ). Now, if  $y$  has two out-neighbours in  $U_2$ , they form two more breakers for  $xy$ , and  $xy$  would have four breakers. Finally, if  $z$  has two out-neighbours in  $U_2$ , one of these is in-neighbour of  $y$  and would form a new breaker for  $yz$ , which had already three.

3. Assume that  $C$  receives at least eight arcs from  $U_1$  and that there is a 2-matching from  $C$  to  $U_2$ . If there exists a 3-matching from  $U_1$  to  $C$ , then we can extend  $C$  using at most four vertices of  $U$ . If not, then we are in the case described in the point 2, and  $C$  has at least five breakers in  $U_1$ , three for  $yz$  and at least two for  $xy$ . We can conclude except if the 2-matching from  $C$  to  $U_2$  starts from  $y$  and  $z$ . We denote it by  $\{yy', zz'\}$ . If  $z'y$  is an arc of  $T$ , then  $yz$  would have four breakers. Then  $yz'$  is an arc of  $T$ , but then, as  $U_2$  dominates  $x$ , the vertices  $y'$  and  $z'$  would be two breakers of  $xy$ , which already has two.  $\diamond$

The two following claims are useful to extend two 3-cycles of  $\mathcal{F}$  in order to form three new 3-cycles.

**Claim 3** *There are no two 3-cycles  $C$  and  $C'$  of  $\mathcal{F}$  with a 3-matching from  $U_1$  to  $C$ , a 3-matching from  $C$  to  $C'$  and a 3-matching from  $C'$  to  $U_2$ .*

**Proof:** If this happens, we respectively denote these matchings by  $\{x_1x, y_1y, z_1z\}$ ,  $\{xx', yy', zz'\}$  and  $\{x'_2x, y'_2y, z'_2z\}$ , where  $V(C) = \{x, y, z\}$ ,  $V(C') = \{x', y', z'\}$ ,  $x_1, y_1, z_1 \in U_1$  and  $x_2, y_2, z_2 \in U_2$ . If all three of  $\{x_2x, y_2y, z_2z\}$  are arcs of  $T$ , then we can extend  $C$  and  $C'$  by  $xx_2x'$ ,  $y_2yy'$  and  $zz_2z'$ . So, we can assume that  $xx_2$  is an arc of  $T$ . If one of the arcs  $yy_2$  or  $zz_2$  exists then, we can extend  $C$ . So,  $xx_2$ ,  $y_2y$  and  $z_2z$  are arcs of  $T$  and we extend  $C$  and  $C'$  using the 3-cycles  $xx_2x_1$ ,  $y_2yy'$  and  $zz_2z'$ .  $\diamond$

**Claim 4** *There are no two 3-cycles  $C$ ,  $C'$  such that  $|E(U_1, C)| \geq 8$ ,  $|E(C, C')| \geq 7$  and  $|E(C', U_2)| \geq 8$ .*

**Proof:** Assume that  $C$  and  $C'$  satisfy the hypothesis of the claim. We denote  $V(C) = \{x, y, z\}$  and  $V(C') = \{x', y', z'\}$ . As  $|E(C, C')| \geq 7$  there is a 3-matching between  $C$  and  $C'$ . By the Claim 3, one cannot both find a 3-matching from  $U_1$  to  $C$  and a 3-matching from  $C'$  to  $U_2$ . By symmetry, two cases arise:

*Case 1:* there are no 3-matching from  $U_1$  to  $C$  and from  $C'$  to  $U_2$ . We fix the orientations of  $C$  and  $C'$ :  $C = xyz$  and  $C' = x'y'z'$ . By Claim 2, up to permutation, we can assume that  $yz$  has three breakers in  $U_1$  and  $xy$  at least two, and that  $x'y'$  has three breakers in  $U_2$  and  $y'z'$  at least two. Furthermore we know, by Claim 2 that  $U_2$  dominates  $x$ ,  $z$  has at most one out-neighbour in  $U_2$ ,  $z'$  dominates  $U_1$  and  $x'$  has at most one in-neighbour in  $U_1$ . We denote then by  $x_1$  a breaker of

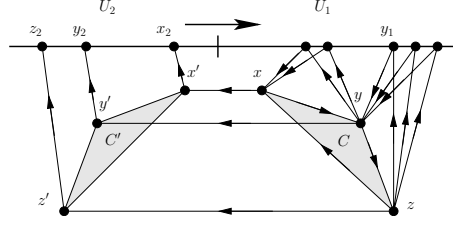


Figure 1: The case 2 of the proof of Claim 4

$xy$  in  $U_1$  which is an out-neighbour of  $x'$ , and by  $z_2$  a breaker of  $y'z'$  in  $U_2$  which is an in-neighbour of  $z$ . We denote also by  $y_2$  and  $y_1$  a breaker of respectively  $x'y'$  and  $yz$ . Now, if  $xz'$  is an arc of  $T$ , then, we form the 3-cycles  $xz'z_2$ ,  $y_1yz$  and  $x'y'y_2$ . If  $xx'$  is an arc of  $T$ , then, we form the 3-cycles  $xx'x_1$ ,  $y'z'z_2$  and  $yz y_1$ . And, if  $zz'$  is an arc of  $T$ , then, we form the 3-cycles  $zz'z_2$ ,  $x'y'y_2$  and  $xyx_1$ . As  $|E(C, C')| \geq 7$ , one of the three arcs  $xz'$ ,  $xx'$  and  $zz'$  exists and we can extend  $C$  and  $C'$ .

*Case 2:* there is no 3-matching from  $U_1$  to  $C$  and there is a 3-matching from  $C'$  to  $U_2$ . We fix the orientation of  $C$ ,  $C = xyz$ , but we do not fix the orientation of  $C'$ . We just assume that  $\{xx', yy', zz'\}$  is a 3-matching between  $C$  and  $C'$ . We denote by  $\{x'x_2, y'y_2, z'z_2\}$  a 3-matching from  $C'$  to  $U_2$ . By Claim 2, up to permutation, we can assume that  $yz$  has three breakers in  $U_1$ , we denote by  $y_1$  one of them, and that  $xy$  has at least two. Furthermore we know, that  $d_{U_1}(x) \geq 5$ , that  $U_2$  dominates  $x$  and that  $y$  and  $z$  have at most one out-neighbour in  $U_2$ . The situation is depicted in Figure 1.

To obtain a contradiction, we follow the next implications:

- $zz_2$  is an arc of  $T$ , otherwise we form the three circuits  $zz'z_2$ ,  $xx'x_2$  and  $yy'y_2y_1$ , which contain three 3-cycles intersecting  $U$  on at most four vertices.
- $yz_2$  is an arc of  $T$ , otherwise  $z_2$  is a fourth breaker of  $yz$ .
- $x_2$  and  $y_2$  dominate  $y$  and  $z$ . Indeed, the only out-neighbour of  $y$  and  $z$  in  $U_2$  is  $z_2$ .
- $\{y', y_2, z, z'\}$  form an acyclic tournament. Indeed if  $\{y', y_2, z, z'\}$  contains a circuit, we pick this circuit,  $xx'x_2$  and  $yz_2y_1$  to extend  $C$  and  $C'$ . In particular, the orientation of  $C'$  is  $x'y'z'$  and  $y'z \in A(T)$ .
- $xy'$  is an arc of  $T$ . Otherwise,  $y'z$  and  $y'x$  are the only arcs from  $C'$  to  $C$  and we form the 3-cycles  $xz'z_2$ ,  $zx'x_2$  and  $yy'y_2$  to extend  $C$  and  $C'$ .
- $z'x_2$  is an arc of  $T$ . Otherwise, we form the 3-cycles  $z'x'x_2$ ,  $xy'y_2$  and  $yz y_1$ .

Finally, we extend  $C$  and  $C'$  using the 3-cycles  $zz'x_2$ ,  $xy'y_2$  and  $yz_2y_1$ .  $\diamond$

Now, we will show that  $k \leq 6$ . For this, we consider the partition  $(U_2, U_1)$  of  $U$  with  $|U_1| = 5$  (as  $W$  contains  $3k - 3$  vertices, and  $T$  has at least  $4k - 1$  vertices,  $U$  contains at least  $k + 2$  vertices, and provided that  $k \geq 3$ , it is possible to consider such a  $U_1$ ). So, we denote by  $\mathcal{I}$  the set of 3-cycles which receive at least 8 arcs each from  $U_1$  (the **in 3-cycles**), by  $\mathcal{O}$  the set of 3-cycles which send at least 8 arcs each to  $U_2$  (the **out 3-cycles**) and by  $\mathcal{R}$  the remaining 3-cycles of  $\mathcal{F} \setminus (\mathcal{I} \cup \mathcal{O})$ . Furthermore,  $i$ ,  $o$  and  $r$  respectively denote the size of  $\mathcal{I}$ ,  $\mathcal{O}$  and  $\mathcal{R}$  (with  $i + o + r = k - 1$  as  $\mathcal{I} \cap \mathcal{O} = \emptyset$  by Claim 2). First, we bound below and above the number of arcs leaving  $U_1$ , and obtain:

$$5(2k - 1) - 10 \leq 15i + 7(k - 1 - i - o) + 3o$$

In the right part, we bound the number of arcs from  $U_1$  to  $\mathcal{I}$ , to  $\mathcal{R}$  and to  $\mathcal{O}$  (using Claim 2). Finally, we have:

$$3k + 4o - 8 \leq 8i \quad (1)$$

Now, we bound below and above the number of arcs leaving  $\mathcal{F} \setminus \mathcal{O}$  and obtain

$$3(k-1-o)(2k-1) - \frac{1}{2}3(k-1-o)(3(k-1-o)-1) \leq 9ro + 6io + 7r + 3i + (15(i+r) - (10k-15-3o))$$

In the right part, we bound the number of arcs from  $\mathcal{R}$  to  $\mathcal{O}$ , from  $\mathcal{I}$  to  $\mathcal{O}$  (using Claim 4), from  $\mathcal{R}$  to  $U_2$ , from  $\mathcal{I}$  to  $U_2$  (using Claim 2) and from  $\mathcal{I} \cup \mathcal{R}$  to  $U_1$ . For the last bound, we know that at least  $5(2k-1) - 10 = 10k - 15$  arcs leave  $U_1$  and that at most  $3o$  of these arcs go to  $\mathcal{O}$ . So, at least  $10k - 15 - 3o$  arcs go from  $U_1$  to  $\mathcal{I} \cup \mathcal{R}$  on the  $15(i+r)$  possible arcs between these two parts. Now, we replace  $r$  by  $k-1-i-o$  and obtain:

$$9o^2 - 12ko + 6io + 41o + 3k^2 - 21k + 8i + 8 \leq 0$$

We bound  $i$  from below using (1) to get (after adjusting to get integral coefficients):

$$16o^2 - 13ko + 52o + 4k^2 - 24k \leq 0$$

This inequality admits solution for  $o$  only if

$$(52 - 13k)^2 - 4 \cdot 16 \cdot (4k^2 - 24k) = -87k^2 + 184k + 2704$$

is positive, that is, if  $k \leq 6$ .

## 2.2 Small cases

Below we handle the cases  $k \leq 6$ . The partition  $(U_1, U_2)$  is no more fixed by  $|U_1| = 5$ , we will specify its size later.

### 2.2.1 Some remarks

We need some more general statements to solve the cases  $k \leq 6$ . For the following Claim 5 and Claim 6, symmetric statements hold if we exchange the roles of  $U_1$  and  $U_2$ , and the bounds on in- and out-neighbours for every vertex.

**Claim 5** *If  $|E(U_1, C)| \geq 10$ , then there exists a 3-matching from  $U_1$  to  $C$ .*

**Proof:** Otherwise, two vertices,  $\{x, y\}$ , belong to all arcs from  $U_1$  to  $C$ . As  $\{x, y\} \subset U_1$  is not possible (otherwise only at most 6 arcs go from  $U_1$  to  $C$ ), either  $x \in U_1$  and  $y \in C$  or  $\{x, y\} \subset C$ . In the first case,  $y$  has at least seven in-neighbours in  $U_1$  distinct of  $x$ , and if  $z$  is the out-neighbour of  $y$  in  $C$ , these seven vertices would be breakers of  $yz$ , contradicting Claim 1. So, we have  $\{x, y\} \subset C$ . We assume that  $x$  dominates  $y$  and that the orientation of  $C$  is  $C = xyz$ . Then  $y$  has at most three in-neighbours in  $U_1$ , otherwise  $yz$  would have four breakers, and  $x$  has at most three in-neighbours in  $U_1$  which are not also in-neighbours of  $y$ , otherwise  $xy$  would have four breakers. But then there are at most nine arcs from  $U_1$  to  $C$ , contradicting the hypothesis.  $\diamond$

As for Claim 2, it is possible to obtain the same result by exchanging  $U_1$  and  $U_2$  and the role of in- and out-neighbours for every vertex.

We say that a 3-cycle  $C$  has a **3-cover** from  $U_1$  if there is a 3-matching from  $U_1$  to  $C$  or two 2-matchings from  $U_1$  to  $C$  which cover all the vertices of  $C$ .

**Claim 6** *For every 3-cycle  $C$  of  $\mathcal{F}$ , if there is a 3-cover from  $U_1$  to  $C$ , then there is no 2-matching from  $C$  to  $U_2$ . In particular,  $|E(C, U_2)| \leq 3$ .*

**Proof:** Assume that  $C = xyz$  and that there is a 2-matching  $\{zz', xx'\}$  from  $C$  to  $U_2$  and a 3-cover from  $U_1$  to  $C$ . If there is a 2-matching from  $U_1$  to  $\{z, x\}$ , we are done. The remaining case occurs when the 3-cover from  $U_1$  to  $C$  is formed by a 2-matching  $\{ax, by\}$  to  $\{x, y\}$  and a 2-matching  $\{cy, dz\}$  to  $\{y, z\}$  with  $a = d$ . In this case, we form the circuits  $axx'$  and  $byzz'$ , which contain two 3-cycles extending  $C$ . The bound on  $|E(C, U_2)|$  follows from Claim 2.  $\diamond$

For a fixed  $U_1$ , we say that a 3-cycle  $C$  of  $\mathcal{F}$  is of type **2-m**, **3-m** or **3-c** if there respectively is a 2-matching, a 3-matching or a 3-cover from  $U_1$  to  $C$ . A 3-cover is useful to extend a 3-cycle, using Claim 6, but not very convenient in the general case, because the number of arcs that forces a 3-cover from  $U_1$  to some 3-cycle  $C$  of  $\mathcal{F}$  is the same than the number of arcs that forces a 3-matching (which is seven). However, to prove the existence of a 3-cover, we have the following statement.

**Claim 7** *If there are three vertices  $a, b, c$  of  $U_1$  such that  $d_Y^+(a) \geq 2p$ ,  $d_Y^+(b) \geq 2p - 1$  and  $d_Y^+(c) \geq 2p - 2$ , where  $Y$  is the set of vertices of a set of  $p$  3-cycles  $\mathcal{F}' \subset \mathcal{F}$ , then  $\mathcal{F}'$  contains a 3-c 3-cycle or all the 3-cycles of  $\mathcal{F}'$  are 2-m.*

**Proof:** We prove it by induction on  $p$ . If  $p = 1$  then there is a 2-matching from  $\{a, b\}$  to the 3-cycle of  $\mathcal{F}'$ . Thus we may assume that  $p \geq 2$ . There is  $6p - 3$  arcs from  $\{a, b, c\}$  to the  $p$  3-cycles of  $\mathcal{F}'$ . Thus there is a 3-cycle  $C$  of  $\mathcal{F}'$  such that there are at least four arcs from  $\{a, b, c\}$  to  $C$  and so there is a 2-matching from  $\{a, b, c\}$  to  $C$ . If  $C$  is 3-c, we are done, otherwise each vertex of  $\{a, b, c\}$  sends at most two arcs to  $C$ . We apply induction on  $\mathcal{F}' \setminus C$ .  $\diamond$

Now we are ready to prove the remaining cases ( $k \leq 6$ ). As mentioned in the beginning of the paper, Conjecture 1.1 is known to hold for all digraphs when  $k \leq 3$ , so we only have to deal with the cases  $k \in \{4, 5, 6\}$ .

We will use several times, without referring explicitly, that a 3-cycle of type respectively 2-m and 3-c or 3-m sends respectively at most 7 and 3 arcs to  $U_2$ , by Claim 6 and 2. For each of the three cases below, we will use the three first vertices of  $U$  for  $U_1$ , that is,  $U_1 = \{u_1, u_2, u_3\}$ .

### 2.2.2 Case $k = 4$

For  $k = 4$ , we have  $\delta^+(T) \geq 7$  and three 3-cycles in  $\mathcal{F}$ . There are:

- at least  $21 - 3 = 18$  arcs from  $U_1$  to  $W$  and then at most 9 arcs from  $W$  to  $U_1$ .
- at least  $9 \cdot 7 - \frac{1}{2}9 \cdot 8 = 27$  arcs from  $W$  to  $U$  and then, at least 18 arcs from  $W$  to  $U_2$ .

So it is not possible to have types 3-c, 2-m and 2-m for the three 3-cycles of  $\mathcal{F}$ , otherwise, they send at most  $3 + 7 + 7 = 17$  arcs to  $U_2$ . Now we prove that there are at least two 3-cycles of type 3-c. As  $u_1$  sends seven arcs to  $W$ , one of the 3-cycle, say  $C_1$  receives 3 arcs. If  $u_2$  or  $u_3$  sends one arc to  $C_1$ , then  $C_1$  is of type 3-c, if not, then  $C_2$  and  $C_3$  are of type 3-c. So, at least one of the three 3-cycle is of type 3-c, we assume that it is  $C_1$ . Note that  $u_1, u_2$  and  $u_3$  send respectively at least 4, 3 and 2 arcs to  $C_2 \cup C_3$ . Using Claim 7, we find a second 3-cycle which is of type 3-c. We assume that this second one is  $C_2$ . Now, we have:

- there is no 2-matching from  $U_1$  to  $C_3$ , then  $C_3$  receives at most 3 arcs from  $U_1$ , and then  $C_1 \cup C_2$  receive at least 15 arcs from  $U_1$ , what means that there is a 3-matching from  $U_1$  to  $C_1$  for instance.
- $C_1 \cup C_2$  sends at least  $6 \cdot 7 - \frac{1}{2}6 \cdot 5 = 27$  arcs to  $U \cup C_3$ , at most 3 to  $U_1$  and 6 to  $U_2$ , what means that there all the arcs from  $C_1 \cup C_2$  to  $C_3$
- $C_3$  sends at least  $18 - 3 - 3 = 12$  arcs to  $U_2$ , then, by Claim 5, there is a 3-matching from  $C_3$  to  $U_2$ .

Finally, using 3-matchings from  $U_1$  to  $C_1$ , from  $C_1$  to  $C_3$  and from  $C_3$  to  $U_2$  and Claim 3, we can extend  $C_1, C_2$  and  $C_3$ .

**2.2.3 Case  $k = 5$** 

For  $k = 5$ , we have  $\delta^+(T) \geq 9$  and four 3-cycles in  $\mathcal{F}$ . There is:

- at least 24 arcs from  $U_1$  to  $W$  and then at most 12 arcs from  $W$  to  $U_1$ .
- at least  $12 \cdot 9 - \frac{1}{2}12 \cdot 11 = 42$  arcs from  $W$  to  $U$  and then, at least 30 arcs from  $W$  to  $U_2$ .

So, it is not possible to have types 2-m, 2-m, 2-m and 2-m for the four 3-cycles of  $\mathcal{F}$ , otherwise, they send at most  $7 + 7 + 7 + 7 = 28$  arcs to  $U_2$ . There are no three type 3-c among the four 3-cycles of  $\mathcal{F}$ . Otherwise, assume that  $C_1, C_2$  and  $C_3$  are of type 3-c, then,  $C_4$  can not be of type 2-m, and there are at most 3 arcs from  $U_1$  to  $C_4$  and at least 21 arcs from  $U_1$  to  $C_1 \cup C_2 \cup C_3$ . Then,  $C_1 \cup C_2 \cup C_3$  sends at most 3 arcs to  $U_1$ , at most 9 arcs to  $U_2$  and at most 27 arcs to  $C_4$ . However, there is at least  $9 \cdot 9 - \frac{1}{2}9 \cdot 8 = 45$  arcs going out of  $C_1 \cup C_2 \cup C_3$ , what gives a contradiction.

Using Claim 7 twice, we find two 3-cycles,  $C_1$  and  $C_2$  for instance, in  $\mathcal{F}$  that are of type 3-c. Now,  $u_1, u_2$  and  $u_3$  respectively send at least 3, 2 and 1 arc to  $C_3$  and  $C_4$  and it is easy to find a 2-matching from  $U_1$  to  $C_3$  or  $C_4$ .

Now, we assume that  $C_1$  and  $C_2$  have a 3-cover from  $U_1$  and that  $C_3$  have a 2-matching from  $U_1$ . We obtain:

- $C_4$  receives at most three arcs from  $U_1$  (otherwise  $C_4$  would be a fourth 3-cycle of type 2-m).
- $U_1$  sends at least 21 arcs to  $C_1 \cup C_2 \cup C_3$ , then there is a 3-matching from  $U_1$  to one of these 3-cycle, say  $C_1$  and there is at most 6 arcs from  $C_1 \cup C_2 \cup C_3$  to  $U_1$ .
- there is at most  $3 + 3 + 7 = 13$  arcs from  $C_1 \cup C_2 \cup C_3$  to  $U_2$ , and then as there is at least  $9 \cdot 9 - \frac{1}{2}9 \cdot 8 = 45$  arcs going out of  $C_1 \cup C_2 \cup C_3$ , there is  $45 - 6 - 13 = 26$  arcs from  $C_1 \cup C_2 \cup C_3$  to  $C_4$ . In particular, there is a 3-matching from  $C_1$  to  $C_4$ .
- there are at most 13 arcs from  $C_1 \cup C_2 \cup C_3$  to  $U_2$ , so, there are at least 17 arcs from  $C_4$  to  $U_2$  and then a 3-matching from  $C_4$  to  $U_2$ .

Finally, we extend  $C_1$  and  $C_4$  using 3-matchings from  $U_1$  to  $C_1$ , from  $C_1$  to  $C_4$  and from  $C_4$  to  $U_2$ .

**2.2.4 Case  $k = 6$** 

For  $k = 6$ , we have  $\delta^+(T) \geq 11$  and five 3-cycles in  $\mathcal{F}$ . There is:

- at least 30 arcs from  $U_1$  to  $W$  and then at most 15 arcs from  $W$  to  $U_1$ .
- at least  $15 \cdot 11 - \frac{1}{2}15 \cdot 14 = 60$  arcs from  $W$  to  $U$  and then, at least 45 arcs from  $W$  to  $U_2$ .

Finding five 3-cycles of type 2-m in  $\mathcal{F}$  is not possible then, because we would have at most  $7 \cdot 5 = 35$  arcs from  $W$  to  $U_2$ . We will see that there are either at least three 3-cycles which are of type 3-c or there are two 3-cycles of type 3-c and two 3-cycles of type 2-m. Using Claim 7 twice, we find two 3-cycles which are of type 3-c, say  $C_1$  and  $C_2$ . There remains at least 5, 4 and 3 arcs from respectively  $u_1, u_2$  and  $u_3$  to  $C_3 \cup C_4 \cup C_5$ . One of the 3-cycles  $C_3, C_4$  or  $C_5$ , say  $C_3$ , receives at least 4 arcs from  $\{u_1, u_2, u_3\}$  and then is of type 2-m. If  $C_3$  is of type 3-c, we are done, otherwise, it receives at most 2 arcs from each of  $u_1, u_2, u_3$ , and  $u_1, u_2$  and  $u_3$  respectively send at least 3, 2 and 1 arcs to  $C_4 \cup C_5$ . We then find another 3-cycle of type 2-m.

First, we consider the case where there are two 3-cycles of type 3-c,  $C_1$  and  $C_2$  and two 3-cycles of type 2-m,  $C_3$  and  $C_4$ . Then, we have:

- $C_5$  receives at most 3 arcs from  $U_1$  (otherwise there is a fifth 3-cycle of type 2-m).
- $U_1$  sends at least 27 arcs to  $C_1 \cup C_2 \cup C_3 \cup C_4$ , thus there is at most 9 arcs from  $C_1 \cup C_2 \cup C_3 \cup C_4$  to  $U_1$ .
- there are at most  $3 + 3 + 7 + 7 = 20$  arcs from  $C_1 \cup C_2 \cup C_3 \cup C_4$  to  $U_2$ .



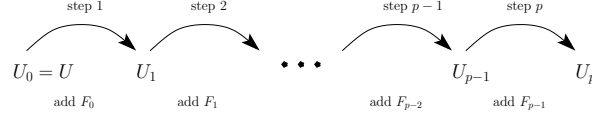


Figure 2: The  $p$  steps in the procedure to define free vertices.

- as there are at least  $11 \cdot 12 - \frac{1}{2}12 \cdot 11 = 66$  arcs going out of  $C_1 \cup C_2 \cup C_3 \cup C_4$ , there are at least  $66 - 9 - 20 = 37$  arcs from  $C_1 \cup C_2 \cup C_3 \cup C_4$  to  $C_5$ , which is not possible.

Now, we treat the case where there are three 3-cycles of type 3-c in  $\mathcal{F}$ ,  $C_1$ ,  $C_2$  and  $C_3$ . Then, we obtain:

- $C_4$  and  $C_5$  receive each at most 3 arcs from  $U_1$  (otherwise we are in one of the previous situations).
- $U_1$  sends at least 24 arcs to  $C_1 \cup C_2 \cup C_3$ . Thus there is a 3-matching from  $U_1$  to two of these 3-cycles, say  $C_1$  and  $C_2$  and there are at most 3 arcs from  $C_1 \cup C_2 \cup C_3$  to  $U_1$ .
- there are at most  $3 + 3 + 3 = 9$  arcs from  $C_1 \cup C_2 \cup C_3$  to  $U_2$ , and then as there are at least  $9 \cdot 11 - \frac{1}{2}9 \cdot 8 = 63$  arcs going out of  $C_1 \cup C_2 \cup C_3$ , there are  $63 - 3 - 9 = 51$  arcs from  $C_1 \cup C_2 \cup C_3$  to  $C_4 \cup C_5$ . In particular, there is, a 3-matching from any of 3-cycle of  $\{C_1, C_2, C_3\}$  to any of the 3-cycle of  $\{C_4, C_5\}$ , excepted possibly for one pair, say  $C_2$  to  $C_4$ , to be in the worst case.
- there are at least  $45 - 9 = 36$  arcs from  $C_4 \cup C_5$  to  $U_2$ , so, there are at least 18 arcs from one of the 3-cycle  $C_4$  or  $C_5$  to  $U_2$ , say from  $C_4$ , and then there is a 3-matching from  $C_4$  to  $U_2$ .

Finally, we extend  $C_1$  and  $C_4$  using 3-matchings from  $U_1$  to  $C_1$ , from  $C_1$  to  $C_4$  and from  $C_4$  to  $U_2$ .

### 3 Proof of Theorem 1.3: An asymptotic better constant

In this part, we will asymptotically ameliorate the result of Theorem 1.2 by proving Theorem 1.3.

Let  $\alpha$  be a real number with  $\alpha > 1.5$ , and  $T$  be a tournament with  $\delta^+(T) \geq \alpha k$ . We assume that  $\alpha < 2$ , otherwise Theorem 1.2 gives

We consider a family  $\mathcal{F}$  of less than  $k$  disjoint 3-cycles in  $T$ . We will see that if  $k$  is great enough, then we can extend  $\mathcal{F}$ . As usual, we denote by  $W$  the set of vertices of all the 3-cycles of  $\mathcal{F}$ , and by  $U$  the other vertices that form an acyclic part (otherwise, we directly extend  $\mathcal{F}$ ). As  $\delta^+(T) \geq \alpha k$ , remark that  $T$  has at least  $2\alpha k$  vertices and then, as  $|W| \leq 3k - 3$ , the size of  $U$  is at least  $(2\alpha - 3)k$ . The main idea of the proof is to obtain (almost) a partition of  $W$  into two parts  $X_1$  and  $X_2$  such that, as previously,  $X_1$  receives many arcs from  $U$  and  $X_2$  sends many arcs to  $U$ , with the requirement that the 3-cycles of  $\mathcal{F}$  behave well with respect to the partition. The 3-cycles (or parts of the 3-cycles) of  $X_1$  will act as in-3-cycles and the 3-cycles of  $X_2$  as out-3-cycles. If we assume that  $\mathcal{F}$  is maximum, a contradiction will result by computing the number of arcs leaving  $X_1$ .

We chose a positive real number  $\epsilon$  such that  $\epsilon < (\alpha - 1.5)/4$ . This value corresponds to the room that we have to ignore some vertices, which we will do several times during the proof. Then we fix an integer  $p$  with  $(3 - \alpha)/p < \epsilon/3$ , and we will repeat  $p$  times the procedure described below to define **free vertices**. We define three families of sets:

- $(F_i)_{0 \leq i \leq p-1}$  the free vertices produced at step  $i$ ,
- $(U_i)_{0 \leq i \leq p}$  the free vertices produced since the beginning (they will form an acyclic part), and
- $(W_i)_{0 \leq i \leq p}$  the remaining vertices, see Figure 2.

We initialize by setting  $U_0 = U$ , and  $W_0 = W$ . For  $0 \leq i \leq p-1$ , a vertex  $x$  of  $W_i$  (resp. an arc  $xy$  of  $W_i$ ) is **good at step  $i$**  if there exists at least  $3^{p+1}$  disjoint pairs of vertices  $\{y, z\}$  (resp. distinct vertices  $z$ ) of  $U_i$  such that  $\{x, y, z\}$  induces a 3-cycle. In other words, an element (vertex or arc) is good if it is contained in at least  $3^{p+1}$  3-cycles which are disjoint on  $U_i$ . When we find good elements, we will split the 3-cycles they are involved in into the good vertices (or vertices belonging to a good arc), that we will keep in  $W_{i+1}$ , and the others, called later free vertices and that we put with the transitive part  $U_{i+1}$ . For a 3-cycle  $C$  of  $\mathcal{F}$ , the vertices of  $C$  which we keep in  $W_{i+1}$  form the **remainder** of  $C$ . The remainder of  $C$  can contain one or two vertices. We use the name a **1-remainder** for a remainder of a 3-cycle with one vertex and a **2-remainder** for a remainder with 2 vertices.

Then, for  $i = 0, \dots, p-1$ , we initialise  $F_i = \emptyset$  and perform the step  $i$  of the procedure below, that is, we apply the first of the following rules as long as possible and then we consider the second rule, apply it as long as possible and proceed similarly for the third and fourth rule. When it is no more possible to apply the fourth rule, the step  $i$  is over, and we deal with the step  $i+1$ .

**Rule 3.1** *If a 3-cycle or a 2-remainder  $C$  belonging to  $W_i$  contains a vertex  $x$  which is good at step  $i$ , then we add  $V(C) \setminus \{x\}$  to  $F_i$ .*

**Rule 3.2** *If  $C, C'$  and  $C''$  are 3-cycles or 2-remainders belonging to  $W_i$  and  $T < V(C) \cup V(C') \cup V(C'') >$  contains three disjoint arcs, say  $xy, x'y'$  and  $x''y''$ , which are good at step  $i$ , then we add  $V(C) \cup V(C') \cup V(C'') \setminus \{x, x', x'', y, y', y''\}$  to  $F_i$ .*

**Rule 3.3** *If  $C$  and  $C'$  are 3-cycles or 2-remainders belonging to  $W_i$  and  $T < V(C) \cup V(C') >$  contains two disjoint arcs, say  $xy$  and  $x'y'$ , which are good at step  $i$ , then we add  $V(C) \cup V(C') \setminus \{x, x', y, y'\}$  to  $F_i$ .*

**Rule 3.4** *If a 3-cycle  $C$  of  $\mathcal{F}$  belonging to  $W_i$  contains a good arc  $xy$  at step  $i$ , then we add  $V(C) \setminus \{x, y\}$  to  $F_i$ .*

Now, we fix the sets  $U_{i+1}$  to  $U_i \cup F_i$  and  $W_{i+1}$  to  $W_i \setminus F_i$ . Furthermore, we call  $U_i$  the **free vertices at step  $i$** . The next claim shows that these vertices are 'free to form a 3-cycle'.

**Claim 8** *If the final set of free vertices,  $U_p$ , contains a 3-cycle, then, we can extend the family  $\mathcal{F}$ .*

**Proof:** Assume that  $U_p$  contains a 3-cycle  $xyz$ , we will build a family  $\mathcal{F}'$  of 3-cycles with  $|\mathcal{F}'| = |\mathcal{F}| + 1$ . The family  $\mathcal{F}'$  initially contains  $xyz$  and all the 3-cycles of  $\mathcal{F}$  that still exist in  $W_p$ . We will inductively complete  $\mathcal{F}'$  with 3-cycles formed from remainings of 3-cycles of  $\mathcal{F}$  that are in  $W_p$  by going step by step backward from the step  $p$  to the initial configuration. A vertex of  $U_p \setminus U_0 = \bigcup_{i=0}^{p-1} F_i$  is called **busy** if it is currently contained in a 3-cycle of  $\mathcal{F}'$ . At the end of step  $p$ , only  $x, y, z$  are possibly busy (and only if they do not belong to  $U_0$ ), and, for  $i = 1, \dots, p$  we will prove the following (where stage  $i$  corresponds to the  $i$ th level of undoing the steps performed above, starting with stage 1 where we undo step  $p$ ):

*At stage  $i$ , every remainder created at step  $p-i+1$  is contained in a 3-cycle of  $\mathcal{F}'$   $(\star)$   
or in a 2-remainder previously created and  $U_{p-i}$  contains at most  $3^{i+1}$  busy vertices.*

Let us see what happens when  $i = 1$ . If,  $\{x, y, z\} \cap F_{p-1} = \emptyset$ , then using the vertices of  $F_{p-1}$  and the corresponding remainders we undo step  $p-1$  to re-create original 3-cycles, which we add to  $\mathcal{F}'$  or 2-remainders previously created (if Rule 3.1 has been used on a 2-remainder at step  $p-1$ ). So, in this case, the only possible busy vertices of  $U_{p-1}$  are  $x, y$  and  $z$  and the property  $(\star)$  holds for  $i = 1$ . Otherwise, consider a busy vertex in  $\{x, y, z\}$  which is contained in  $F_{p-1}$ . It became free through the application of one of the Rules 3.1, 3.2, 3.3 or 3.4. In each of these cases, it has been separated from good elements (vertex or arc(s)), and these good elements can be re-completed into 3-cycles by adding at most three vertices (two for Rule 3.1, three for Rule 3.2, two for Rule 3.3 and one for Rule 3.4). Each of these good elements can be completed into at least  $3^{p+1}$  disjoint (on  $U_{p-1}$ ) 3-cycles. Hence, it is always possible to complete them disjointly with vertices of  $U_{p-1}$ . In the worst

case, 3 vertices were busy in the beginning ( $x$ ,  $y$  and  $z$ ) and each of the corresponding good element needs 3 vertices in  $U_{p-1}$  to be completed, producing 9 busy elements in  $U_{p-1}$ . Finally, the vertices of  $F_{p-1}$  that are not busy are used to re-create 3-cycles or 2-remainders destroyed at step  $p$ . For  $i = 2, \dots, p-1$ , we apply exactly same arguments to pass from stage  $i$  to stage  $i+1$ , provided that at each stage  $i$  at most  $3^{i+1} \leq 3^{p+1}$  busy vertices are present in  $U_{p-i}$ . For the last stage, that is to undo step 1, everything is similar, except that, by definition,  $U_0$  contains no busy vertices and hence the corresponding vertices can be directly taken to form the last 3-cycles of  $\mathcal{F}'$ . Finally,  $\mathcal{F}'$  contains one 3-cycle for each remainder in  $W_p$  and  $xyz$ , so  $|\mathcal{F}'| = |\mathcal{F}| + 1$ .  $\diamond$

An immediate consequence of Claim 8 is that the size of set  $W_p$  can not be less than  $\alpha \cdot k$ , because the first vertex of  $U_p$  has its out-neighbour-hood contained in  $W_p$ . So, the number of free vertices added to  $U_0 = U$ , that is  $\cup_{i=0}^{p-1} F_i$ , is at most  $(3-\alpha)k$ , and thus there is a step  $i_0+1$  with  $0 \leq i_0 \leq p-1$ , with  $|F_{i_0}| < (3-\alpha)k/p < \epsilon k/3$ . We stop just before this step  $i_0+1$ , and denote by  $\mathcal{R}$  the set of 3-cycles or 2-remainders with at least one vertex in  $F_{i_0}$ . So, the size of  $R = V(\mathcal{R})$ , is at most  $\epsilon k$ . We symbolically remove the small set  $R$  and go on working on the other 3-cycles and remainders. Remark that, now, in  $W_{i_0} \setminus R$  there are no more free elements.

For any  $q \leq p$ , we say that a set of vertices (or abusively a sub-digraph)  $S$  of  $W_q$  is **insertable in  $U_q$  up to  $l$  vertices**, if there exists a partition of  $U_q$  into three sets  $Z_1, Z_2$  and  $Z$  such that: there is no arc from  $Z_1$  to  $Z_2$ ,  $|Z| \leq l$  and there is no arc from  $Z_1$  to  $S$  and no arc from  $S$  to  $Z_2$ .

**Claim 9** *Every vertex  $x \in W_{i_0} \setminus R$  belonging to a 3-cycle of  $\mathcal{F}$  or a 2-remainder is insertable in  $U_{i_0}$  up to  $3^{p+1}$  vertices. Furthermore, every 3-cycle of  $\mathcal{F}$  contained in  $W_{i_0} \setminus R$  is insertable in  $U_{i_0}$  up to  $5 \cdot 3^{p+1}$  vertices.*

**Proof:** Consider  $C$  a 3-cycle of  $\mathcal{F}$  or a 2-remainder which is contained in  $W_{i_0} \setminus R$  and let  $x$  be a vertex of  $C$ . As  $U_{i_0}$  is an acyclic tournament by Claim 8, we denote by  $\{u_1, u_2, \dots, u_r\}$  its vertices in such way that  $U_{i_0}$  contains no arc  $u_i u_j$  with  $i < j$ . Among all the  $r+1$  cuts of type  $(Z_1 = \{u_1, \dots, u_i\}, Z_2 = \{u_{i+1}, \dots, u_r\})$ , we choose one for which  $d^+(x, Z_2) + d^+(Z_1, x)$  is minimum<sup>1</sup> and abusively denote it by  $(Z_1, Z_2)$  with  $Z_1 = \{u_1, \dots, u_i\}$ . If  $d^+(Z_1, x) = l$  then it is possible to build  $l$  3-cycles containing  $x$  and some vertices of  $Z_1$  which are all disjoint on  $Z_1$ . Indeed, we denote by  $(u_{in(j)})_{1 \leq j \leq l}$  (resp.  $(u_{out(j)})_{1 \leq j \leq l}$ ) the in-neighbours of  $x$  in  $Z_1$  (resp. the out-neighbours of  $x$  in  $Z_1$ ) sorted according to the order  $(u_i, u_{i-1}, \dots, u_1)$ . Then, assume that for some  $j$ ,  $x u_{out(j)} u_{in(j)}$  is not a 3-cycle (because  $u_{out(j)}$  is after  $u_{in(j)}$ , or because  $u_{out(j)}$  does not exist), it means that  $x$  has more in-neighbours than out-neighbours in the set  $\{u_i, u_{i-1}, \dots, u_{in(j)}\}$ , which contradicts the choice of the partition  $(Z_1, Z_2)$ . So, it is possible to form all the 3-cycles  $(x u_{out(j)} u_{in(j)})_{1 \leq j \leq l}$ . Similarly, the same statement holds with  $Z_2$ , and globally it is possible to provide  $d^+(x, Z_2) + d^+(Z_1, x)$  3-cycles containing  $x$  and all disjoint on  $U_{i_0}$ . Then, as  $x \in W_{i_0} \setminus R$  we have  $d^+(x, Z_2) + d^+(Z_1, x) \leq 3^{p+1}$  and hence  $x$  is insertable in  $U_{i_0}$  up to  $3^{p+1}$  vertices.

For the second part of the claim, consider a 3-cycle  $C = xyz$  which is contained in  $W_{i_0} \setminus R$ . By the first part of the claim, we know that there exist three sets of vertices  $Z_x, Z_y$  and  $Z_z$  in  $U_{i_0}$  of size at most  $3^{p+1}$  such that  $(U_{i_0} \setminus \{Z_x \cup Z_y \cup Z_z\}) - \{xy, yz, zx\}$  forms an acyclic digraph. We consider an acyclic ordering of this digraph. If one of three arcs  $xy, yz$  or  $zx$ , say  $xy$ , is backward in this ordering and 'jumps' across more than  $3^{p+1}$  vertices of  $U_{i_0}$ , then the arc  $xy$  is good and  $C$  should have been put in  $\mathcal{R}$ . So, as  $C$  can have one or two backward arcs with respect to this order, it is possible to remove from  $U_{i_0} \setminus \{Z_x \cup Z_y \cup Z_z\}$  two further sets of vertices of size at most  $3^{p+1}$  to insert  $C$ .  $\diamond$

Now, using Claim 9, we give to every vertex  $x$  of a 2-remainder which is contained in  $W_{i_0} \setminus R$  a position  $p(x)$  in the ordering of  $U_{i_0}$ . More precisely, there exists a set  $Z$  of at most  $3^{p+1}$  vertices of  $U_{i_0}$  such that there is no arc from  $\{u_1, \dots, u_{p(x)}\} \setminus Z$  to  $x$  and no arc from  $x$  to  $\{u_{p(x)+1}, \dots, u_r\} \setminus Z$ . If there is several possibilities to choose  $p(x)$ , we pick one arbitrarily. Similarly, for a 3-cycle  $C = xyz$  which lies in  $W_{i_0} \setminus R$ , we assign to each of its vertices a position  $p(x) = p(y) = p(z)$ , such that up to  $5 \cdot 3^{p+1}$  vertices,  $C$  is insertable between  $\{u_1, \dots, u_{p(x)}\}$  and  $\{u_{p(x)+1}, \dots, u_r\}$ .

<sup>1</sup>Here for two disjoint sets of vertices  $R, S$   $d^+(R, S)$  denotes the number of arcs from  $R$  to  $S$ .

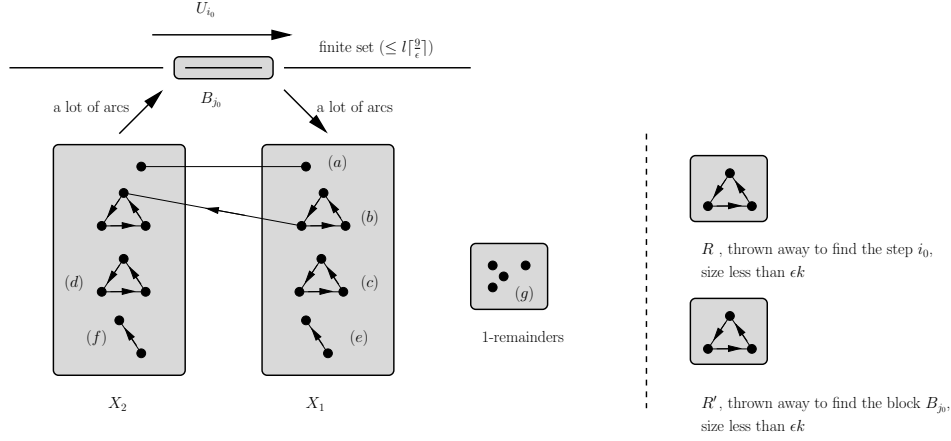


Figure 3: The situation in the proof of Theorem 1.3.

Then, we fix an integer  $l$  such that  $(\alpha - 1.5)l > 2 \cdot 3^{p+2}$  and  $l > 3 \cdot 3^{p+1}$  and we consider a partition of the first vertices of  $U_{i_0}$  into  $\lceil 9/\epsilon \rceil$  blocks of  $l$  vertices, provided that  $U_{i_0}$  is large enough. This is insured if  $U$ , of size at least  $(2\alpha - 3)k$ , is large enough, that is if  $(2\alpha - 3)k > l \lceil 9/\epsilon \rceil$ , what is possible as  $l$  and  $\epsilon$  only depend on  $\alpha$ . Exactly, for  $j = 1, \dots, \lceil 9/\epsilon \rceil$ , the block  $B_j$  is  $B_j = \{u_{(j-1)l+1}, u_{(j-1)l+2}, \dots, u_{jl}\}$ . As  $W_{i_0} \setminus R$  contains at most  $3k$  vertices, there is at most  $3k$  different values  $p(x)$  for  $x \in W_{i_0} \setminus R$ . So, one of the  $\lceil 9/\epsilon \rceil$  blocks  $B_j$ , say  $B_{j_0}$ , contains at most  $3k / \lceil 9/\epsilon \rceil \leq \epsilon k / 3$  values  $p(x)$  for  $x$  being a vertex of a 2-remainder or of a 3-cycle of  $W_{i_0} \setminus R$ . We call  $\mathcal{R}'$  these 2-remainders and 3-cycles of  $W_{i_0} \setminus R$  and denote by  $R'$  the set  $V(\mathcal{R}')$ . Remark that  $R'$  has size at most  $\epsilon k$ .

So, we partition the remaining vertices of  $W_{i_0} \setminus (R \cup R')$  into two parts:  $X_1 = \{x \in W_{i_0} : p(x) \leq (j_0 - 1)l\}$  and  $X_2 = \{x \in W_{i_0} : p(x) > j_0 l\}$ . By the definition of  $p$ , a 3-cycle  $C$  of  $\mathcal{F}$  which lies in  $W_{i_0} \setminus (R \cup R')$  satisfies  $V(C) \subseteq X_1$  or  $V(C) \subseteq X_1$ . Whereas the 2-remainders of  $W_{i_0} \setminus (R \cup R')$  can intersect both parts of the partition  $(X_1, X_2)$  of  $W_{i_0} \setminus (R \cup R')$ . The situation is depicted in Figure 3.

We have the following property on the partition  $((X_1, X_2)$  of  $W_{i_0} \setminus (R \cup R')$ .

**Claim 10** *Every arc from  $X_1$  to  $X_2$  is a good arc.*

**Proof:** Let  $xy$  be an arc from  $X_1$  to  $X_2$ . By definition, we know that  $p(x) \leq (j_0 - 1)l$  and that  $p(y) > j_0 l$ . That means that all the vertices of  $B_{j_0}$  dominate  $x$  except for at most  $3^{p+1}$  of them, and that all the vertices of  $B_{j_0}$  are dominated by  $y$  except for at most  $3^{p+1}$  of them. As  $|B_{j_0}| > 3 \cdot 3^{p+1}$ , we can find  $3^{p+1}$  vertices of  $B_{j_0}$  that are dominated by  $y$  and that dominate  $x$ , implying that  $xy$  is a good arc.  $\diamond$

Now, according to their behaviour, we classify the 2-remainders and 3-cycles which are in  $W_{i_0} \setminus (R \cup R')$ :

- A 2-remainders which have one vertex in  $X_1$  and the other in  $X_2$  is of type (a).
- We consider a maximal collection of disjoint pairs of 3-cycles  $\{C, C'\}$  where  $C$  is in  $X_1$ ,  $C'$  is in  $X_2$  and there is at least one arc from  $C$  to  $C'$ . All the 3-cycles involved in this collection are of type (b).
- A 3-remainder included in  $X_1$  is of type (c) if it is not of type (b).
- A 3-remainder included in  $X_2$  is of type (d) if it is not of type (b).

- A 2-remainder included in  $X_1$  is of type (e).
- A 2-remainder included in  $X_2$  is of type (f).

We abusively denote by  $a$  (resp.  $c$ ,  $d$ ,  $e$  and  $f$ ) the number of remainders of type (a) (resp. (c), (d), (e) and (f)). We denote by  $b$  the number of pairs of 3-remainders of type (b). Finally, we denote by  $g$  the number of 1-remainders. At this point of the proof, any of this value is an integer between 0 and  $k-1$  and  $X_1$  or  $X_2$  could be empty. This will be settled by the computation at the end of the proof. For the moment, we have the following properties.

**Claim 11** *We have the following bounds on the number of arcs from  $X_1$  to  $X_2$ :*

- The number of arcs from  $X_1$  to  $X_2$  linking a 3-cycle with another 3-cycle or a 2-remainder is at most 3.
- The number of arcs going from  $X_1$  to  $X_2$  and linking a vertex of an element of type (a) and a vertex of an element of type (b) is at most  $4ab$ .
- There is no arc from a 3-cycle of type (c) to a 3-cycle of type (d).

**Proof:** To prove the first point, consider  $C$  a 3-cycle of  $X_1$  and  $C'$  a 3-cycle or a 2-remainder of  $X_2$ . If there are more than three arcs from  $C$  to  $C'$ , then we find a 2-matching from  $C$  to  $C'$ . By Claim 10, this 2-matching is made of good arcs and the Rule 3.3 could apply to find a free vertex in  $C$ , contradicting that  $W_{i_0} \setminus R$  has no free elements.

For the second point, consider a 2-remainder of type (a) on vertices  $v_1$  and  $v_2$  with  $v_1 \in X_1$  and  $v_2 \in X_2$  and a pair  $(C_1, C_2)$  of 3-cycles of type (b) with  $C_1 = x_1y_1z_1$  contained in  $X_1$  and  $C_2 = x_2y_2z_2$  contained in  $X_2$  and  $x_1x_2$  being an arc of  $T$ . To find a contradiction, assume that the number of arcs from  $v_1$  to  $C_2$  plus the number of arcs from  $C_1$  to  $v_2$  is at least 5. It means that either  $v_1$  dominates  $C_2$  or  $C_1$  dominates  $v_2$ , say that  $v_1$  dominates  $C_2$ . Now,  $v_2$  is dominated by at least two vertices of  $C_1$ , and one of these two is not  $x_1$ , say that it is  $y_1$ . But, the arcs  $x_1x_2$ ,  $y_1v_2$  and  $v_1y_2$  are good by Claim 10, and  $z_1$  and  $z_2$  should be free by Rule 3.2, contradicting that  $W_{i_0} \setminus R$  has no free elements.

The third point follows from the definition of 3-cycles of type (c) and (d).  $\diamond$

Now, we can derive a number of in-equalities from the structure derived so far, in order to obtain a contradiction, knowing that it has not been possible to increase the size of  $\mathcal{F}$  above.

The first in-equality comes from the fact that there is a most  $k-1$  remainders and 3-cycles in  $W_{i_0}$ .

$$a + 2b + c + d + e + f + g < k \quad (2)$$

For the second one, we compute the number of arcs going outside of  $B_{j_0}$ , which has size  $l$ . There are at least  $\alpha kl - l(l-1)/2$  such arcs. The number of arcs from  $B_{j_0}$  to  $\cup_{j=1}^{j_0-1} B_j$  is at most  $l^2 \lceil 9/\epsilon \rceil$ . There is no arc from  $B_{j_0}$  to  $U_{i_0} \setminus (\cup_{j=1}^{j_0} B_j)$ . The number of arcs from  $B_{j_0}$  to  $X_2$  is at most  $|X_2|3^{p+1}$ , because every vertex of  $X_2$  is insertable into  $U_{i_0}$  before  $B_{j_0}$  up to  $3^{p+1}$  vertices. As  $X_2 \subseteq W$  we can bound this number by  $3k3^{p+1}$ . Finally, the remaining arcs going outside of  $B_{j_0}$  are at most  $l(a + 3b + 3c + 2e + g + |R| + |R'|)$ , and we obtain:

$$\alpha kl - \frac{l(l-1)}{2} \leq \lceil \frac{9}{\epsilon} \rceil l^2 + k3^{p+2} + l(a + 3b + 3c + 2e + g + 2\epsilon k),$$

which we rewrite as

$$(\alpha - 1.5)kl - \frac{l(l-1)}{2} - \lceil \frac{9}{\epsilon} \rceil l^2 - k3^{p+2} - 2\epsilon k \leq l(a + 3b + 3c + 2e + g - 1.5k)$$

And finally arrange in:

$$\left( \frac{(\alpha - 1.5)l}{2} - 3^{p+2} \right) k + \left( \left( \frac{(\alpha - 1.5)}{2} - 2\epsilon \right) k - \frac{l-1}{2} - \lceil \frac{9}{\epsilon} \rceil l \right) l \leq l(a + 3b + 3c + 2e + g - 1.5k)$$

By choice of  $l$ , the first term is positive, and as  $\epsilon < (\alpha - 1.5)/4$ , if  $k$  is large enough, the second term is strictly positive too, implying that:

$$a + 3(b + c) + 2e + g > 1.5k \quad (3)$$

For the last in-equality, we compute the number of arcs going outside of  $X_1$ . As previously, we first show that, if  $k$  is large enough, we have  $\alpha k|X_1| - d^+(X_1, U_{i_0}) - d^+(X_1, R \cup R') - 1.5k|X_1|$  is positive. Indeed, this term is greater than  $(\alpha - 1.5)k|X_1| - \lceil \frac{9}{\epsilon} \rceil l|X_1| - 2\epsilon k|X_1|$  which is

$$|X_1| \left( ((\alpha - 1.5) - 2\epsilon)k - \lceil \frac{9}{\epsilon} \rceil l \right)$$

and this is positive if  $k$  is large enough. Now, we have to take in account the arcs inside  $X_1$  and those from  $X_1$  to  $X_2$  and to the 1-remainders. By the calculation above we still have at least  $1.5k|X_1|$  arcs incident with vertices of  $X_1$  to account for. Using Claim 11 we obtain

$$\begin{aligned} & 1.5k(a + 3(b + c) + 2e) - \frac{1}{2}(a + 3(b + c) + 2e)^2 < \\ & a(a + 4b + 3d + 2f + g) + b(3b + 3d + 3f + 3g) + c(3a + 3b + 3f + 3g) + e(2a + 3b + 3d + 4f + 2g) = \\ & a(a + 4b + 3c + 3d + 2e + 2f + g) + b(3b + 3c + 3d + 3e + 3f + 3g) + c(3f + 3g) + e(3d + 4f + 2g) \end{aligned} \quad (4)$$

Considering (2), (3) and (4), an equation solver leads to a contradiction. We just indicate how to manage the computation 'by hand'. Suppose that there exists a solution  $X = (a, b, c, d, e, f, g)$  to these three in-equalities, we will show that then  $X' = (a + b + f + g, 0, b + c + d + e, 0, 0, 0, 0)$  is also a solution to these equations. It is easy to check that  $X'$  is a solution to (2) and (3). For (4), we denote by  $\phi(a, b, c, d, e, f, g)$  the value

$$\begin{aligned} & 2a(a + 4b + 3c + 3d + 2e + 2f + g) + 2b(3b + 3c + 3d + 3e + 3f + 3g) + 2c(3f + 3g) + 2e(3d + 4f + 2g) \\ & + (a + 3(b + c) + 2e)^2 - 3k(a + 3(b + c) + 2e) \end{aligned}$$

Then, we compute  $\phi(a + b + f + g, 0, b + c + d + e, 0, 0, 0, 0) - \phi(a, b, c, d, e, f, g)$  and obtain:

$$\begin{aligned} & 2a(2b + 3d + 2e + f + 2g) + 3b(3b + 2c + 8d + 4e + 4f + 4g) + 6c(3d + e + f + g) + 3d(3d + 4e + 4f + 4g) \\ & + e(5e + 4f + 8g) + 3f(f + 2g) + 3g^2 - 3k(b + 3d + e + f + g) \end{aligned}$$

Using the fact that  $X$  is a solution of (3), we have  $-3k > -2(a + 3(b + c) + 2e + g)$  and so  $\phi(a + b + f + g, 0, b + c + d + e, 0, 0, 0, 0) - \phi(a, b, c, d, e, f, g)$  is greater than

$$2a(b + e + g) + b(3b + 6d + 2e + 6f + 4g) + 3d(3d + 4f + 2g) + e(e + 2g) + f(3f + 4g) + g^2$$

which is positive. So,  $\phi(a + b + f + g, 0, b + c + d + e, 0, 0, 0, 0)$  is strictly positive and  $X'$  is a solution of (4).

Now, there is a solution to the in-equations (2), (3) and (4) of type  $(a', 0, c', 0, 0, 0, 0)$ , what is impossible: (2) gives  $a' + c' < 1$  and (4) gives  $3(a' + 3c')(a' + c' - 1) > 0$ .

This concludes the proof of Theorem 1.3. As a last remark, note that  $k_\alpha$  is larger than a polynomial function in  $l$ , which is larger than an exponential in  $p$ , itself larger than a linear function in the inverse of  $\alpha - 1.5$ . So,  $k_\alpha$  is an exponential function in the inverse of  $\alpha - 1.5$ .

## 4 Some Remarks

It is perhaps worth pointing out that the following obvious idea does not lead to a proof of Conjecture 1.1 for tournaments: find a 3-cycle  $C$  which is not dominated by any vertex of  $V(T) - V(C)$ , remove  $C$  and apply induction. This approach does not work because of the following<sup>2</sup>.

**Proposition 4.1** *For infinitely many  $k \geq 3$  there exists a tournament  $T$  with  $\delta^+(T) = 2k - 1$  such that every 3-cycle  $C$  is dominated by at least one vertex of minimum out-degree.*

**Proof:** Consider the quadratic residue tournament  $T$  on 11 vertices  $V(T) = \{1, 2, \dots, 11\}$  and arcs  $A(T) = \{i \rightarrow i + p \pmod{11} : i \in V(T), p \in \{1, 3, 4, 5, 9\}\}$ . The possible types of 3-cycles in  $T$  are  $i \rightarrow i + 1 \rightarrow i + 10 \rightarrow i$ ,  $i \rightarrow i + 1 \rightarrow i + 6 \rightarrow i$ ,  $i \rightarrow i + 3 \rightarrow i + 6 \rightarrow i$ ,  $i \rightarrow i + 3 \rightarrow i + 7 \rightarrow i$ , where indices are taken modulo 11. These are dominated by the vertices  $i - 3, i - 3, i + 2, i + 2$  respectively. By substituting an arbitrary tournament for each vertex of  $T$ , we can obtain a tournament with the property that every 3-cycle is dominated by some vertex of minimum out-degree.  $\diamond$

On the other hand, removing a 2-cycle from a digraph  $D$  with  $\delta^+(D) \geq 2k - 1$  clearly results in a new digraph  $D'$  with  $\delta^+(D') \geq 2(k - 1) - 1$  and hence, when trying to prove Conjecture 1.1, we may always assume that the digraph in question has no 2-cycles. In particular the following follows directly from Theorem 1.2<sup>3</sup>.

**Corollary 4.2** *Every semicomplete digraph  $D$  with  $\delta^+(D) \geq 2k - 1$  contains  $k$  disjoint cycles.*  $\diamond$

A **chordal bipartite digraph** is a bipartite digraph with no induced cycle of length greater than 4. Note that in particular semicomplete bipartite digraphs [3, page 35] are chordal bipartite. It is easy to see that Conjecture 1.1 holds for chordal bipartite digraphs.

**Proposition 4.3** *Every chordal bipartite digraph  $D$  with  $\delta^+(D) \geq 2k - 1$  contains  $k$  disjoint cycles.*

**Proof:** This follows from the fact that such a digraph contains a directed cycle  $C$  of length 2 or 4 as long as  $k \geq 1$ . As  $D$  is bipartite, no vertex dominates more than half of the vertices on  $C$  and so we have  $\delta^+(D - C) \geq 2(k - 1) - 1$  and the result follows by induction on  $k$ .  $\diamond$

An **extension** of a digraph  $D = (V, A)$  is any digraph which can be obtained by substituting an independent set  $I_v$  for each vertex  $v \in V$ . More precisely we replace each vertex  $v$  of  $V$  by an independent set  $I_v$  and then add all arcs from  $I_u$  to  $I_v$  precisely if  $uv \in A$ .

**Proposition 4.4** *Let  $D = T[I_{n_1}, I_{n_2}, \dots, I_{n_{|V(T)|}}]$  be an extension of a tournament  $T$  such that  $I_{n_i}$  is an independent set on  $n_i$  vertices for  $i \in \{1, 2, \dots, |V(T)|\}$ . If  $\delta^+(D) \geq 2k - 1$ , then  $D$  contains  $k$  disjoint 3-cycles.*

**Proof:** Let  $T'$  be the tournament that we obtain from  $D$  by replacing each  $I_{n_i}$  by a transitive tournament on  $n_i$  vertices. Then  $\delta^+(T') \geq 2k - 1$  and hence, by theorem 1.2,  $T'$  contains  $k$  disjoint 3-cycles  $C_1, C_2, \dots, C_k$ . By the definition of an extension and the fact that we replaced independent sets by acyclic digraphs, no  $C_i$  can contain more than one vertex from any  $I_{n_i}$ , implying that  $C_1, C_2, \dots, C_k$  are also cycles in  $D$ .  $\diamond$

## References

- [1] N. Alon. Disjoint directed cycles. *J. Combin. Theory Ser. B*, 68(2):167–178, 1996.

<sup>2</sup>See also [2, Section 9.1].

<sup>3</sup>A digraph is semicomplete if there is at least one arc between any pair of vertices.

- [2] N. Alon and J.H. Spencer. *The Probabilistic Method*. John Wiley & Sons, New York, 2nd edition, 2000.
- [3] J. Bang-Jensen and G. Gutin. *Digraphs: Theory, Algorithms and Applications, 2nd Edition*. Springer-Verlag, London, 2009.
- [4] J.-C. Bermond and C. Thomassen. Cycles in digraphs—a survey. *J. Graph Theory*, 5(1):1–43, 1981.
- [5] S. Bessy, N. Lichiardopol, and J.-S. Sereni. Two proofs of the bermond-thomassen conjecture for tournaments with bounded minimum in-degree. *Discrete Mathematics*, 310:557–560, 2010.
- [6] N. Lichiardopol. Vertex-disjoint directed cycles of prescribed length in tournaments with given minimum out-degree and in-degree. *Discrete Mathematics*, 310:2567–2570, 2010.
- [7] N. Lichiardopol, A. Pór, and J.-S. Sereni. A step towards the Bermond-Thomassen conjecture about disjoint cycles in digraphs. *Siam J. Discrete Math.*, 23:979–992, 2009.
- [8] C. Thomassen. Disjoint cycles in digraphs. *Combinatorica*, 3(3–4):393–396, 1983.



## 2.4 Arc-chromatic number for digraphs

Arc-chromatic number of digraphs in which every vertex has  
bounded outdegree or bounded indegree

S. Bessy and F. Havet,  
Projet Mascotte, CNRS/INRIA/UNSA,  
INRIA Sophia-Antipolis,  
2004 route des Lucioles BP 93,  
06902 Sophia-Antipolis Cedex,  
France  
`sbessy,fhavet@sophia.inria.fr`  
and  
E. Birmelé  
Laboratoire MAGE, Université de Haute-Alsace,  
6, rue des frères Lumière,  
68093 Mulhouse Cedex,  
France  
`Etienne.Birmele@uha.fr`  
November 8, 2004

### Abstract

A  $k$ -digraph is a digraph in which every vertex has outdegree at most  $k$ . A  $(k \vee l)$ -digraph is a digraph in which a vertex has either outdegree at most  $k$  or indegree at most  $l$ . Motivated by function theory, we study the maximum value  $\Phi(k)$  (resp.  $\Phi^\vee(k, l)$ ) of the arc-chromatic number over the  $k$ -digraphs (resp.  $(k \vee l)$ -digraphs). El-Sahili [3] showed that  $\Phi^\vee(k, k) \leq 2k + 1$ . After giving a simple proof of this result, we show some better bounds. We show  $\max\{\log(2k+3), \theta(k+1)\} \leq \Phi(k) \leq \theta(2k)$  and  $\max\{\log(2k+2l+4), \theta(k+1), \theta(l+1)\} \leq \Phi^\vee(k, l) \leq \theta(2k+2l)$  where  $\theta$  is the function defined by  $\theta(k) = \min\{s : \binom{s}{\lceil \frac{s}{2} \rceil} \geq k\}$ . We then study in more details properties of  $\Phi$  and  $\Phi^\vee$ . Finally, we give the exact values of  $\Phi(k)$  and  $\Phi^\vee(k, l)$  for  $l \leq k \leq 3$ .

## 1 Introduction

A *directed graph* or *digraph*  $D$  is a pair  $(V(D), E(D))$  of disjoint sets (of *vertices* and *arcs*) together with two maps  $tail : E(D) \rightarrow V(D)$  and  $head : E(D) \rightarrow V(D)$  assigning to every arc  $e$  a *tail*,  $tail(e)$ , and a *head*,  $head(e)$ . The tail and the head of an arc are its *ends*. An arc with tail  $u$  and head  $v$  is denoted by  $uv$ ; we say that  $u$  *dominates*  $v$  and write  $u \rightarrow v$ . We also say that  $u$  and  $v$  are adjacent. The *order* of a digraph is its number of vertices. In this paper, all the digraph we consider are *loopless*, that is that every arc has its tail distinct from its head.

Let  $D$  be a digraph. The *line-digraph* of  $D$  is the digraph  $L(D)$  such that  $V(L(D)) = E(D)$  and an arc  $a \in E(D)$  dominates an arc  $b \in E(D)$  in  $L(D)$  if and only if  $head(a) = tail(b)$ .

A *vertex-colouring* or *colouring* of  $D$  is an application  $c$  from the vertex-set  $V(D)$  into a set of colours  $S$  such that for any arc  $uv$ ,  $c(u) \neq c(v)$ . The *chromatic number* of  $D$ , denoted  $\chi(D)$ , is the minimum number of colours of a colouring of  $D$ .

An *arc-colouring* of  $D$  is an application  $c$  from the arc-set  $E(D)$  into a set of colours  $S$  such that if the tail of an arc  $e$  is the head of an arc  $e'$  then  $c(e) \neq c(e')$ . Trivially, there is a one-to-one correspondence between arc-colourings of  $D$  and colourings of  $L(D)$ . The *arc-chromatic number* of  $D$ , denoted  $\chi_a(D)$ , is the minimum number of colours of an arc-colouring of  $D$ . Clearly  $\chi_a(D) = \chi(L(D))$ .

A *k-digraph* is a digraph in which every vertex has outdegree at most  $k$ . A  $(k \vee l)$ -*digraph* is a digraph in which a vertex has either outdegree at most  $k$  or indegree at most  $l$ .

For any digraph  $D$  and set of vertices  $V' \subset V(D)$ , we denote by  $D[V']$ , the subdigraph of  $D$  induced by the vertices of  $V'$ . For any subdigraph  $F$  of  $D$ , we denote by  $D - F$  the digraph  $D[(V(D) \setminus V(F))]$ . For any arc-set  $E' \subset E$ , we denote by  $D - E'$  the digraph  $(V(D), E(D) \setminus E')$  and for any vertex  $x \in V(D)$ , we denote by  $D - x$  the digraph induced by  $V(D) \setminus \{x\}$ .

Let  $D$  be a  $(k \vee l)$ -digraph. We denote by  $V^+(D)$ , or  $V^+$  if  $D$  is clearly understood, the subset of the vertices of  $D$  with outdegree at most  $k$ , and by  $V^-(D)$  or  $V^-$  the complementary of  $V^+(D)$  in  $V(D)$ . Also  $D^+$  (resp.  $D^-$ ) denotes  $D[V^+]$  (resp.  $D[V^-]$ ).

In this paper, we study the arc-chromatic number of  $k$ -digraphs and  $(k \vee l)$ -digraphs. This is motivated by the following interpretation in function theory as shown by El-Sahili in [3].

Let  $f$  and  $g$  be two maps from a finite set  $A$  into a set  $B$ . Suppose that  $f$  and  $g$  are *nowhere coinciding*, that is for all  $a \in A$ ,  $f(a) \neq g(a)$ . A subset  $A'$  of  $A$  is  $(f, g)$ -*independant* if  $f(A') \cap g(A') = \emptyset$ . We are interested by finding the largest  $(f, g)$ -independant subset of  $A$  and the minimum number of  $(f, g)$ -independant subsets to partition  $A$ . As shown by El-Sahili [3], this can be translated into an arc-colouring problem.

Let  $D_{f,g}$  and  $H_{f,g}$  be the digraphs defined as follows :

- $V(D_{f,g}) = B$  and  $(b, b') \in E(D_{f,g})$  if there exists an element  $a$  in  $A$  such that  $g(a) = b$  and  $f(a) = b'$ . Note that if for all  $a$ ,  $f(a) \neq g(a)$ , then  $D_{f,g}$  has no loop.
- $V(H_{f,g}) = A$  and  $(a, a') \in E(H_{f,g})$  if  $f(a) = g(a')$ .

We associate to each arc  $(b, b')$  in  $D_{f,g}$  the vertex  $a$  of  $A$  such that  $g(a) = b$  and  $f(a) = b'$ . Then  $(a, a')$  is an arc in  $H_{f,g}$  if, and only if,  $head(a) = tail(a')$  (as arcs in  $D_{f,g}$ ). Thus  $H_{f,g} = L(D_{f,g})$ . Note that for every digraph  $D$ , there exists maps  $f$  and  $g$  such that  $D = D_{f,g}$ .

It is easy to see that an  $(f, g)$ -independant subset of  $A$  is an independant set in  $H_{f,g}$ . In [2] El-Sahili proved the following :

**Theorem 1 (El-Sahili [2])** *Let  $f$  and  $g$  be two nowhere coinciding maps from a finite set  $A$  into a set  $B$ . Then there exists an  $(f, g)$ -independant subset  $A'$  of cardinality at least  $|A|/4$ .*

Let  $f$  and  $g$  be two nowhere coinciding maps from a finite set  $A$  into  $B$ . We define  $\phi(f, g)$  as the minimum number of  $(f, g)$ -independant sets to partition  $A$ . Then  $\phi(f, g) = \chi(H_{f,g}) = \chi_a(D_{f,g})$ .

Let  $\Phi(k)$  (resp.  $\Phi^\vee(k, l)$ ) be the maximum value of  $\phi(f, g)$  for two nowhere coinciding maps  $f$  and  $g$  from  $A$  into  $B$  such that for every  $z$  in  $B$ ,  $g^{-1}(z) \leq k$  (resp. either  $g^{-1}(z) \leq k$  or

$f^{-1}(z) \leq l$ ). The condition  $f^{-1}(z)$  (resp.  $g^{-1}(z)$ ) has at most  $k$  elements means that each vertex has indegree (resp. outdegree) at most  $k$  in  $D_{f,g}$ . Hence  $\Phi(k)$  (resp.  $\Phi^\vee(k, l)$ ) is the maximum value of  $\chi_a(D)$  for  $D$  a  $k$ -digraph (resp.  $(k \vee l)$ -digraph).

**Remark 2** Let  $f$  and  $g$  be two nowhere coinciding maps from  $A$  into  $B$ . Then  $A$  may be partitionned into  $\Phi(|A| - 1)$   $(f, g)$ -independant sets.

The functions  $\Phi^\vee$  and  $\Phi$  are very close to each other:

**Proposition 3**

$$\Phi(k) \leq \Phi^\vee(k, 0) \leq \dots \leq \Phi^\vee(k, k) \leq \Phi(k) + 2$$

**Proof.** The sole inequality that does not immediatly follow the definitions is  $\Phi^\vee(k, k) \leq \Phi(k) + 2$ . Let us prove it.

Let  $D$  be a  $(k \vee k)$ -digraph. One can colour the arcs in  $D^+ \cup D^-$  with  $\Phi(k)$  colours. It remains to colour the arcs with tail in  $V^-$  and head in  $V^+$  with one new colour and the arcs with tail in  $V^+$  and head in  $V^-$  with a second new colour.  $\square$

Moreover, we conjecture that  $\Phi^\vee(k, k)$  is never equal to  $\Phi(k) + 2$ .

**Conjecture 4**

$$\Phi^\vee(k, k) \leq \Phi(k) + 1$$

In [3], El-Sahili gave the following upper bound on  $\Phi^\vee(k, k)$ :

**Theorem 5 (El-Sahili [3])**  $\Phi^\vee(k, k) \leq 2k + 1$

In this paper, we first give simple proofs of Theorems 1 and 5. Then, in Section 3, we improve the upper bounds on  $\Phi(k)$  and  $\Phi^\vee(k, l)$ . We show (Theorem 18) that  $\Phi(k) \leq \theta(2k)$  if  $k \geq 2$ , and  $\Phi^\vee(k, l) \leq \theta(2k + 2l)$  if  $k + l \geq 3$ , where  $\theta$  is the function defined by  $\theta(k) = \min\{s : \binom{s}{\lceil s/2 \rceil} \geq k\}$ . Since  $2^s/s \leq \binom{s}{\lceil s/2 \rceil} \leq 2^s/\sqrt{s}$  for  $s \geq 2$ , once can obtain the following equivalent for  $\theta$  as  $k \rightarrow \infty$ :

$$\theta(k) = \log(k) + \Theta(\log(\log(k)))$$

Lower bounds for  $\Phi$  and  $\Phi^\vee$  are stated by Corollaries 14 and 15:  $\max\{\log(2k+3), \theta(k+1)\} \leq \Phi(k)$  and  $\max\{\log(2k+2l+4), \theta(k+1), \theta(l+1)\} \leq \Phi^\vee(k, l)$ .

We also establish (Corollary 21) that  $\Phi^\vee(k, l) \leq \theta(2k)$  if  $\theta(2k) \geq 2l + 1$ .

In Section 4, we study in more details the relations between  $\Phi^\vee(k, l)$  and  $\Phi(k)$ . We conjecture that if  $k$  is very large compared to  $l$  then  $\Phi^\vee(k, l) = \Phi(k)$ . We prove that  $\Phi^\vee(k, 0) = \Phi(k)$  and conjecture that  $\Phi^\vee(k, 1) = \Phi(k)$  if  $k \geq 1$ . We prove that for a fixed  $k$  either this latter conjecture holds or Conjecture 4 holds. This implies that  $\Phi^\vee(k, 1) \leq \Phi(k) + 1$ .

Finally, in Section 5, we give the exact values of  $\Phi(k)$  and  $\Phi^\vee(k, l)$  for  $l \leq k \leq 3$ . They are summarized in the following table :

$\Phi^\vee(0, 0) = 1$	$\Phi^\vee(1, 0) = \Phi(1) = 3$	$\Phi^\vee(2, 0) = \Phi(2) = 4$	$\Phi^\vee(3, 0) = \Phi(3) = 4$
	$\Phi^\vee(1, 1) = 3$	$\Phi^\vee(2, 1) = 4$	$\Phi^\vee(3, 1) = 4$
		$\Phi^\vee(2, 2) = 4$	$\Phi^\vee(3, 2) = 5$
			$\Phi^\vee(3, 3) = 5$

## 2 Simple proofs of Theorems 1 and 5

**Proof of Theorem 1.** Let  $D = D_{f,g}$ . Let  $(V_1, V_2)$  be a partition of  $V(D)$  that maximizes the number of arcs  $a$  with one end in  $V_1$  and one end in  $V_2$ . It is well-known that  $a \geq |E(D)|/2$ . Now let  $A_1$  be the set of arcs with head in  $V_1$  and tail in  $V_2$  and  $A_2$  be the set of arcs with head in  $V_2$  and tail in  $V_1$ . Then  $A_1$  and  $A_2$  corresponds to independant sets of  $L(D)$  and  $|A_1| + |A_2| = a$ . Hence one of the  $A_i$  has cardinality at least  $a/2 = \frac{|E(D)|}{4}$ .  $\square$

Before giving a short proof of Theorem 5, we precise few standard definitions.

**Definition 6** A *path* is a non-empty digraph  $P$  of the form

$$V(P) = \{v_0, v_1, \dots, v_k\} \quad E(P) = \{v_0 v_1, v_1 v_2, \dots, v_{k-1} v_k\},$$

where the  $v_i$  are all distinct. The vertices  $v_0$  and  $v_k$  are respectively called the *origin* and *terminus* of  $P$ .

A *circuit* is a non-empty digraph  $C$  of the form

$$V(C) = \{v_0, v_1, \dots, v_k\} \quad E(C) = \{v_0 v_1, v_1 v_2, \dots, v_{k-1} v_k, v_k v_0\},$$

where the  $v_i$  are all distinct.

A digraph is *strongly connected* or *strong* if for every two vertices  $u$  and  $v$  there is a path with origin  $u$  and terminus  $v$ . A maximal strong subdigraph of a digraph  $D$  is called a *strong component* of  $D$ . A component  $I$  of  $D$  is *initial* if there is no arc with tail in  $V(D) \setminus V(I)$  and head in  $V(I)$ . A component  $I$  of  $D$  is *terminal* if there is no arc with tail in  $V(I)$  and head in  $V(D) \setminus V(I)$ . A digraph is *connected* if its underlying graph is connected.

A digraph  $D$  is  *$l$ -degenerate* if every subdigraph  $H$  has a vertex of degree at most  $l$ .

The following lemma corresponding to the greedy colouring algorithm is a piece of folklore.

**Lemma 7** Every  $l$ -degenerate digraph is  $(l+1)$ -colourable.

**Proof of Theorem 5.** Let  $D$  be a  $(k \vee k)$ -digraph. According to Lemma 7, it suffices to prove that  $L(D)$  is  $2k$ -degenerate.

In every initial strong component  $C$  there is a vertex with indegree at most  $k$ . Indeed if there is no such vertex then  $(k+1)|C| \leq \sum_{v \in C} d^-(v) \leq \sum_{v \in C} d^+(v) \leq k|C|$ . Analogously, in every terminal strong component there is a vertex with outdegree at most  $k$ .

Now, there is a path originating in a minimal component and terminating in a terminal one. Hence there is a path whose origin has indegree at most  $k$  and whose terminus has outdegree at most  $k$ . Hence there is an arc  $e$  whose tail has indegree at most  $k$  and whose head has outdegree at most  $k$ . Thus  $e$  has degree at most  $2k$  in  $L(D)$ .  $\square$

## 3 Lower and upper bounds for $\Phi$ and $\Phi^\vee$

We will now search for bounds on  $\Phi$  since they also give bounds on  $\Phi^\vee$ .

Theorem 1 and an easy induction yields  $\chi_a(D) \leq \log_{4/3} |D|$ . However there exists better upper bounds stated by Poljak and Rödl [5]. For sake of completeness and in order to introduce useful tools, we provide a proof of Theorem 11.

**Definition 8** We denote by  $\overline{H}_k$  the complementary of the hypercube of dimension  $k$ , that is the digraph with vertex-set all the subsets of  $\{1, \dots, k\}$  and with arc-set  $\{xy : x \not\subseteq y\}$ .

A homomorphism  $h : D \rightarrow D'$  is a mapping  $h : V(D) \rightarrow V(D')$  such that for every arc  $xy$  of  $D$ ,  $h(x)h(y)$  is an arc of  $D'$ .

Let  $c$  be an arc-colouring of a digraph  $D$  into a set of colours  $S$ . For any vertex  $x$  of  $D$ , we denote by  $Col_c^+(x)$  or simply  $Col^+(x)$  the set of colours assigned to the arcs with tail  $x$ . We define  $Col^-(x) = S \setminus Col^+(x)$ . Note that  $Col^-(x)$  contains (but may be bigger than) the set of colours assigned to the arcs with head  $x$ . The cardinality of  $Col^+(x)$  (resp.  $Col^-(x)$ ) is denoted by  $col^+(x)$  (resp.  $col^-(x)$ ).

**Theorem 9** For every digraph  $D$ ,  $\chi_a(D) = \min\{k : D \rightarrow \overline{H}_k\}$ .

**Proof.** Assume that  $D$  admits an arc-colouring with  $\{1, \dots, k\}$ . It is easy to check that  $Col^+$  is a homomorphism from  $D$  to  $\overline{H}_k$ .

Conversely, suppose that there exists a homomorphism  $h$  from  $D$  to  $\overline{H}_k$ . Assign to each arc  $xy$  an element of  $h(y) \setminus h(x)$ , which is not empty. This provides an arc-colouring of  $D$ .  $\square$

**Definition 10** The complete digraph of order  $n$ , denoted  $\vec{K}_n$ , is the digraph with vertex-set  $\{v_1, v_2, \dots, v_n\}$  and arc-set  $\{v_i v_j : i \neq j\}$ .

The transitive tournament of order  $n$ , denoted  $TT_n$ , is the digraph with vertex-set  $\{v_1, v_2, \dots, v_n\}$  and arc-set  $\{v_i v_j : i < j\}$ .

The following corollary of Theorem 9 provides bounds on the arc-chromatic number of a digraph according to its chromatic number.

**Theorem 11 (Poljak and Rödl [5])** For every digraph  $D$ ,

$$\lceil \log(\chi(D)) \rceil \leq \chi_a(D) \leq \theta(\chi(D)).$$

**Proof.** By definition of the chromatic number,  $D \rightarrow \vec{K}_{\chi(D)}$ . As the subsets of  $\{1, \dots, k\}$  with cardinality  $\lceil \frac{k}{2} \rceil$  induce a complete digraph on  $\binom{k}{\lceil \frac{k}{2} \rceil}$  vertices in  $\overline{H}_k$ , we obtain a homomorphism from  $D$  to  $\overline{H}_{\theta(\chi(D))}$ . So  $\chi_a(D) \leq \theta(\chi(D))$ .

By Theorem 9, we have  $D \rightarrow \overline{H}_{\chi_a(D)}$ . As  $\chi(\overline{H}_{\chi_a(D)}) = 2^{\chi_a(D)}$ , we obtain  $D \rightarrow \vec{K}_{2^{\chi_a(D)}}$ .  $\square$

These bounds are tight since the lower one is achieved by transitive tournaments and the upper one by complete digraphs by Sperner's Lemma (see [6]). However, the lower bound may be increased if the digraph has no sink (vertex with outdegree 0) or/and no source (vertex with indegree 0).

**Theorem 12** Let  $D$  be a digraph.

(i) If  $D$  has no sink then  $\log(\chi(D) + 1) \leq \chi_a(D)$ .

(ii) If  $D$  has no source and no sink then  $\log(\chi(D) + 2) \leq \chi_a(D)$ .

**Proof.** The proof is identical to the proof of Theorem 11. But if a digraph has no source (resp. no sink) then for every  $v$ ,  $Col^+(v) \neq S$  (resp.  $Col^+(v) \neq \emptyset$ ).  $\square$

Again, these two lower bounds are also tight. Let  $Q_n$  (resp.  $W_n$ ) be the tournament of order  $n$  obtained from  $TT_n$  by reversing the arc  $v_1v_n$  (resp.  $v_2v_n$ ). One can easily check that  $\chi_a(W_n) = \lceil \log(n+1) \rceil = \lceil \log(\chi(W_n) + 1) \rceil$  and  $\chi_a(Q_n) = \lceil \log(n+2) \rceil = \lceil \log(\chi(Q_n) + 2) \rceil$ .

**Proposition 13** *Every  $k$ -digraph is  $2k$ -degenerate.*

**Proof.** Every subdigraph of a  $k$ -digraph is also a  $k$ -digraph. Hence it suffices to prove that every  $k$ -digraph has a vertex with degree at most  $2k$ . Since the sum of outdegrees equals the sum of indegrees, there is a vertex with indegree at most  $k$  and thus with degree at most  $2k$ .  $\square$

**Corollary 14**

$$\max\{\log(2k+3), \theta(k+1)\} \leq \Phi(k) \leq \theta(2k+1)$$

**Proof.** The upper bound follows from Proposition 13, Lemma 7 and Theorem 11. The lower bound comes from a regular tournament on  $2k+1$  vertices  $T_{2k+1}$  and the complete digraph on  $k+1$  vertices  $\vec{K}_{k+1}$ . Indeed  $\chi(T_{2k+1}) = 2k+1$ , so  $\chi_a(T_{2k+1}) \geq \log(2k+3)$  by Theorem 12 and  $\chi_a(\vec{K}_{k+1}) = \theta(k+1)$ .  $\square$

**Corollary 15**

$$\max\{\log(2k+2l+4), \theta(k+1), \theta(l+1)\} \leq \Phi^\vee(k, l) \leq \theta(2k+2l+2)$$

**Proof.** The upper bound follows Proposition 13 and Theorem 11 since every  $(k \vee l)$ -digraph  $D$  is  $2k+2l+2$ -colourable ( $D^+$  is  $2k$ -degenerate and so  $(2k+1)$ -colourable and  $D^-$  is  $2l$ -degenerate and so  $(2l+1)$ -colourable). The lower bound comes from  $\vec{K}_{k+1}$ ,  $\vec{K}_{l+1}$  and a tournament  $T$  composed of a regular tournament on  $2l+1$  vertices dominating a regular tournament on  $2k+1$  vertices. Indeed,  $\chi_a(\vec{K}_{k+1}) = \theta(k+1)$ ,  $\chi_a(\vec{K}_{l+1}) = \theta(l+1)$  and  $T$  has no source, no sink and chromatic number  $2k+2l+2$ , so, by Theorem 12,  $\chi_a(T) \geq \log(2k+2l+4)$ .  $\square$

We can obtain a slightly better upper bound on  $\Phi$ . Bounds on  $\Phi^\vee$  follow.

**Definition 16** For any integer  $k \geq 1$ , let  $T_k^+$  ( $k \geq 1$ ) be the complete digraph on  $t_1^+, \dots, t_{2k+1}^+$  minus the arcs  $\{t_1^+t_2^+, t_1^+t_3^+, \dots, t_1^+t_{k+1}^+\}$ .

**Lemma 17** *Let  $k \geq 1$  be an integer. If  $D$  is a  $k$ -digraph, then there exists a homomorphism  $h^+$  from  $D$  to  $T_k^+$  such that if  $h^+(x) = t_1^+$  then  $d^+(x) = k$ .*

**Proof.** For a fixed  $k$ , we prove it by induction on  $|V(D)|$ .

First, suppose that there exists in  $D$  a vertex  $x$  with  $d^-(x) < k$ . Then,  $d^+(x) + d^-(x) < 2k$ . By induction on  $D - x$ , there is a homomorphism  $h$  from  $D - x$  to  $T_k^+$  such that if  $h^+(v) = t_1^+$  then  $d_{D-x}^+(v) = k$ . Hence,  $h^+(y) \neq t_1^+$  for every inneighbour  $y$  of  $x$ , because  $d_{D-x}^+(y) < k$ . As  $x$  has at most  $2k-1$  neighbours, we find  $i \in \{2, \dots, 2k+1\}$  such that no neighbour  $y$  of  $x$  satisfies  $h^+(y) = t_i^+$ . So,  $h^+(x) = t_i^+$  extends  $h^+$  to a homomorphism from  $D$  to  $T_k^+$ .

Suppose now that every vertex  $v$  of  $D$  has indegree at least  $k$ . Since  $\sum_{v \in V(D)} d^-(v) = \sum_{v \in V(D)} d^+(v) \leq k|V(D)|$ , every vertex has indegree and outdegree  $k$ . Hence, by Brooks Theorem (see [1]) either  $D$  is  $2k$ -colourable and  $D \rightarrow T_k^+[\{t_2^+, \dots, t_{2k+1}^+\}]$ , or  $D$  is a regular tournament on  $2k+1$  vertices. In this latter case, label the vertices of  $D$  with  $v_1, v_2, \dots, v_{2k+1}$  such that  $N^-(v_1) = \{v_2, \dots, v_{k+1}\}$ . Then  $h^+$  defined by  $h^+(v_i) = t_i^+$  is the desired homomorphism.  $\square$

**Theorem 18** *Let  $k$  and  $l$  be two positive integers.*

(i) *If  $k \geq 2$ , then  $\Phi(k) \leq \theta(2k)$ .*

(ii) *If  $k+l \geq 3$ , then  $\Phi^\vee(k, l) \leq \theta(2k+2l)$ .*

**Proof.** (i) If  $k = 2$ , the result follows Corollary 14 since  $\theta(4) = \theta(5) = 4$ . Suppose now that  $k \geq 3$ . Let  $D$  be a  $k$ -digraph. By Lemma 17 there is a homomorphism from  $D$  to  $T_k^+$ . We will provide a homomorphism  $g$  from  $T_k^+$  to  $\overline{H}_{\theta(2k)}$ .

Fix  $S_1, \dots, S_{2k}$ ,  $2k$  subsets of  $\{1, \dots, \theta(2k)\}$  with cardinality  $\lfloor \theta(2k)/2 \rfloor$  and  $S$  a subset of  $\{1, \dots, 2k\}$  with cardinality  $\lfloor \theta(2k)/2 \rfloor - 1$ . Without loss of generality, the  $S_i$  containing  $S$  are  $S_1, \dots, S_l$  with  $l \leq \lceil \theta(2k)/2 \rceil + 1 \leq k$ . (One can easily check that  $\theta(2k)/2 + 1 \leq k$  provided that  $k \geq 3$ .) Now, set  $g(t_1^+) = S$  and  $g(t_{i+1}^+) = S_i$  for  $1 \leq i \leq 2k$ . It is straightforward to check that  $g$  is a homomorphism.

(ii) Let  $D$  be a  $(k \vee l)$ -digraph. By Lemma 17, there exists a homomorphism  $h^+$  from  $D^+$  to  $T_k^+$  such that if  $h^+(x) = t_1^+$  then  $d^+(x) = k$  and, by symmetry, a homomorphism  $h^-$  from  $D^-$  to  $T_l^-$ , the complete digraph on  $\{t_1^-, \dots, t_{2l+1}^-\}$  minus the arcs  $\{t_2^- t_1^-, t_3^- t_1^-, \dots, t_{l+1}^- t_1^-\}$ , such that if  $h^-(x) = t_1^-$  then  $d^-(x) = l$ . We now provide a homomorphism from  $D$  to  $\overline{H}_{\theta(2k+2l)}$ .

Fix  $S^+$  and  $S^-$ , two subsets of  $\{1, \dots, \theta(2k+2l)\}$  with cardinality  $\lfloor \theta(2k+2l)/2 \rfloor - 1$  for  $S^+$  and  $\lfloor \theta(2k+2l)/2 \rfloor + 1$  for  $S^-$  such that  $S^+ \not\subseteq S^-$ . (This is possible since  $\theta(2k+2l) \geq 4$ , because  $k+l \geq 2$ .) Set  $\mathcal{N} = \{U \subset \{1, \dots, \theta(2k+2l)\} : |U| = \lfloor \theta(2k+2l)/2 \rfloor\}$ . We want a partition of  $\mathcal{N}$  into two parts  $\mathcal{A}$  and  $\mathcal{B}$  with  $|\mathcal{A}| \geq 2k$  and  $|\mathcal{B}| \geq 2l$ , such that  $S^+$  is included in at most  $k$  sets of  $\mathcal{A}$  and  $S^-$  contains at most  $l$  sets of  $\mathcal{B}$ . Let  $\mathcal{N}_{S^+}$  (resp.  $\mathcal{N}_{S^-}$ ) be the set of elements of  $\mathcal{N}$  containing  $S^+$  (resp. contained in  $S^-$ ). We have  $|\mathcal{N}_{S^+}| = \lfloor \theta(2k+2l)/2 \rfloor + 1$  and  $|\mathcal{N}_{S^-}| = \lfloor \theta(2k+2l)/2 \rfloor + 1$ ; because  $k+l \geq 3$ , it follows  $|\mathcal{N}_{S^+}| \leq k+l$  and  $|\mathcal{N}_{S^-}| \leq k+l$ . Moreover, the sets  $\mathcal{N}_{S^+}$  and  $\mathcal{N}_{S^-}$  are disjoint. Let us sort the elements of  $\mathcal{N}$  beginning with those of  $\mathcal{N}_{S^-}$  and ending with those of  $\mathcal{N}_{S^+}$ . Let  $\mathcal{A}$  be the  $2k$  first sets in this sorting and  $\mathcal{B}$  what remains ( $|\mathcal{B}| \geq 2l$ ). We claim that  $\mathcal{A}$  contains at most  $k$  elements of  $\mathcal{N}_{S^+}$ . If not, then  $|\mathcal{A}| > \binom{\theta(2k+2l)}{\lfloor \theta(2k+2l)/2 \rfloor} - |\mathcal{N}_{S^+}| + k$ . We obtain  $2k > 2k + 2l - |\mathcal{N}_{S^+}| + k$  which contradicts  $|\mathcal{N}_{S^+}| \leq k+l$ . With same argument,  $\mathcal{B}$  contains at most  $l$  elements of  $\mathcal{N}_{S^-}$ .

Finally, set  $\mathcal{A} = \{A_1, \dots, A_{2k}\}$  such that none of  $A_{k+1}, \dots, A_{2k}$  contains  $S^+$  and  $\{B_1, \dots, B_{2l}\}$   $2l$  sets of  $\mathcal{B}$  such that none of  $B_{l+1}, \dots, B_{2l}$  is contained in  $S^-$ .

Let us define  $h : D \rightarrow \overline{H}_{2k+2l}$ . If  $x \in V^+$  and  $h^+(x) = t_i^+$  then  $h(x) = S^+$  if  $i = 1$  and  $h(x) = A_{i-1}$  otherwise. If  $x \in V^-$  and  $h^-(x) = t_i^-$  then  $h(x) = S^-$  if  $i = 1$  and  $h(x) = B_{i-1}$  otherwise. Let us check that  $h$  is a homomorphism. Let  $xy$  be an arc of  $D$ .  $T_k^+$  is a subdigraph of  $\overline{H}_{2k+2l}[\{A_1, \dots, A_{2k}, S^+\}]$  and  $T_l^-$  is a subdigraph of  $\overline{H}_{2k+2l}[\{B_1, \dots, B_{2l}, S^-\}]$ . So,  $h(x)h(y)$  is an arc of  $\overline{H}_{2k+2l}$  if  $x$  and  $y$  are both in  $V^+$  or both in  $V^-$ . Suppose now that  $x \in V^+$  and  $y \in V^-$ , then  $h^+(x) \neq t_1^+$  because  $d_{D^+}(x) < k$  and  $h(x) \neq S^+$ . Similarly,  $h^-(y) \neq S^-$ . Thus  $h(x)$  and  $h(y)$  are elements of  $\mathcal{N}$ , so  $h(x)h(y) \in E(\overline{H}_{2k+2l})$ . Finally, suppose that  $x \in V^-$  and

$y \in V^+$ . Then  $h(x)h(y) \in E(\overline{H}_{2k+2l})$  because no element of  $\{B_1, \dots, B_{2l}, S^-\}$  is a subset of an element of  $\{A_1, \dots, A_{2k}, S^+\}$ .  $\square$

**Remark 19** Note that the homomorphism provided in (i) has for image subsets of  $\{1, \dots, \theta(2k)\}$  with cardinality at most  $\lfloor \theta(2k)/2 \rfloor$ . So, using the method developped in Theorem 9, we provide an arc-colouring of a  $k$ -digraph  $D$  with  $\theta(2k)$  colours which satisfies  $col^+(x) \leq \lfloor \theta(2k)/2 \rfloor$ , so  $col^-(x) \geq \lceil \theta(2k)/2 \rceil$ , for every vertex  $x$  of  $D$ .

We will now improve the bound (ii) of Theorem 18 when  $l$  is very small compared to  $k$ .

**Lemma 20** *Let  $D$  be a  $(k \vee l)$ -digraph and  $D^1$  the subdigraph of  $D$  induced by the arcs with tail in  $V^+$ . If there exists an arc-colouring of  $D^1$  with  $m \geq 2l + 1$  colours such that for every  $x$  in  $V^+$ ,  $col^-(x) \geq l + 1$  then  $\chi_a(D) = m$ .*

**Proof.** We will extend the colouring as stated into an arc-colouring of  $D$ .

First, we extend this colouring to the arcs of  $D^-$ . Since  $\sum_{v \in V^-} d_{D^-}^+(v) = \sum_{v \in V^-} d_{D^-}^-(v) \leq l|V^-|$ , there is a vertex  $v_1 \in V^-$  such that  $d_{D^-}^+(v_1) \leq l$ . And so on, by induction, there is an ordering  $(v_1, v_2, \dots, v_p)$  of the vertices of  $D^-$  such that, for every  $1 \leq i \leq p$ ,  $v_i$  dominates at most  $l$  vertices in  $\{v_j : j > i\}$ . Let us colour the arcs of  $D^-$  in decreasing order of their head; that is first colour the arcs with head  $v_p$  then those with head  $v_{p-1}$ , and so on. This is possible since at each stage  $i$ , an arc  $uv_i$  has at most  $2l < m$  forbidden colours ( $l$  ingoing  $u$  and  $l$  outgoing  $v_i$  to a vertex in  $\{v_j : j > i\}$ ).

It remains to colour the arcs with tail in  $V^-$  and head in  $V^+$ . Let  $v^-v^+$  be such an arc. Since  $col^-(v^+) \geq l + 1$  and  $d^-(v^-) \leq l$ , there is a colour  $\alpha$  in  $Col^-(v^+)$  that is assigned to no arc ingoing  $v^-$ . Hence, assigning  $\alpha$  to  $v^-v^+$ , we extend the arc-colouring to  $v^-v^+$ .  $\square$

**Corollary 21** *If  $\theta(2k) \geq 2l + 1$ , then  $\phi^\vee(k, l) \leq \theta(2k)$ .*

**Proof.** The digraph  $D^1$ , as defined in Lemma 20, is a  $k$ -digraph. The result follows directly from Remark 19 and Lemma 20.  $\square$

## 4 Relations between $\Phi(k)$ and $\Phi^\vee(k, l)$

**Conjecture 22** *Let  $l$  be a positive integer. There exists an integer  $k_l$  such that if  $k \geq k_l$  then  $\Phi^\vee(k, l) = \Phi(k)$ .*

We now prove Conjecture 22, for  $l = 0$ , showing that  $k_0 = 1$ .

**Theorem 23** *If  $k \geq 1$ ,*

$$\Phi^\vee(k, 0) = \Phi(k).$$

**Proof.** Let  $D = (V, A)$  be a  $(k \vee 0)$  digraph. Let  $V_0$  be the set of vertices with indegree 0. Let  $D'$  be the digraph obtained from  $D$  by splitting each vertex  $v$  of  $V_0$  into  $d^+(v)$  vertices with outdegree 1. Formally, for each vertex  $v \in V_0$  incident to the arcs  $vw_1, \dots, vw_{d^+(v)}$ , replace  $v$



by  $\{v_1, v_2, \dots, v_{d^+(v)}\}$  and  $vw_i$  by  $v_i w_i$ ,  $1 \leq i \leq d^+(v)$ . By construction,  $D'$  is a  $k$ -digraph and  $L(D) = L(D')$ . So  $\chi_a(D) = \chi_a(D') \leq \Phi(k)$ .  $\square$

We conjecture that if  $l = 1$ , Conjecture 22 holds with  $k_1 = 1$ .

**Conjecture 24** *If  $k \geq 1$ ,*

$$\Phi^\vee(k, 1) = \Phi(k)$$

**Theorem 25** *If  $\Phi(k) = \Phi(k-1)$  or  $\Phi(k) = \Phi(k+1)$  then  $\Phi^\vee(k, 1) = \Phi(k)$ .*

**Proof.** By Lemma 20, it suffices to prove that if  $\Phi(k) = \Phi(k-1)$  or  $\Phi(k) = \Phi(k+1)$  every  $k$ -digraph admits an arc-colouring with  $\Phi(k)$  colours such that for every vertex  $x$ ,  $\text{col}^-(x) \geq 2$ .

Suppose that  $\Phi(k) = \Phi(k-1)$ . Let  $D$  be a  $k$ -digraph and  $D'$  be a  $(k-1)$ -digraph such that  $\chi_a(D') = \Phi(k)$ . Let  $C$  be the digraph constructed as follows: for every vertex  $x \in V(D)$  add a copy  $D'(x)$  of  $D'$  such that every vertex of  $D'(x)$  dominates  $x$ . Then  $C$  is a  $k$ -digraph, so it admits an arc-colouring  $c$  with  $\Phi(k)$  colours. Note that  $c$  is also an arc-colouring of  $D$  which is a subdigraph of  $C$ . Let us prove that for every vertex  $x \in V(D)$ ,  $\text{col}^-(x) \geq 2$ .

Suppose, reductio ad absurdum, that there is a vertex  $x \in V(D)$  such that  $\text{col}^-(x) \leq 1$ . Since there are arcs ingoing  $x$  in  $C$  (those from  $V(D'(x))$ ), then  $\text{Col}^-(x)$  is a singleton  $\{\alpha\}$ . Now every arc  $vx$  with  $v \in D'(x)$  is coloured  $\alpha$  so any arc  $uv \in E(D'(x))$  is not coloured  $\alpha$ . Hence  $c$  is an arc-colouring with  $\Phi(k) - 1$  colours which is a contradiction.

The proof is analogous if  $\Phi(k) = \Phi(k+1)$  with  $D'$  a  $k$ -digraph such that  $\chi_a(D') = \Phi(k)$ . Then  $C$  is a  $(k+1)$ -digraph and we get the result in the same way.  $\square$

The next theorem shows that for a fixed integer  $k$ , one of the Conjectures 24 and 4 holds.

**Theorem 26** *Let  $k$  be an integer. Then  $\Phi^\vee(k, 1) = \Phi(k)$  or  $\Phi^\vee(k, k) \leq \Phi(k) + 1$ .*

**Proof.** Suppose that  $\Phi^\vee(k, 1) \neq \Phi(k)$ . Let  $C$  be a  $(k \vee 1)$ -digraph such that  $\chi_a(C) = \Phi(k) + 1$  and  $C^1$  the subdigraph of  $C$  induced by the arcs with tail in  $V^+(C)$ . By Lemma 20, for every arc-colouring of  $C^1$  with  $\Phi(k)$  colours there exists a vertex  $x$  of  $C^+$  with  $\text{col}^-(x) \leq 1$ .

Let  $D$  be a  $(k \vee k)$ -digraph. Let  $D^1$  (resp.  $D^2$ ) be the subdigraph of  $D$  induced by the arcs with tail in  $V^+(D)$  (resp. head in  $V^-(D)$ ). Let  $E'$  be the set of arcs of  $D$  with tail in  $V^-$  and head in  $V^+$ . Let  $F^1$  be the digraph constructed from  $C^1$  as follows : for every vertex  $x \in V^+(C)$ , add a copy  $D^+(x)$  of  $D^+$  and the arcs  $\{u(x)x : uv \in E(D), u \in V^+(D), \text{ and } v \in V^-(D)\}$ . Then  $F^1$  is a  $k$ -digraph so it admits an arc-colouring  $c_1$  with  $\{1, \dots, \Phi(k)\}$ . Now there is a vertex  $x \in V^+(C)$  such that  $\text{col}^-(x) \leq 1$ . So all the arcs from  $D^+(x)$  to  $x$  are coloured the same. Free to permute the labels, we may assume they are coloured 1. Since  $F^1[V(D^+(x)) \cup x]$  has the same line-digraph than  $D^1$ , the arc-colourings of  $F^1[V(D^+(x)) \cup x]$  is in one-to-one correspondence with the arc-colourings of  $D^1$ . So  $D^1$  admits an arc-colouring  $c^1$  with  $\{1, \dots, \Phi(k)\}$  such that every arc with head in  $V^-$  is coloured 1.

Analogously,  $D^2$  admits an arc-colouring  $c^2$  with  $\{1, \dots, \Phi(k)\}$  such that every arc with head in  $V^-$  is coloured 1. The union of  $c_1$  and  $c_2$  is an arc-colouring of  $D - E'$  with  $\{1, \dots, \Phi(k)\}$ . Hence assigning  $\Phi(k) + 1$  to every arc of  $E'$ , we obtain an arc-colouring of  $D$  with  $\Phi(k) + 1$  colours.  $\square$

**Corollary 27**  $\Phi^\vee(k, 1) \leq \Phi(k) + 1$ .

Note that since  $\Phi(k)$  is bounded by  $\theta(2k)$ , the condition  $\Phi^\vee(k, 1) = \Phi(k)$  or  $\Phi^\vee(k, k) \leq \Phi(k) + 1$  is very often true. Indeed, we conjecture that it is always true and that  $\Phi$  behaves “smoothly”.

**Conjecture 28** (i) If  $k \geq 1$ ,  $\Phi(k + 1) \leq \Phi(k) + 1$ .

(ii) If  $k \geq 1$ ,  $\Phi(k + 2) \leq \Phi(k) + 1$ .

(iii)  $\Phi(k_1 k_2) \leq \Phi(k_1) + \Phi(k_2)$ .

Note that (ii) implies (i) and Conjecture 24.

The arc-set of a  $(k_1 + k_2)$ -digraph  $D$  may trivially be partitioned into two sets  $E_1$  and  $E_2$  such that  $(V(D), E_1)$  is a  $k_1$ -digraph and  $(V(D), E_2)$  is a  $k_2$ -digraph. So  $\Phi(k_1 + k_2) \leq \Phi(k_1) + \Phi(k_2)$ . In particular,  $\Phi(k + 1) \leq \Phi(k) + \Phi(1) = \Phi(k) + 3$ . Despite we were not able to prove Conjecture 28-(i), we now improve the above trivial result.

**Theorem 29** If  $k \geq 1$  then,  $\Phi(k + 1) \leq \Phi(k) + 2$ .

**Proof.** Let  $D$  be a  $(k + 1)$ -digraph. Free to add arcs, we may assume that  $d^+(v) = k + 1$  for every  $v \in V(D)$ . Let  $T_1, \dots, T_p$  be the terminal components of  $D$ . Each  $T_i$  contains a circuit  $C_i$  which has a chord. Indeed consider a maximal path  $P$  in  $T_i$  and two arcs with tail its terminus and head in  $P$ , by maximality. One can extend  $\bigcup C_i$  into a subdigraph  $F$  spanning  $D$  such that  $d_F^+(v) \geq 1$  for every  $v \in V(D)$  and the sole circuits are the  $C_i$ ,  $1 \leq i \leq p$ . In fact,  $F$  is the union of  $p$  connected components  $F_1, \dots, F_p$ , each  $F_i$  being the union of  $C_i$  and inarborescences  $A_i^1, \dots, A_i^{q_i}$  with roots  $r_i^1, \dots, r_i^{q_i}$  in  $C_i$  such that  $(V(C_i), V(A_i^1) \setminus \{r_i^1\}, \dots, V(A_i^p) \setminus \{r_i^p\})$  is a partition of  $V(D)$ .

Now  $D - F$  is a  $k$ -digraph. So we colour the arcs of  $D - F$  with  $\Phi(k)$  colours. Let  $\alpha$  and  $\beta$  be two new colours. Let us colour the arcs of  $F$ . Let  $1 \leq i \leq p$ . If  $C_i$  is an even circuit then  $F_i$  is bipartite and its arcs may be coloured by  $\alpha$  and  $\beta$ . If  $C_i$  is an odd circuit, consider its chord  $xy$  in  $E(D - F)$ . In the colouring of  $D - F$ ,  $Col^+(x) \not\subseteq Col^+(y)$  thus there is an arc  $x'y'$  of  $E(C_i)$  such that  $Col^+(x') \not\subseteq Col^+(y')$ . Hence we may assign to  $x'y'$  a colour of  $Col^+(x') \setminus Col^+(y')$ . Now  $F_i - x'y'$  is bipartite and its arcs may be coloured by  $\alpha$  and  $\beta$ .  $\square$

## 5 $\Phi$ and $\Phi^\vee$ for small value of $k$ or $l$ .

### 5.1 $\Phi(1)$ , $\Phi^\vee(1, 0)$ and $\Phi^\vee(1, 1)$ .

**Theorem 30**

$$\Phi^\vee(1, 1) = \Phi^\vee(1, 0) = \Phi(1) = 3$$

**Proof.** By Theorem 5,  $\Phi^\vee(1, 1) \leq 3$ . The 3-circuit is its own line-digraph and is not 2-colourable.  $\square$

### 5.2 $\Phi(2)$ and $\Phi^\vee(2, l)$ , for $l \leq 2$ .

The aim of this subsection is to prove Theorem 35, that is  $\Phi(2) = \Phi^\vee(2, 0) = \Phi^\vee(2, 1) = \Phi^\vee(2, 2) = 4$ . Therefore, we first exhibit a 2-digraph which is not 3-arc-colourable. Then we show that  $\Phi^\vee(2, 2) \leq 4$ .

**Definition 31** For any integer  $k \geq 1$ , the *rotative tournament* on  $2k+1$  vertices, denoted  $R_{2k+1}$ , is the tournament with vertex-set  $\{v_1, \dots, v_{2k+1}\}$  and arc-set  $\{v_i v_j : j - i \pmod{2k+1} \in \{1, \dots, k\}\}$ .

**Proposition 32** *The tournament  $R_5$  is not 3-arc-colourable. So  $\Phi(2) \geq 4$ .*

**Proof.** Suppose that  $R_5$  admits a 3-arc-colouring  $c$  in  $\{1, 2, 3\}$ . Then, for any two vertices  $x$  and  $y$ ,  $Col^+(x) \neq Col^+(y)$  and  $1 \leq col^+(x) \leq 2$ . Hence there is a vertex, say  $v_1$ , such that  $col^+(v_1) = 1$ , say  $Col^+(v_1) = \{1\}$ . Then  $Col^+(v_2)$  and  $Col^+(v_3)$  are subsets of  $\{2, 3\}$  and  $Col^+(v_2) \not\subset Col^+(v_3)$ . It follows that  $col^+(v_3) = 1$ . Repeating the argument for  $v_3$ , we obtain  $col^+(v_5) = 1$  and then  $col^+(v_i) = 1$ , for every  $1 \leq i \leq 5$ , which is a contradiction.  $\square$

In order to prove that  $\Phi^\vee(2, 2) \leq 4$ , we need to show that every  $(2 \vee 2)$ -digraph admits homomorphism  $h$  into  $\overline{H}_4$ . In order to exhibit such a homomorphism, we first show that there is a homomorphism  $h^+$  from  $D^+$  into a subdigraph  $S_2^+$  of  $\overline{H}_4$  and a homomorphism  $h^-$  from  $D^-$  into a subdigraph  $S_2^-$  of  $\overline{H}_4$  with specific properties allowing us to extend  $h^+$  and  $h^-$  into a homomorphism  $h$  from  $D$  into  $\overline{H}_4$ .

**Definition 33** Let  $S_2^+$  be the digraph with vertex-set  $\{s_1^+, \dots, s_6^+\}$  with arc-set  $\{s_i^+ s_j^+ : i \neq j\} \setminus \{s_2^+ s_1^+, s_4^+ s_3^+, s_6^+ s_5^+\}$ .

**Lemma 34** *Let  $D$  be a 2-digraph. There exists a homomorphism  $h^+$  from  $D$  to  $S_2^+$  such that the vertices  $x$  with  $h^+(x) \in \{s_2^+, s_4^+, s_6^+\}$  have outdegree 2.*

**Proof.** Let us prove it by induction on  $|V(D)|$ . If  $d^+(x) \leq 1$  for every vertex  $x$  of  $D$  then  $D$  is 3-colourable and  $D \rightarrow S_2^+[\{s_1^+, s_3^+, s_5^+\}]$ . So, we assume that there exists a vertex  $x$  with outdegree 2. By induction hypothesis, there is a homomorphism  $h^+ : D - x \rightarrow S_2^+$  with the required condition. Note that every inneighbour of  $x$  has outdegree at most 1 in  $D - x$  and thus can not have image  $s_2^+, s_4^+$  or  $s_6^+$  by  $h^+$ . Denote by  $y$  and  $z$  the outneighbours of  $x$ . The set  $\{h^+(y), h^+(z)\}$  does not intersect one set of  $\{s_1^+, s_2^+\}$ ,  $\{s_3^+, s_4^+\}$  and  $\{s_5^+, s_6^+\}$ , say  $\{s_1^+, s_2^+\}$ . Then, setting  $h^+(x) = s_2^+$ , we extend  $h^+$  into a homomorphism from  $D$  to  $S_2^+$  with the required condition.  $\square$

**Theorem 35**

$$\Phi(2) = \Phi^\vee(2, 0) = \Phi^\vee(2, 1) = \Phi^\vee(2, 2) = 4$$

**Proof.**

By Proposition 32,  $4 \leq \Phi(2) \leq \Phi^\vee(2, 0) \leq \Phi^\vee(2, 1) \leq \Phi^\vee(2, 2)$ .

Let us prove that  $\Phi^\vee(2, 2) \leq 4$ . Let  $D$  be a  $(2 \vee 2)$ -digraph. We will provide a homomorphism from  $D$  to  $\overline{H}_4$ .

Let  $S_2^-$  be the dual of  $S_2^+$ , that is the digraph on  $\{s_1^-, \dots, s_6^-\}$  with arc-set  $\{s_i^- s_j^- : i \neq j\} \setminus \{s_1^- s_2^-, s_3^- s_4^-, s_5^- s_6^-\}$ . By Lemma 34, there is a homomorphism  $h^+ : D^+ \rightarrow S_2^+$  such that if  $h^+(x) \in \{s_2^+, s_4^+, s_6^+\}$  then  $d_{D^+}^+(x) = 2$ . Symmetrically, there exists a homomorphism  $h^- : D^- \rightarrow S_2^-$  such that if  $h^-(x) \in \{s_2^-, s_4^-, s_6^-\}$  then  $d_{D^-}^-(x) = 2$ .

Let  $S_2$  be the digraph obtained from the disjoint union of  $S_2^+$  and  $S_2^-$  by adding the arcs of  $\{s_i^- s_j^+ : 1 \leq i \leq 6, 1 \leq j \leq 6\} \cup \{s_i^+ s_j^- : i = 1, 3, 5, j = 1, 3, 5\}$ . The mapping  $h : D \rightarrow S_2$  defined by  $h(x) = h^+(x)$  if  $x \in V^+$  and  $h(x) = h^-(x)$  if  $x \in V^-$  is a homomorphism. Indeed if  $xy$  is an arc of  $D$  with  $x \in V^+$  and  $y \in V^-$ , conditions on  $h^+$  and  $h^-$  imply that  $h(x) = h^+(x) \in \{s_1^+, s_3^+, s_5^+\}$  and  $h(y) = h^-(y) \in \{s_1^-, s_3^-, s_5^-\}$ . To conclude, Figure 1 provides a homomorphism  $g$  from  $S_2$  to  $\overline{H}_4$ . The non-oriented arcs on the figure corresponds to circuits of length 2 and all the arcs from  $S_2^-$  to  $S_2^+$  are not represented.  $\square$

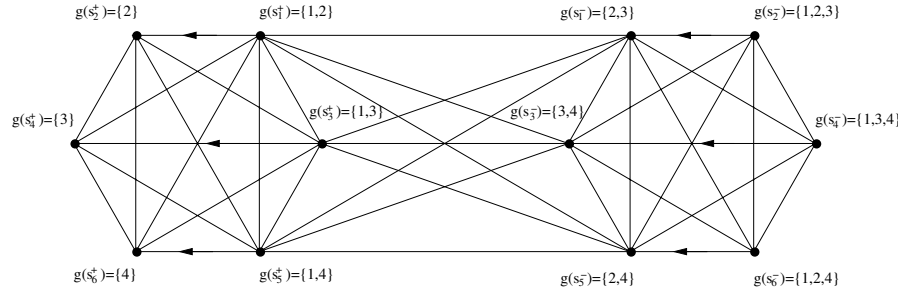


Figure 1: The homomorphism  $g$  from  $S_2$  to  $\overline{H}_4$ .

### 5.3 $\Phi(3)$ and $\Phi^\vee(3, l)$ , for $l \leq 3$ .

#### Theorem 36

$$\Phi(3) = \Phi^\vee(3, 0) = \Phi^\vee(3, 1) = 4$$

**Proof.**  $4 \leq \Phi(2) \leq \Phi(3) \leq \Phi^\vee(3, 0) \leq \Phi^\vee(3, 1) \leq \theta(6) = 4$  by Corollary 21.  $\square$

In the remaining of this subsection, we shall prove Theorem 42, that is  $\Phi^\vee(3, 2) = \Phi^\vee(3, 3) = 5$ . Therefore, we first exhibit a  $(3 \vee 2)$ -digraph which is not 4-arc-colourable. Then we show that  $\Phi^\vee(3, 3) \leq 5$ .

**Definition 37** Let  $G^-$  be the digraph obtained from the rotative tournament on five vertices  $R_5$ , with vertex set  $\{v_1^-, \dots, v_5^-\}$  and arc-set  $\{v_i^- v_j^- : j - i \pmod{5} \in \{1, 2\}\}$  and five copies of the 3-circuits  $R_3^1, \dots, R_3^5$  by adding, for  $1 \leq i \leq 5$ , the arcs  $vv_i^-$ , for  $v \in R_3^i$ .

Let  $G^+$  be the digraph obtained from the rotative tournament on seven vertices  $R_7$ , with vertex set  $\{v_1^+, \dots, v_7^+\}$  and arc-set  $\{v_i^+ v_j^+ : j - i \pmod{7} \in \{1, 2, 3\}\}$  and seven copies of the rotative tournament of,  $R_5^1, \dots, R_5^7$  by adding, for  $1 \leq i \leq 7$ , the arcs  $vv_i^+$ , for  $v \in R_5^i$ .

Finally, let  $G$  be the  $(3 \vee 2)$ -digraph obtained from the disjoint union of  $G^-$  and  $G^+$  by adding all the arcs of the form  $v^- v^+$  with  $v^- \in V(G^-)$  and  $v^+ \in V(G^+)$ . See Figure 2.

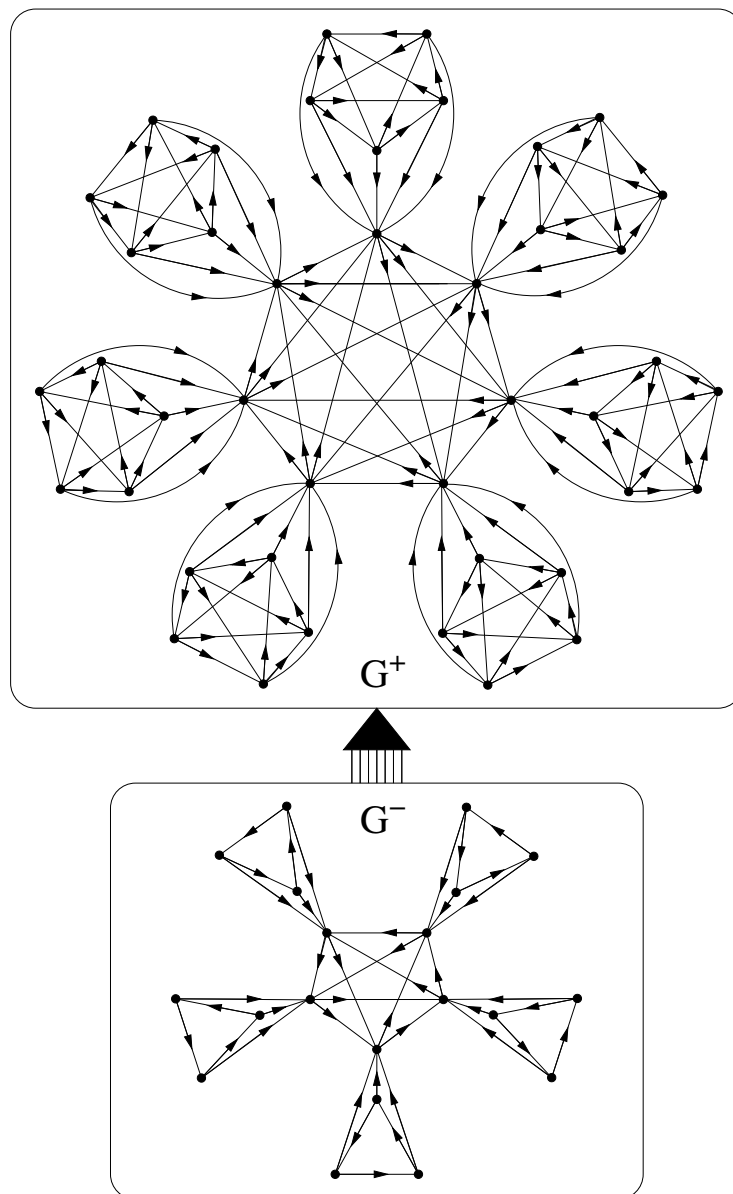


Figure 2: The non 4-arc-colourable  $(3 \vee 2)$ -digraph  $G$ .

**Proposition 38** *The digraph  $G$  is not 4-arc-colourable. So  $\Phi(3, 2) \geq 5$ .*

**Proof.** Suppose for a contradiction that  $G$  admits an arc-colouring  $c$  in  $\{1, 2, 3, 4\}$ .

Let  $v^+$  be a vertex of  $G^+$  and  $v^-$  a vertex of  $G^-$ . Then since  $v^-v^+$  is an arc,  $Col^+(v^+) \neq Col^+(v^-)$ . We will show:

- (1) there are at least two 2-subsets  $S$  of  $\{1, 2, 3, 4\}$  such that a vertex  $v^- \in G^-$  satisfies  $Col^+(v^-) = S$ ;
- (2) there are at least five 2-subsets  $S$  of  $\{1, 2, 3, 4\}$  such that a vertex  $v^+ \in G^+$  satisfies  $Col^+(v^+) = S$ .

This gives a contradiction since there are only six 2-subsets in  $\{1, 2, 3, 4\}$ .

Let us first show (1). Every vertex of  $G^-$  satisfies  $col^+ \geq 2$  otherwise all the arcs of  $R_7$  in  $G^+$  must be coloured with three colours, a contradiction to Theorem 12. Hence, since in  $R_5$  all the  $Col^+$  are distinct and not  $\{1, 2, 3, 4\}$ , a vertex of  $R_5$ , say  $v_1^-$ , has  $col^+ = 2$ , say  $Col^+(v_1^-) = \{1, 2\}$ . Consider now the vertices of  $R_3^1$ . None of them has  $Col^+ = \{1, 2, 3\}$  nor  $Col^+ = \{1, 2, 4\}$  since they are dominated by  $v_1^-$ . Moreover they all have different  $Col^+$  since  $R_3^1$  is a tournament. Hence one of them, say  $v$ , satisfies  $col^+(v) = 2$ . Now  $Col^+(v) \neq Col^+(v_1^-)$  since  $v_1^- \rightarrow v$ .

Let us now prove (2). Let  $\mathcal{S} = \{2\text{-subsets } S \text{ such that } \exists v^+ \in G^+, Col^+(v^+) = S\}$  and suppose that  $|\mathcal{S}| \leq 4$ . Every vertex of  $G^+$  has  $col^+ \leq 2$  otherwise all the arcs of  $R_5$  in  $G^-$  must be coloured with three colours, a contradiction to Proposition 32. Now, all the vertices of  $R_7$  have distinct and non-empty  $Col^+$ . So at least three vertices of  $R_7$  have  $col^+ = 2$  and  $|\mathcal{S}| \geq 3$ . Thus, without loss of generality, we are in one these three following cases:

- (a)  $\mathcal{S} \subset \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{1, 4\}\}$  and  $Col^+(v_1^+) = \{1, 2\}$ ;
- (b)  $\mathcal{S} \subset \{\{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 4\}\}$  and  $Col^+(v_1^+) = \{2, 3\}$ ;
- (c)  $\mathcal{S} \subset \{\{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 4\}\}$  and  $Col^+(v_1^+) = \{1, 4\}$ .

Let  $x_1, \dots, x_5$  be the vertices of  $R_5^1$  such that  $x_i x_j$  is an arc if and only if  $j = i + 1 \pmod 5$  or  $j = i + 2 \pmod 5$  and  $\mathcal{F} = \{Col^+(x_i) : 1 \leq i \leq 5\}$ . Recall that  $|\mathcal{F}| = 5$  since  $R_5^1$  is a tournament and that every element  $S$  of  $\mathcal{F}$  is not included in  $Col^+(v_1^+)$  since  $x_i \rightarrow v_1^+$  for every  $1 \leq i \leq 5$ .

Case (a): We have  $\mathcal{F} = \{\{3\}, \{4\}, \{2, 3\}, \{3, 4\}, \{1, 4\}\}$ . So we may assume that  $Col^+(x_1) = \{3\}$ . Now because  $x_1 \rightarrow x_2$ ,  $x_1 \rightarrow x_3$  and  $x_2 \rightarrow x_3$ ,  $Col^+(x_1) \not\subset Col^+(x_2)$ ,  $Col^+(x_1) \not\subset Col^+(x_3)$  and  $Col^+(x_2) \not\subset Col^+(x_3)$ . It follows that  $Col^+(x_2) = \{1, 4\}$  and  $Col^+(x_3) = \{4\}$ . Hence, none of  $Col^+(x_4)$  and  $Col^+(x_5)$  is  $\{3, 4\}$  since  $x_3 \rightarrow x_4$  and  $x_3 \rightarrow x_5$ , a contradiction.

Case (b): We have  $\mathcal{F} = \{\{1\}, \{4\}, \{1, 2\}, \{1, 3\}, \{1, 4\}\}$ . So we may assume that  $Col^+(x_1) = \{1\}$ . Since  $x_1 \rightarrow x_2$ ,  $Col^+(x_1) \not\subset Col^+(x_2)$ , so  $Col^+(x_2) = \{4\}$ . Similarly,  $Col^+(x_3) = \{4\}$  which is a contradiction.

Case (c): We have  $\mathcal{F} = \{\{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}$ . So we may assume that  $Col^+(x_1) = \{2\}$ . It follows that  $Col^+(x_2) = \{1, 3\}$  and  $Col^+(x_3) = \{3\}$ . Hence, none of  $Col^+(x_4)$  and  $Col^+(x_5)$  is  $\{2, 3\}$  since  $x_3 \rightarrow x_4$  and  $x_3 \rightarrow x_5$ , a contradiction.  $\square$

We will now prove that  $\Phi^\vee(3, 3) \leq 5$ . As in the proof of Theorem 35, in order to exhibit a homomorphism from a  $(3\vee 3)$ -digraph  $D$  to  $\overline{H}_5$ , we first show that there are two homomorphisms,  $h^+$  from  $D^+$  into a subdigraph  $S_3^+$  of  $\overline{H}_5$  and  $h^-$  from  $D^-$  into another subdigraph  $S_3^-$  of  $\overline{H}_5$ , with specific properties.

**Definition 39** Let  $S_3^+$  be the complete digraph with vertex-set  $\{s_1^+, \dots, s_7^+\}$ . Let  $S_3^-$  be the digraph with vertex-set  $\{s_1^-, \dots, s_9^-\}$  and arc-set  $\{s_i^- s_j^- : i \neq j\} \setminus \{s_2^- s_1^-, s_3^- s_1^-\}$ .

**Lemma 40** Let  $D$  be a 3-digraph. There exists a homomorphism  $h^+$  from  $D$  to  $S_3^+$  such that the vertices  $x$  with  $h^+(x) \in \{s_6^+, s_7^+\}$  have outdegree 3.

**Proof.** Let us prove it by induction on  $n = |V(D)|$ . If there exists a vertex  $x$  with  $d^+(x) + d^-(x) \leq 4$  then we obtain the desired homomorphism  $h^+$  from  $D - x$  to  $S_3^+$  and extend it with a suitable choice of  $h^+(x)$  in  $\{s_1^+, \dots, s_5^+\}$ .

Assume now that  $d^+(x) + d^-(x) \geq 5$  for every  $x$ . Let  $n_i$  be the number of vertices with outdegree  $i$ . Clearly,  $n = n_0 + n_1 + n_2 + n_3$ . Moreover, we have:

$$3n \geq \sum_{x \in V} d^+(x) = \sum_{x \in V} d^-(x) = \sum_{d^+(x)=0} d^-(x) + \sum_{d^+(x)=1} d^-(x) + \sum_{d^+(x)=2} d^-(x) + \sum_{d^+(x)=3} d^-(x)$$

Then, by assumption:

$$3n \geq 5n_0 + 4n_1 + 3n_2 + \sum_{d^+(x)=3} d^-(x)$$

If there is no vertex with outdegree 3, then  $D$  is 5-colourable and there is an homomorphism  $h^+$  from  $D$  to  $S_3^+[\{s_1^+, \dots, s_5^+\}]$ . Suppose now that there exists a vertex with outdegree 3. Then, there exists a vertex with outdegree 3 and indegree at most 3. If not,  $d^-(x) \geq 4$  for every  $x$  with  $d^+(x) = 3$  and the previous inequality implies  $3n \geq 5n_0 + 4n_1 + 3n_2 + 4n_3$  with  $n_3 \neq 0$ , what contradicts  $n = n_0 + n_1 + n_2 + n_3$ .

Finally, let  $x$  be a vertex with outdegree 3 and indegree at most 3. By induction hypothesis, there is a homomorphism  $h^+$  from  $D - x$  to  $S_3^+$  with the required property. As  $x$  has at most 6 neighbours, we extend  $h^+$  with a suitable choice for  $h^+(x)$  in  $\{s_1^+, \dots, s_7^+\}$ .  $\square$

**Lemma 41** Let  $D$  be a digraph with maximal indegree at most 3. There exists a homomorphism  $h^-$  from  $D$  to  $S_3^-$  such that the vertices  $x$  with  $h^-(x) \in \{s_6^-, \dots, s_9^-\}$  have indegree 3.

**Proof.** We prove the result by induction on  $|V(D)|$ .

If every vertex have indegree at most 2 then, by the dual form of the Lemma 17, there exists a homomorphism from  $D$  to  $S_3^-[\{s_1^-, \dots, s_5^-\}]$ .

Now, let  $x$  be a vertex with indegree 3. Let  $y_1, y_2$  and  $y_3$  be the outneighbours of  $x$ . By induction, there is a homomorphism  $h^-$  from  $D - x$  to  $S_3^-$  with the required property. In particular, as the vertex  $y_i$ ,  $1 \leq i \leq 3$ , has indegree at most 2 in  $D - x$ , we have  $h^-(y_i) \in \{s_1^-, \dots, s_5^-\}$ . So, as  $x$  has 3 inneighbours, we can extend  $h^-$  with a suitable choice for  $h^-(x)$  in  $\{s_6^-, \dots, s_9^-\}$ .  $\square$

**Theorem 42**

$$\Phi^\vee(3, 2) = \Phi^\vee(3, 3) = 5$$

**Proof.** By Proposition 38,  $5 \leq \Phi^\vee(3, 2) \leq \Phi^\vee(3, 3)$ . We will prove that  $\Phi^\vee(3, 3) \leq 5$ . Let  $D$  be a  $(3 \vee 3)$ -digraph, we will provide a homomorphism from  $D$  to  $\overline{H}_5$ .

By Lemma 40, there is a homomorphism  $h^+ : D^+ \rightarrow S_3^+$  such that if  $h^+(x) \in \{s_6^+, s_7^+\}$  then  $d_{D^+}^+(x) = 3$ . Moreover by Lemma 41, there is a homomorphism  $h^- : D^- \rightarrow S_3^-$  such that if  $h^-(x) \in \{s_6^-, \dots, s_9^-\}$ , then  $d_{D^-}^-(x) = 3$ .

Let  $S_3$  be the digraph obtained from the disjoint union of  $S_3^+$  and  $S_3^-$  by adding the arcs of  $\{s_i^- s_j^+ : 1 \leq i \leq 9, 1 \leq j \leq 7\} \cup \{s_i^+ s_j^- : i = 1, \dots, 5 \quad j = 1, \dots, 5\}$ . The mapping  $h : D \rightarrow S_3$  defined by  $h(x) = h^+(x)$  if  $x \in V^+$  and  $h(x) = h^-(x)$  if  $x \in V^-$  is a homomorphism. Indeed if  $xy$  is an arc of  $D$  with  $x \in V^+$  and  $y \in V^-$ , conditions on  $h^+$  and  $h^-$  imply that  $h(x) = h^+(x) \in \{s_1^+, \dots, s_5^+\}$  and  $h(y) = h^-(y) \in \{s_1^-, \dots, s_5^-\}$ . To conclude, Figure 3 provides a homomorphism  $g$  from  $S_3$  to  $\overline{H}_5$ . Inside  $S_3^-$  and  $S_3^+$ , only the arcs which are not in a circuit of length 2 are represented, every pair of not adjacent vertices are, in fact, linked by two arcs, one in each way.  $\square$

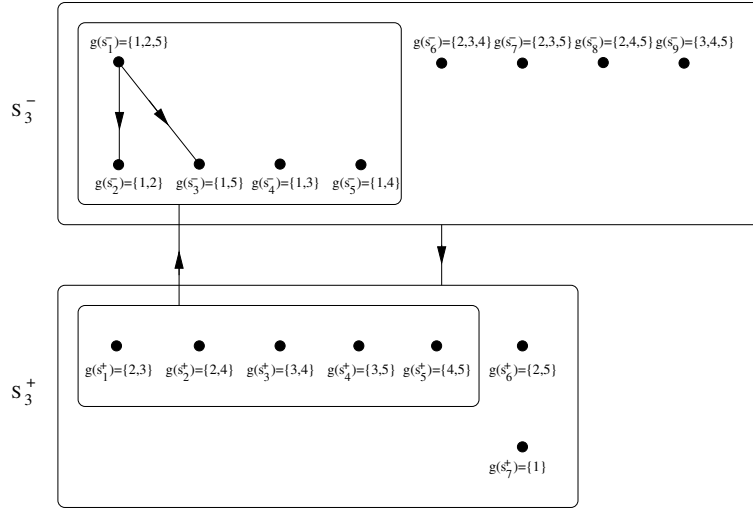


Figure 3: The homomorphism  $g$  from  $S_3$  to  $\overline{H}_5$ .

## References

- [1] R. L. Brooks, On colouring the nodes of a network. *Proc Cambridge Phil. Soc.* **37** (1941), 194–197.
- [2] A. El-Sahili, Fonctions de graphes disjoints, *C. R. Acad. Sci. Paris* **319** Série I (1994), 519–521.



- [3] A. El-Sahili, Functions and Line Digraphs, *J. of Graph Theory* **44** (2003), no. 4, 296–303.
- [4] S. Poljak and V. Rödl, On the arc-chromatic number of a digraph, *J. Combin. Theory Ser. B* **31** (1981), no. 2, 190–198.
- [5] E. Sperner, Ein Satz über Untermengen einer endlichen Menge, *Math. Zeit.* **27** (1928).

## 2.5 Wavelengths Division Multiplexing Networks

### Optical index of fault tolerant routings in WDM networks

S. Bessy

Laboratoire LIRMM - Université Montpellier 2,

161, rue Ada, 34000 Montpellier,

France

`bessy@lirmm.fr`

and

C. Lepelletier,

Projet Mascotte, CNRS/INRIA/UNSA,

INRIA Sophia-Antipolis,

2004 route des Lucioles BP 93,

06902 Sophia-Antipolis Cedex,

France

June 9, 2009

#### Abstract

Mañuch and Stacho [7] introduced the problem of designing  $f$ -tolerant routings in optical networks, i.e., routings which still satisfy the given requests even if  $f$  failures occur in the network. In this paper, we provide  $f$ -tolerant routings in complete and complete balanced bipartite optical networks, optimal according to two parameters: the arc-forwarding index and the optical index. These constructions use tools from design theory and graph theory and improve previous results of Dinitz, Ling and Stinson [4] for the complete network, and Gupta, Mañuch and Stacho [5] for the complete balanced bipartite network.

**Keywords:** optical networks, forwarding and optical indices, routing, fault tolerance

## 1 Introduction

In this paper, we are interested in a problem arising in the design of optical networks. Using models of graph theory and design theory, this topic has been of considerable interest over the last decade (see [1], [2] or [5] for instance). Readers may refer to [1] for a background review of optical networks. The model studied in this article is valid for the so-called *wavelength division multiplexing* (or *WDM*) optical network. Such a network is modeled by a symmetric directed graph with arcs representing the fiber-optic links. A *request* in the network is an ordered pair of graph nodes, representing a possible communication in the network. A set of different requests is an *instance* in the network. For each request of the instance, we have to select a routing directed path to satisfy it, and the set of all selected paths forms a *routing set* according to the instance. To make the communications possible, a wavelength is allocated to each routing path, such that two paths sharing an arc do not carry the same wavelength; otherwise the corresponding communications could be perturbed. Given a routing set related to the wavelength assignment, we can define two classical invariants. The *arc-forwarding index* of the routing set is the maximum number of paths sharing the same arc. In the network, there is a general bound on the number of wavelengths which can transit at the same time in a fiber-optic link, corresponding to the admissible maximal arc-forwarding index. The other invariant, called the *optical index* of the routing set, is the minimum number of wavelengths to assign to the routing paths in order to ensure that there is no interference in the network. The main challenge here is to provide, for a given instance, a routing set which minimizes the arc-forwarding index or the optical index, or both if possible.

Our work is a contribution to a variant of this problem, introduced by Mañuch and Stacho [7], in which we focus on possible breakdowns of nodes in the network. Precisely, for a given fixed integer  $f$ , we have to provide, for every request, not just one directed path to satisfy it, but rather a set of  $f+1$  directed paths with the same beginning and end nodes (corresponding to the request) and which are internally disjoint. In this routing, if  $f$  nodes break down, every request between the remaining nodes could still be satisfied by a previously selected routing path which contains no failed component. Such a routing set of directed paths is called an  *$f$ -fault tolerant routing* or an  *$f$ -tolerant routing*.

In this paper, we focus on the very special cases of complete symmetric directed graphs and

complete balanced bipartite symmetric directed graphs. Moreover, we only study the case of *all-to-all* communication, i.e., where the instance of the problem is the set of all ordered pairs of nodes of the network. Some results on these problems were presented by Gupta, Mañuch and Stacho [5] and Dinitz, Ling and Stinson [4]. We improve these results: for complete symmetric directed graphs, we show that optimal routings for the arc-forwarding index given in [5] are also optimal for the optical index. And for complete balanced bipartite symmetric directed graphs, we provide routings that are optimal for both parameters.

## 2 Preliminaries

In this section, we specify the previous definitions and formalize the problem. For the purpose of the paper, we only describe the case of all-to-all communication, but the notions can be extended to any kind of instances. We mainly use the notations proposed in [5].

We model an all-optical network as a symmetric directed graph  $D = (V(D), A(D))$ , where  $V(D)$  is the vertex set of  $D$  and  $A(D)$  is the arc set with the additional property that if  $(u, v) \in A(D)$  then  $(v, u) \in A(D)$ . If no confusion is possible, we simply write  $V$  and  $A$  instead of  $V(D)$  and  $A(D)$ , respectively. All paths and circuits are considered as oriented.

A directed graph  $D$  is *strongly connected* if, for every two vertices  $x$  and  $y$  of  $D$ , there is a path from  $x$  to  $y$  in  $D$ . In a symmetric directed graph, strong connectivity is equivalent to connectivity of the underlying non-oriented graph. So, for an integer  $k \geq 1$ , a symmetric directed graph  $D$  is *k-connected* if, for every set  $\{x_1, \dots, x_{k-1}\}$  of vertices of  $D$ ,  $D \setminus \{x_1, \dots, x_{k-1}\}$  is strongly connected.

For a fixed positive integer  $f$ , an *f-tolerant routing* in  $D$  is a set of paths:

$$\mathcal{R} = \{P_i(u, v) : u, v \in V, u \neq v, i = 0, \dots, f\}$$

where, for each pair of distinct vertices  $u, v \in V$ , the paths  $P_0(u, v), \dots, P_f(u, v)$  are internally vertex disjoint. Note that such a set of paths exists if and only if the connectivity of the directed graph is large enough (at least  $f + 1$ ), which will be the case in complete and complete bipartite networks for suitable  $f$ .

The basic parameters for WDM optical networks, the arc-forwarding index and the optical index, are generalized in *f-tolerant routings*. The *load* of an arc in  $\mathcal{R}$  is the number of directed

paths of  $\mathcal{R}$  containing it. By extension, the maximum load over all the arcs of  $D$  is the *load* of the routing, which is also called the *arc-forwarding index* of  $\mathcal{R}$  and is denoted by  $\pi(\mathcal{R})$ . Each path of  $\mathcal{R}$  receives a wavelength in the network to enable the communication and, to avoid interference, two paths sharing an arc do not receive the same wavelength. Like graph coloring, we speak about wavelengths as colors to assign to the paths of  $\mathcal{R}$ . Finally, the *optical index* of  $\mathcal{R}$ , denoted  $w(\mathcal{R})$ , is the minimum number of wavelengths to assign to paths of  $\mathcal{R}$  so that no two paths sharing an arc receive the same wavelength. In other words,  $w(\mathcal{R})$  is exactly the chromatic number of the graph with vertex set  $\mathcal{R}$  and where two paths of  $\mathcal{R}$  are linked if they share the same arc of  $D$  (known as the *path graph* of  $\mathcal{R}$ ).

The goal is to minimize  $\pi(\mathcal{R})$  and  $w(\mathcal{R})$ . So the *f-tolerant arc-forwarding index* of  $D$  and the *f-tolerant optical index* of  $D$  are respectively defined by:

$$\pi_f(D) = \min_{\mathcal{R}} \pi(\mathcal{R})$$

$$w_f(D) = \min_{\mathcal{R}} w(\mathcal{R})$$

where the minima span all the possible routing sets  $\mathcal{R}$ . A routing set achieving one of the bounds is said to be *optimal for the arc-forwarding index* or *optimal for the optical index*, respectively.

For a routing set  $\mathcal{R}$ , all paths sharing the same arc must receive different wavelengths in the computation of  $w(\mathcal{R})$ . In particular, we have  $\pi(\mathcal{R}) \leq w(\mathcal{R})$ . By considering a routing set which is optimal for the optical index, we obtain  $\pi_f(D) \leq w_f(D)$ . The equality was conjectured by Mañuch and Stacho [7].

**Conjecture 1 (J. Mañuch, L. Stacho, 2003, [7])** *Let  $D$  be a symmetric directed  $k$ -connected graph. For any  $f$ ,  $0 \leq f < k$ , we have  $\pi_f(D) = w_f(D)$ .*

For  $f = 0$  (without tolerating any faults), the conjecture was previously raised by Beauquier et al. [1].

Let  $K_n^*$  denote the *complete symmetric directed graph* with vertex set  $\{x_1, \dots, x_n\}$  and arc set  $\{x_i x_j : i \neq j\}$ . The *complete balanced bipartite symmetric digraph*  $K_{n,n}^*$  is the directed graph on vertex set  $X \cup Y$  with  $X = \{x_1, \dots, x_n\}$  and  $Y = \{y_1, \dots, y_n\}$  and arc set  $\{xy, yx : x \in X, y \in Y\}$ . The arc-forwarding indices of  $K_n^*$  and  $K_{n,n}^*$  were computed by Gupta, Mañuch and Stacho in [5].

Moreover, they give  $w_f(K_n^*)$  and  $w_f(K_{n,n}^*)$  up to a multiplicative factor. In [4], Dinitz, Ling and Stinson compute, among other things,  $w_f(K_n^*)$  up to an additive constant in some cases.

In the next two sections, we provide exact values for  $w_f(K_n^*)$  and  $w_f(K_{n,n}^*)$  and hence prove Conjecture 1 for the complete symmetric directed graph and the complete balanced bipartite symmetric directed graph.

### 3 Complete optical network

It is easy to provide a lower bound for the arc-forwarding index of  $K_n^*$ . Indeed, any two vertices  $x$  and  $y$  of  $K_n^*$  have to be linked in an  $f$ -tolerant routing by  $f + 1$  internally disjoint paths. If one of these paths has length one (the direct arc  $xy$ ), all the others have length at least two, and at least  $2f + 1$  arcs are needed to ensure  $f$ -tolerant communication from  $x$  to  $y$ . So, by an average argument, one arc of  $K_n^*$  must have load at least  $2f + 1$ , providing  $w_f(K_n^*) \geq 2f + 1$ . In the case of  $K_n^*$ , Gupta, Mañuch and Stacho prove that this lower bound gives exactly the value of the arc-forwarding index. Indeed, they construct  $f$ -tolerant routings through families of independent idempotent Latin squares in [5], which are optimal for the arc-forwarding index.

**Theorem 2 (A. Gupta, J. Mañuch, L. Stacho, 2005, [5])** *For every  $f$  with  $0 \leq f \leq n - 2$ , we have  $\pi_f(K_n^*) = 2f + 1$ .*

They also partially bound the optical index of their  $f$ -tolerant routings, proving that  $w_f(K_n^*) \leq 3f + 1$  for some  $f$ . This result was improved by Dinitz, Ling and Stinson [4], who gave a better multiplicative factor for some infinite sets of values of  $n$  and the optimal index up to an additive constant for another infinite set of values of  $n$ . We improve these results by showing that every  $f$ -tolerant routing set of  $K_n^*$  which is optimal for the arc-forwarding index is also optimal for the optical index.

**Theorem 3** *For every  $f$ ,  $0 \leq f \leq n - 2$ , and every  $f$ -tolerant routing set  $\mathcal{R}$  of  $K_n^*$  with  $\pi(\mathcal{R}) = \pi_f(K_n^*) = 2f + 1$ , we have  $w(\mathcal{R}) = 2f + 1$ . In particular, we have  $w_f(K_n^*) = \pi_f(K_n^*) = 2f + 1$ .*

**Proof.** Let  $f$  be fixed with  $0 \leq f \leq n - 2$  and consider an  $f$ -tolerant routing  $\mathcal{R}$  of  $K_n^*$  which is optimal for the arc-forwarding index, i.e., of value  $2f + 1$ . By the tightness of the lower bound, for any two vertices  $x$  and  $y$  of  $K_n^*$  there is exactly one path  $xy$  of length 1 and  $f$  paths of length

2 from  $x$  to  $y$  in  $\mathcal{R}$  (otherwise, summing up the total load gives  $\pi(\mathcal{R}) \geq 2f + 2$ ). Hence, every arc of  $K_n^*$  has a load of exactly  $2f + 1$  and appears in one path of length 1 and  $2f$  paths of length two in  $\mathcal{R}$ . Now, define the graph  $H$  with vertex set being the set of the arcs of  $K_n^*$  and link two arcs of  $K_n^*$  if they belong to the same path of  $\mathcal{R}$  of length 2. Thus, we have a one-to-one correspondence between the edge set of  $H$  and the paths of length 2 of  $\mathcal{R}$ . Since each arc of  $K_n^*$  belongs to exactly  $2f$  paths of  $\mathcal{R}$  of length 2,  $H$  is regular with degree  $2f$ . By Vizing's Theorem (see [3] or [9]), the edges of  $H$  can be colored with  $2f + 1$  distinct colors such that any two adjacent edges receive distinct colors. This provides a coloring of the paths of length 2 of  $\mathcal{R}$  with  $2f + 1$  colors. To conclude, a path of length 1 of  $\mathcal{R}$  intersects exactly  $2f$  paths of length 2 and we can color this path with the remaining color.  $\square$

Moreover, the edge-coloring provided by Vizing's Theorem can be computed in polynomial time (polynomial in the size of the input graph,  $H$  here). So, given an optimal routing for the arc-forwarding index, this proof gives a polynomial algorithm (polynomial in  $n$  and  $f$ ) to obtain a wavelength assignment for a routing which is optimal for the optical index.

## 4 Complete balanced bipartite optical network

In this section, we compute the exact optical index of  $K_{n,n}^*$  and thus prove Conjecture 1 for this family of graphs. This improves the result in [5], where the upper bound given on the optical index of  $K_{n,n}^*$  is 20% higher than the conjectured optimal value.

**Theorem 4** *For any  $n \geq 1$  and any  $f$  with  $0 \leq f \leq n - 1$ , we have  $w_f(K_{n,n}^*) = \pi_f(K_{n,n}^*)$ .*

To prove Theorem 4, we provide a routing set for  $K_{n,n}^*$  which is optimal both for the arc-forwarding index and the optical index. The construction depends on the values of  $n$  and  $f$ . Recall that  $X \cup Y$  denotes the canonical partition of  $K_{n,n}^*$ , with  $X = \{x_1, \dots, x_n\}$  and  $Y = \{y_1, \dots, y_n\}$ . For convenience, indices of the vertices of  $X$  and  $Y$  are computed modulo  $n$ .

### 4.1 Routing set and arc-forwarding index

We use the paths of minimum length to route in  $K_{n,n}^*$ . Indeed, for two vertices  $x$  and  $y$ , we use paths of length 2 when  $x$  and  $y$  belong to the same partite set of  $K_{n,n}^*$  and one path of length 1

and other paths of length 3 (if  $f > 0$ ) when  $x$  and  $y$  do not belong to the same partite set. The main challenge here is to pack the paths of length 3 in order to provide the right optical index. This will be achieved by using the following decomposition result.

**Theorem 5 (Tillson, 1980, [8])** *The arcs of  $K_n^*$  can be partitioned into Hamiltonian circuits if and only if  $n \neq 4$  and  $n \neq 6$ .*

For  $n \neq 4$  and  $n \neq 6$ ,  $\{C_1, \dots, C_{n-1}\}$  denotes a set of  $n - 1$  Hamiltonian circuits which partition the arcs of  $K_n^*$ . Let  $\{1, \dots, n\}$  denote the vertex set of  $K_n^*$ . Moreover, for a vertex  $i$  of  $K_n^*$  and a circuit  $C_k$ , the out-neighbor of  $i$  in  $C_k$  is denoted  $C_k(i)$ . We use  $C_k$  as a functional notation: for  $p \geq 1$ ,  $C_k^p(i) = C_k(C_k^{p-1}(i))$ . Moreover, we compute the powers of  $C_k$  modulo  $n$ , in particular  $C_k^0(i) = i$  and  $C_k^{-1}(i)$  is the in-neighbor of  $i$  in  $C_k$ .

The previous Hamiltonian decomposition is used to route paths of length 3 in  $K_{n,n}^*$ . For paths of length 2, we use a *Latin square*  $A$  of order  $n$ , i.e., a  $n \times n$  matrix in which each row and each column is a permutation of the set  $\{1, \dots, n\}$ . Moreover, we require  $A$  to be *idempotent*: for every  $i$ ,  $1 \leq i \leq n$ , we have  $A(i, i) = i$ . An idempotent Latin square exists for each value of  $n$ , except for  $n = 2$  (see [6], Chapter 2, for an explicit construction). For  $0 \leq k \leq n - 1$ ,  $M_k = \{x_i y_{i+k}, y_{i+k} x_i : 1 \leq i \leq n\}$ , denotes the symmetric orientations of  $n$  disjoint matchings which partition the arcs of  $K_{n,n}^*$  (see Figure 1). The indices of  $M_k$  are computed modulo  $n$ . In our figures, we represent two symmetric arcs by a (non-oriented) edge.

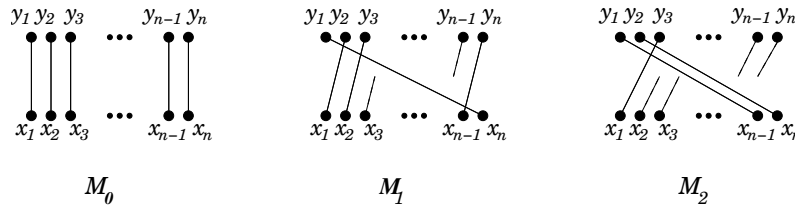


Figure 1: Some matchings  $M_k$ .

Now, for  $n \neq 2$ ,  $n \neq 4$  and  $n \neq 6$ , we describe three kinds of paths to construct the routing set:



- Paths of length 1 between vertices which belong to different partite sets of  $K_{n,n}^*$ :

$$\mathcal{D}_0[X, Y] = \{x_i y_j : 1 \leq i \leq n, 1 \leq j \leq n\}$$

$$\mathcal{D}_0[Y, X] = \{y_i x_j : 1 \leq i \leq n, 1 \leq j \leq n\}$$

$$\mathcal{D}_0 = \mathcal{D}_0[X, Y] \cup \mathcal{D}_0[Y, X]$$

- Paths of length 3 between vertices which belong to different partite sets of  $K_{n,n}^*$ , for  $1 \leq k \leq n-1$ :

$$\mathcal{D}_k[X, Y] = \{x_i y_{C_k^{-1}(j)} x_{C_k(i)} y_j : 1 \leq i \leq n, 1 \leq j \leq n\}$$

$$\mathcal{D}_k[Y, X] = \{y_i x_{C_k^{-1}(j)} y_{C_k(i)} x_j : 1 \leq i \leq n, 1 \leq j \leq n\}$$

$$\mathcal{D}_k = \mathcal{D}_k[X, Y] \cup \mathcal{D}_k[Y, X]$$

- Paths of length 2 between vertices which belong to the same partite set of  $K_{n,n}^*$ , for  $0 \leq k \leq n-1$ :

$$\mathcal{S}_k[X, X] = \{x_i y_{A(i,j)+k} x_j : 1 \leq i \leq n, 1 \leq j \leq n, i \neq j\}$$

$$\mathcal{S}_k[Y, Y] = \{y_i x_{A(i,j)+k+\lceil n/2 \rceil} y_j : 1 \leq i \leq n, 1 \leq j \leq n, i \neq j\}$$

$$\mathcal{S}_k = \mathcal{S}_k[X, X] \cup \mathcal{S}_k[Y, Y]$$

And for a fixed  $f$ ,  $0 \leq f \leq n-1$ , we define the routing set  $\mathcal{R}_f$  by

$$\mathcal{R}_f = \bigcup_{k=0}^f (\mathcal{S}_k \cup \mathcal{D}_k)$$

Note that, by construction, for distinct vertices  $x$  and  $y$  of  $K_{n,n}^*$ ,  $\mathcal{R}_f$  contains exactly  $f+1$  internally disjoint paths from  $x$  to  $y$ . So,  $\mathcal{R}_f$  is an  $f$ -tolerant routing for an all-to-all instance in  $K_{n,n}^*$ . Moreover, note that every arc of  $K_{n,n}^*$  appears exactly once in  $\mathcal{D}_0$  and three times in  $\mathcal{D}_k$  for  $1 \leq k \leq n-1$  (for instance the arc  $x_i y_j$  appears in  $\mathcal{D}_k$  in paths from  $x_i$  to  $y_{C_k(j)}$ , from  $x_{C_k^{-1}(i)}$  to  $y_j$  and from  $y_{C_k^{-1}(j)}$  to  $x_{C_k(i)}$ ). For  $0 \leq k \leq n$ , the routing  $\mathcal{S}_k$  behaves slightly differently:  $\mathcal{S}_k[X, X]$  contains only pairwise arc-disjoint paths and the same holds for  $\mathcal{S}_k[Y, Y]$ . Moreover,  $\mathcal{S}_k[X, X]$  and  $M_k$  are disjoint and every arc not in  $M_k$  appears exactly once in  $\mathcal{S}_k[X, X]$ ; on the other hand,  $\mathcal{S}_k[Y, Y]$  and  $M_{k+\lceil n/2 \rceil}$  are disjoint and every arc not in  $M_{k+\lceil n/2 \rceil}$  appears exactly once in  $\mathcal{S}_k[Y, Y]$ . Using these remarks, we can give the arc-forwarding index of  $\mathcal{R}_f$ . The computation of  $\pi_f(K_{n,n}^*)$  was obtained in [5].

**Lemma 6 (A. Gupta, J. Mañuch, L. Stacho, 06, [5])** *The arc-forwarding of  $K_{n,n}^*$  is:*

$$\pi_f(K_{n,n}^*) = \begin{cases} 5f + 3 & \text{for } 0 \leq f \leq \lceil n/2 \rceil - 2 \\ 5f + 2 & \text{for } \lceil n/2 \rceil - 1 \leq f \leq n - 2 \\ 5f + 1 & \text{for } f = n - 1 \end{cases}$$

In fact, we prove that  $\mathcal{R}_f$  is optimal for the arc-forwarding index.

**Lemma 7** *For every  $f$ ,  $0 \leq f \leq n - 1$ , the routing set  $\mathcal{R}_f$  satisfies  $\pi(\mathcal{R}_f) = \pi_f(K_{n,n}^*)$ .*

**Proof.** For  $0 \leq f \leq \lceil n/2 \rceil - 2$ , every arc of  $K_{n,n}^*$  appears at least 5 times in  $\mathcal{S}_k \cup \mathcal{D}_k$  for  $1 \leq k \leq f$  and 3 times in  $\mathcal{S}_0 \cup \mathcal{D}_0$ , so the computation of  $\pi(\mathcal{R}_f)$  is clear. For  $f \geq \lceil n/2 \rceil - 1$ , every arc of  $K_{n,n}^*$  is in a matching  $M_k$  or  $M_{k+\lceil n/2 \rceil}$  for some  $k$ ,  $0 \leq k \leq f$  and thus is not in a path of one of the  $\mathcal{S}_k[X, X]$  or  $\mathcal{S}_k[Y, Y]$ . We then save 1 in the computation of  $\pi(\mathcal{R}_f)$  in these cases. Finally, if  $f = n - 1$ , every arc of  $K_{n,n}^*$  is in a matching  $M_k$  and in a matching  $M_{k'+\lceil n/2 \rceil}$  for some suitable  $k$  and  $k'$  in  $\{0, \dots, n - 1\}$ . So, every arc of  $K_{n,n}^*$  is not in any path of  $\mathcal{S}_k[X, X]$  and not in any path of  $\mathcal{S}_{k'}[Y, Y]$ , and we save two in the computation of  $\pi(K_{n,n}^*)$ .  $\square$

We then have a lower bound for the optical index of  $K_{n,n}^*$ , and now we prove that the routing set  $\mathcal{R}_f$  achieves this bound.

#### 4.2 Packing the paths of $\mathcal{R}_f$

A *color class* of paths of  $\mathcal{R}_f$  is set of paths which are pairwise arc-disjoint. To construct the different color classes, we need the following notations and definitions. To indicate a path or a set of paths of  $\mathcal{R}_f$ , we always specify the subset  $\mathcal{D}_k$  or  $\mathcal{S}_k$  it belongs to. For instance,  $\mathcal{S}_0[X, X]\{x_1 y_{A(1,2)} x_2\}$  is the path  $x_1 y_{A(1,2)} x_2$  of  $\mathcal{S}_0[X, X]$  and  $\mathcal{D}_0\{M_1\}$  is the set of paths in  $M_1$  of  $\mathcal{D}_0$ . We use the notation  $\star$  as a ‘joker’ instead of all the possible path names. For instance,  $\mathcal{S}_0[X, X]\{\star y_1 \star\}$  stands for all paths of  $\mathcal{S}_0[X, X]$  whose intermediate vertex is  $y_1$ . We specially focus on two particular subsets of paths. For  $k \geq 1$  and  $x_i$ , a vertex of  $K_{n,n}^*$ ,  $\mathcal{D}_k[X, Y]\{x_i \star x_{C_k(i)} \star\}$  contains the  $n$  paths of  $\mathcal{D}_k$  which start at  $x_i$  and whose third vertex is  $x_{C_k(i)}$ . As  $(C_k(1), C_k(2), \dots, C_k(n))$  is a permutation of  $(1, 2, \dots, n)$ , note that these paths are pairwise arc-disjoint and that they cover exactly all arcs beginning at  $x_i$ , all arcs beginning at  $x_{C_k(i)}$  and all arcs ending at  $x_{C_k(i)}$ . Moreover, we have

$\bigcup_{l=0}^{n-1} \mathcal{D}_k[X, Y]\{x_{C_k^l(1)} \star x_{C_k^{l+1}(1)} \star\} = \mathcal{D}_k[X, Y]$ . The set  $\mathcal{D}_k[X, Y]\{x_i \star x_{C_k(i)} \star\}$  and its representation on  $X$ , which shows the saturated in and out-neighborhood, are depicted in Figure 2. Similarly, we will use the sets  $\mathcal{D}_k[X, Y]\{\star y_i \star y_{C_k(i)}\}$ ,  $\mathcal{D}_k[Y, X]\{\star x_i \star x_{C_k(i)}\}$  and  $\mathcal{D}_k[Y, X]\{y_i \star y_{C_k(i)} \star\}$ .

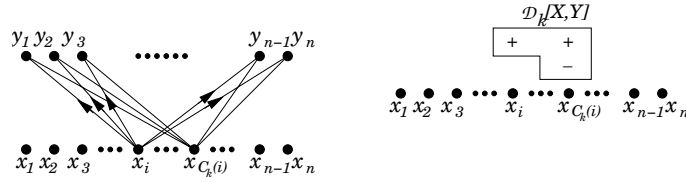


Figure 2: The set of paths  $\mathcal{D}_k[X, Y]\{x_i \star x_{C_k(i)} \star\}$  and its representation on  $X$ .

In addition, for  $0 \leq k \leq n-1$  and  $1 \leq i \leq n$ ,  $\overline{\mathcal{S}_k[Y, Y]}\{\star x_i \star\}$  denotes the set of paths  $\mathcal{S}_k[Y, Y]\{\star x_i \star\} \cup \mathcal{D}_0\{x_i y_{i+k+\lceil n/2 \rceil}, y_{i+k+\lceil n/2 \rceil} x_i\}$ . As  $\mathcal{S}_k[Y, Y]$  and  $M_{k+\lceil n/2 \rceil}$  are disjoint,  $\overline{\mathcal{S}_k[Y, Y]}\{\star x_i \star\}$  contains only arc-disjoint paths. Moreover, it contains exactly all arcs beginning and ending at  $x_i$ , and we have  $\bigcup_{i=1}^n \overline{\mathcal{S}_k[Y, Y]}\{\star x_i \star\} = \mathcal{S}_k[Y, Y] \cup \mathcal{D}_0\{M_{k+\lceil n/2 \rceil}\}$ . This set and the representation of  $\overline{\mathcal{S}_k[Y, Y]}\{\star x_i \star\}$  on  $X$  are shown in Figure 3. Similarly, we will use the sets  $\overline{\mathcal{S}_k[X, X]}\{\star y_i \star\}$ , disjoint from  $\mathcal{D}_0\{M_k\}$ .

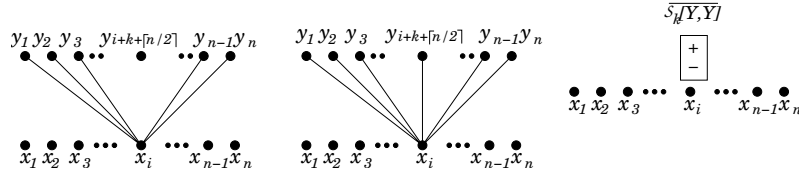


Figure 3: The sets  $\mathcal{S}_k[Y, Y]\{\star x_i \star\}$ ,  $\overline{\mathcal{S}_k[Y, Y]}\{\star x_i \star\}$  and the representation of  $\overline{\mathcal{S}_k[Y, Y]}\{\star x_i \star\}$  on  $X$ .

Now we can define the colors classes. They are constructed differently according to the value of  $f$ : we distinguish three main cases following the residue of  $n$  modulo 3. However, some particular cases occur:  $n = 2$  due to the non-existence of an idempotent Latin square of order

2,  $n = 4$  and  $n = 6$  which are exceptions to Tillson's Theorem, and  $n = 1$  and  $n = 5$  for which the techniques used in general cases  $n \equiv 1 \pmod{3}$  and  $n \equiv 2 \pmod{3}$  cannot be applied. Moreover, for any value of  $n$ , the case  $f = 0$  is different from the other cases.

#### 4.3 Routings without fault

The case  $f = 0$  is simpler than other cases in the sense that the paths of length 1 form a proper color class, whereas for  $f > 0$  these paths are used to complete other color classes. More precisely,  $\mathcal{R}_0 = \mathcal{D}_0 \cup \mathcal{S}_0$  and to obtain the optical index, we assign one different color to each of these sets of paths:  $\mathcal{D}_0$ ,  $\mathcal{S}_0[X, X]$  and  $\mathcal{S}_0[Y, Y]$ . So, for every  $n \geq 1$ , we have  $w_0(K_{n,n}^*) = \pi_0(K_{n,n}^*) = 3$ .

#### 4.4 Optical index in case $n \equiv 0 \pmod{3}$

This is the simplest case, so paths of length 2 and 3 can be packed separately. For any  $k$ ,  $1 \leq k \leq n - 1$ , to color paths of length 3, we define:

$$\begin{aligned} c_k^3 &= \bigcup_{t=0}^{n/3-1} \left( \mathcal{D}_k[Y, X] \{ \star x_{C_k^{3t}(1)} \star x_{C_k^{3t+1}(1)} \} \cup \mathcal{D}_k[X, Y] \{ x_{C_k^{3t+1}(1)} \star x_{C_k^{3t+2}(1)} \star \} \right) \\ c_k^4 &= \bigcup_{t=0}^{n/3-1} \left( \mathcal{D}_k[Y, X] \{ \star x_{C_k^{3t+1}(1)} \star x_{C_k^{3t+2}(1)} \} \cup \mathcal{D}_k[X, Y] \{ x_{C_k^{3t+2}(1)} \star x_{C_k^{3t+3}(1)} \star \} \right) \\ c_k^5 &= \bigcup_{t=0}^{n/3-1} \left( \mathcal{D}_k[Y, X] \{ \star x_{C_k^{3t+2}(1)} \star x_{C_k^{3t+3}(1)} \} \cup \mathcal{D}_k[X, Y] \{ x_{C_k^{3t+3}(1)} \star x_{C_k^{3t+4}(1)} \star \} \right) \end{aligned}$$

In Figure 4, we give the representation on  $X$  of  $c_k^3$ , where the vertices of  $X$  are sorted along  $C_k$ . A shift of one vertex on the right side (modulo  $n$ ) of the sets of paths gives the representation of  $c_k^4$ , for the same  $k$ . And a shift of two vertices gives the representation of  $c_k^5$ . Then, note that every path of  $\mathcal{D}_k[X, X]$  belongs to one of the class  $c_k^3$ ,  $c_k^4$  or  $c_k^5$ . For  $k \in \{0, \dots, n - 1\}$ , we also define  $c_k^1 = \overline{\mathcal{S}_k[Y, Y]}$ . Now, we pack the remaining paths of length 1 and 2 in classes  $c_k^2$  according to the value of  $f$ .

- If  $1 \leq f \leq \lceil n/2 \rceil - 2$ , for  $0 \leq k \leq f$ , we fix:

$$c_k^2 = \mathcal{S}_k[X, X]$$

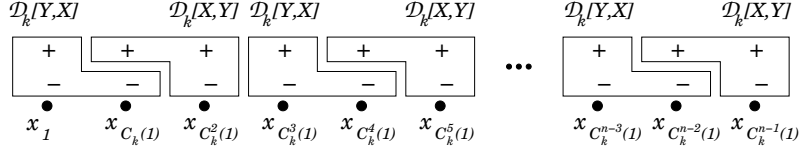


Figure 4: The representation on  $X$  of the color class  $c_k^3$ , where the vertices of  $X$  are sorted along  $C_k$ .

All paths of  $\mathcal{D}_0$  have not been used in the sets  $c_0^1 \cup c_0^2 \cup \bigcup_{k=1}^f \{c_k^1, c_k^2, c_k^3, c_k^4, c_k^5\}$  (sole  $\mathcal{D}_0\{M_{\lceil n/2 \rceil}\}, \mathcal{D}_0\{M_{1+\lceil n/2 \rceil}\}, \dots, \mathcal{D}_0\{M_{f+\lceil n/2 \rceil}\}$  are used). So, we put the remaining paths of  $\mathcal{D}_0$  in a class  $c_0^3$ , and we obtain, in this case,  $w_f(K_{n,n}^*) = 5f + 3$ .

- If  $\lceil n/2 \rceil - 1 \leq f \leq n - 2$ , we have enough space to pack the paths of length one in the other classes. Indeed,  $c_0^1 \cup c_0^2 \cup \bigcup_{k=1}^f \{c_k^1, c_k^3, c_k^4, c_k^5\}$  does not contain the direct paths  $\mathcal{D}_0\{M_{f+1+\lceil n/2 \rceil}\}, \mathcal{D}_0\{M_{f+2+\lceil n/2 \rceil}\}, \dots, \mathcal{D}_0\{M_{\lceil n/2 \rceil-1}\}$ , with  $n \leq f+1+\lceil n/2 \rceil \leq n+\lceil n/2 \rceil-1$  (note that indices of  $M_i$  are computed modulo  $n$ ). So, for  $0 \leq k \leq f$ , we set:

$$\begin{aligned} c_k^2 &= \overline{\mathcal{S}_k[X, X]} & \text{if } f+1+\lceil n/2 \rceil - n \leq k \leq \lceil n/2 \rceil - 1 \\ c_k^2 &= \mathcal{S}_k[X, X] & \text{else} \end{aligned}$$

So, all paths of  $\mathcal{D}_0$  are used, and we obtain  $w_f(K_{n,n}^*) = 5f + 2$ .

- If  $f = n - 1$ , we should save another color. This time, all paths of length 1 are packed in  $\bigcup_{k=0}^{n-1} c_k^2$  and the second saved color is obtained by optimally packing the set of paths  $\{\mathcal{S}_k[X, X] : 0 \leq k \leq n - 1\}$  using only  $n - 1$  colors. First, we complete the color class  $c_1^2$ . We start by covering all the arcs beginning at  $x_1$  and all the arcs ending at  $x_2$ , which can be done by placing in  $c_1^2$  all the paths from  $x_1$  to  $x_2$ :  $\bigcup_{i=0}^{n-1} \mathcal{S}_i[X, X]\{x_1 y_{A(1,2)+i} x_2\}$  (corresponding to the paths of  $\bigcup_{i=0}^{n-1} \mathcal{S}_i[X, X]$  which use the value  $A(1, 2)$  in the matrix  $A$ ). Then, we focus on the arcs beginning at  $x_2$  and the arcs ending at  $x_3$  using the paths:  $\bigcup_{i=0}^{n-1} \mathcal{S}_i[X, X]\{x_2 y_{A(2,3)+i} x_3\}$  (paths using  $A(2, 3)$ ). Subsequently, we cover the arcs beginning at  $x_p$  and the arcs ending at  $x_{p+1}$  for  $p = 1, 2, \dots, n$  (corresponding to all paths using values  $A(p, p+1)$  in  $A$ ). So,  $c_1^2$  forms a color class which covers exactly once all arcs of  $K_{n,n}^*$ . Once  $c_1^2$  is complete, we proceed in the same way to obtain  $c_2^2$ , using, this time, for  $p = 1, 2, \dots, n$ , the paths of  $\bigcup_{i=0}^{n-1} \mathcal{S}_i[X, X]$  obtained through the values  $A(p, p+2)$

to cover the arcs from  $x_p$  and those to  $x_{p+2}$ .

In general way,  $1 \leq k \leq n-1$ ,  $c_k^2$  is constructed using the paths obtained through the values  $A(p, p+k)$ , for  $p = 1, 2, \dots, n$ . More precisely, we define:

$$c_k^2 = \bigcup_{j=1}^n \bigcup_{i=0}^{n-1} \mathcal{S}_i[X, X] \{x_j \star x_{j+k}\}$$

Finally, we obtain the color classes  $\{c_0^1\} \cup \bigcup_{k=1}^{n-1} \{c_k^1, c_k^2, c_k^3, c_k^4, c_k^5\}$  which give the optimal value for the optical index:  $w_f(K_{n,n}^*) = 5f + 1$ .

#### 4.5 Optical index in case $n \equiv 1 \pmod{3}$

This time a color class cannot be composed with only paths of length 3. For  $1 \leq k \leq n-1$ , we pack together a maximum number of paths of length 3 (i.e., from  $\mathcal{D}_k$ ) in classes  $c_k^2, c_k^4$  and  $c_k^5$ , which we supplement with paths of length 2 from  $c_k^1$  (i.e., from  $\mathcal{S}_k$ ). Consequently,  $c_k^1$  contains the main part of paths of length 2 and the remaining paths of length 3, which is possible as soon as  $n \geq 3$ . Precisely, for  $n \geq 3$ , we construct  $c_0^1 = \overline{\mathcal{S}_0[Y, Y]}$ , and, for  $1 \leq k \leq n-1$ , we set:

$$c_k^3 = \left[ \bigcup_{t=0}^{(n-1)/3-1} \left( \mathcal{D}_k[Y, X] \{ \star x_{C_k^{3t+1}(1)} \star x_{C_k^{3t+2}(1)} \} \cup \mathcal{D}_k[X, Y] \{ x_{C_k^{3t+2}(1)} \star x_{C_k^{3t+3}(1)} \star \} \right) \right] \cup \left[ \overline{\mathcal{S}_k[Y, Y]} \{ \star x_1 \star \} \right]$$

$$c_k^4 = \left[ \bigcup_{t=0}^{(n-1)/3-1} \left( \mathcal{D}_k[Y, X] \{ \star x_{C_k^{3t+2}(1)} \star x_{C_k^{3t+3}(1)} \} \cup \mathcal{D}_k[X, Y] \{ x_{C_k^{3t+3}(1)} \star x_{C_k^{3t+4}(1)} \star \} \right) \right] \cup \left[ \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k(1)} \star \} \right]$$

$$c_k^5 = \left[ \bigcup_{t=0}^{(n-1)/3-1} \left( \mathcal{D}_k[Y, X] \{ \star x_{C_k^{3t+3}(1)} \star x_{C_k^{3t+4}(1)} \} \cup \mathcal{D}_k[X, Y] \{ x_{C_k^{3t+4}(1)} \star x_{C_k^{3t+5}(1)} \star \} \right) \right] \cup \left[ \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k^2(1)} \star \} \right]$$

$$c_k^1 = \left[ \left( \overline{\mathcal{S}_k[Y, Y]} \right) \setminus \left( \overline{\mathcal{S}_k[Y, Y]} \{ \star x_1 \star \} \cup \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k(1)} \star \} \cup \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k^2(1)} \star \} \right) \right] \cup$$

$$\left[ \mathcal{D}_k[Y, X] \{ \star x_1 \star x_{C_k(1)} \} \cup \mathcal{D}_k[X, Y] \{ x_{C_k(1)} \star x_{C_k^2(1)} \star \} \right]$$

As previously, Figure 5 shows the representation on  $X$  of  $c_k^3$ , where the vertices of  $X$  are sorted along  $C_k$ . A shift of one vertex on the right (modulo  $n$ ) of the sets of paths gives the representation of  $c_k^4$  for the same  $k$ . And a shift of two vertices gives the representation of  $c_k^5$ .

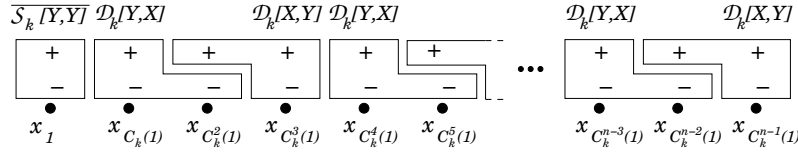


Figure 5: The representation on  $X$  of the color class  $c_k^3$ , where the vertices of  $X$  are sorted along  $C_k$ .

Figure 6 gives the representation on  $X$  of the color classes  $c_k^1$ . Note that  $c_k^1$  is well defined only if  $n \geq 3$ .

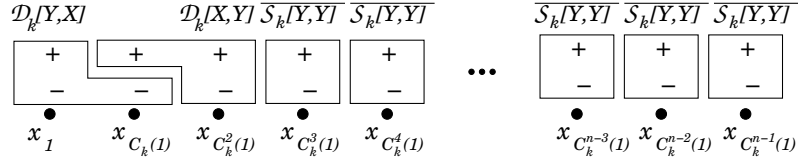


Figure 6: The representation on  $X$  of the color class  $c_k^1$ , where the vertices of  $X$  are sorted along  $C_k$ .

Now, the exact optical index will be obtained by packing the remaining paths of length 1 and 2, as done previously: the classes  $c_k^2$  for  $0 \leq k \leq f$  are defined exactly as in the case  $n \equiv 0 \pmod{3}$ . Indeed, even if the definitions of classes  $c_k^1$ ,  $c_k^3$ ,  $c_k^4$  and  $c_k^5$  have changed, the use of the paths  $\mathcal{D}_0\{M_i\}$  is still the same: for every  $k$ ,  $0 \leq k \leq f$ , we have  $\mathcal{D}_0\{M_k\} \subset c_k^1 \cup c_k^3 \cup c_k^4 \cup c_k^5$ . Finally, we obtain:

- If  $1 \leq f \leq \lceil n/2 \rceil - 2$ ,  $w_f(K_{n,n}^*) = 5f + 3$ .
- If  $\lceil n/2 \rceil - 1 \leq f \leq n - 2$ ,  $w_f(K_{n,n}^*) = 5f + 2$ .

- And, if  $f = n - 1$ ,  $w_f(K_{n,n}^*) = 5f + 1$ .

#### 4.6 Optical index in case $n \equiv 2 \pmod{3}$

We proceed as in case  $n \equiv 1 \pmod{3}$ , except that this time we supplement each class of paths of length 3 with two sets of paths from  $\overline{\mathcal{S}_k[Y, Y]}$ . Conversely, the classes  $c_k^1$  contain the remaining paths of  $\overline{\mathcal{S}_k[Y, Y]}$  and 4 sets of paths of length 3 from  $\mathcal{D}_k$ , provided that  $n \geq 6$ . Precisely, for  $n \geq 6$ , we construct  $c_0^1 = \overline{\mathcal{S}_0[Y, Y]}$ , and for  $1 \leq k \leq n - 1$ , we set:

$$c_k^3 = \left[ \bigcup_{t=0}^{(n-2)/3-1} \left( \mathcal{D}_k[Y, X] \{ \star x_{C_k^{3t+2}(1)} \star x_{C_k^{3t+3}(1)} \} \cup \mathcal{D}_k[X, Y] \{ x_{C_k^{3t+3}(1)} \star x_{C_k^{3t+4}(1)} \star \} \right) \right] \cup \left[ \overline{\mathcal{S}_k[Y, Y]} \{ \star x_1 \star \} \cup \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k(1)} \star \} \right]$$

$$c_k^4 = \left[ \bigcup_{t=0}^{(n-2)/3-1} \left( \mathcal{D}_k[Y, X] \{ \star x_{C_k^{3t+4}(1)} \star x_{C_k^{3t+5}(1)} \} \cup \mathcal{D}_k[X, Y] \{ x_{C_k^{3t+5}(1)} \star x_{C_k^{3t+6}(1)} \star \} \right) \right] \cup \left[ \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k^2(1)} \star \} \cup \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k^3(1)} \star \} \right]$$

$$c_k^5 = \left[ \bigcup_{t=0}^{(n-2)/3-1} \left( \mathcal{D}_k[Y, X] \{ \star x_{C_k^{3t+6}(1)} \star x_{C_k^{3t+7}(1)} \} \cup \mathcal{D}_k[X, Y] \{ x_{C_k^{3t+7}(1)} \star x_{C_k^{3t+8}(1)} \star \} \right) \right] \cup \left[ \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k^4(1)} \star \} \cup \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k^5(1)} \star \} \right]$$

$$c_k^1 = \left[ \left( \overline{\mathcal{S}_k[Y, Y]} \right) \setminus \left( \overline{\mathcal{S}_k[Y, Y]} \{ \star x_1 \star \} \cup \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k(1)} \star \} \cup \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k^2(1)} \star \} \cup \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k^3(1)} \star \} \cup \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k^4(1)} \star \} \cup \overline{\mathcal{S}_k[Y, Y]} \{ \star x_{C_k^5(1)} \star \} \right) \right] \cup \left[ \mathcal{D}_k[Y, X] \{ \star x_1 \star x_{C_k(1)} \} \cup \mathcal{D}_k[X, Y] \{ x_{C_k(1)} \star x_{C_k^2(1)} \star \} \cup \mathcal{D}_k[Y, X] \{ \star x_{C_k^3(1)} \star x_{C_k^4(1)} \} \cup \mathcal{D}_k[X, Y] \{ x_{C_k^4(1)} \star x_{C_k^5(1)} \star \} \right]$$



Once again, we give in Figure 7 the representation on  $X$  of  $c_k^3$ , assuming that vertices of  $X$  are sorted along  $C_k$ . A shift of two vertices on the right side (modulo  $n$ ) of the sets of paths gives the representation of  $c_k^4$ , for the same  $k$ . And a shift of four vertices gives the representation of  $c_k^5$ .

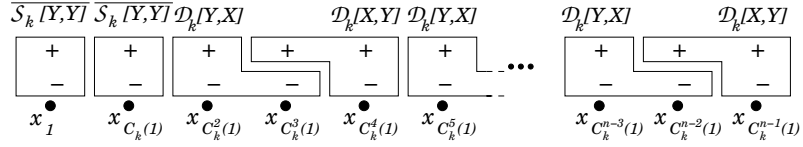


Figure 7: The representation on  $X$  of the color class  $c_k^3$ , where the vertices of  $X$  are sorted along  $C_k$ .

Figure 8 gives the representation on  $X$  of the color classes  $c_k^1$ . Note that  $c_k^1$  is well defined only if  $n \geq 6$ .

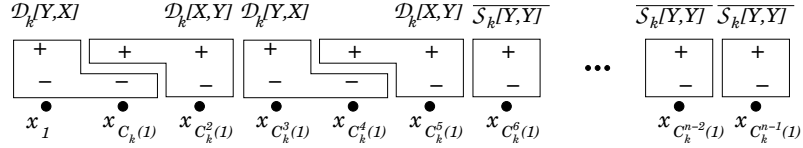


Figure 8: The representation on  $X$  of the color class  $c_k^1$ , where the vertices of  $X$  are sorted along  $C_k$ .

To conclude, the classes  $c_k^2$  for  $0 \leq k \leq f$  are defined exactly as in the cases  $n \equiv 0 \pmod{3}$  and  $n \equiv 1 \pmod{3}$ . Once again, the paths  $\mathcal{D}_0\{M_k\}$  are used as previously, and we obtain:

- If  $1 \leq f \leq \lceil n/2 \rceil - 2$ ,  $w_f(K_{n,n}^*) = 5f + 3$ .
- If  $\lceil n/2 \rceil - 1 \leq f \leq n - 2$ ,  $w_f(K_{n,n}^*) = 5f + 2$ .
- And, if  $f = n - 1$ ,  $w_f(K_{n,n}^*) = 5f + 1$ .

#### 4.7 Remaining particular cases

As previously noted, five cases were not treated in the previous study:  $n = 2$  because of the non-existence of a Latin square of order 2,  $n = 4$  and  $n = 6$  due to the exception to Tillson's Theorem and  $n = 1$  and  $n = 5$ , where the general techniques for  $n = 1 \bmod 3$  and  $n = 2 \bmod 3$  fail. For all of these cases, we provide routings that give, for every  $f$ ,  $0 \leq f \leq n - 1$ ,  $w_f(K_{n,n}^*) = \pi_f(K_{n,n}^*)$ . These cases illustrate the above-mentioned general method and are studied in an appendix, available at: <http://www.lirmm.fr/~bessy/publis.html>.

#### Acknowledgment

We thank Jean-Claude Bermond for introducing the problem to us and for motivating our research.

#### References

- [1] B. Beauquier, J.-C. Bermond, L. Gargano, P. Hell, S. Pérennes, and U. Vaccaro, Graph problems arising from wavelength-routing in all-optical networks, INRIA technical report 3165, May 1997 (presented at the Second IEEE Workshop on Optics and Computer Science).
- [2] B. Beauquier, P. Hell, and S. Pérennes, Optimal wavelength-routed multicasting, *Discrete Applied Mathematics* 84 (1998), 15–20
- [3] R. Diestel, *Graph theory*, Third edition, Graduate Texts in Mathematics, 173, Springer-Verlag, Berlin, 2005.
- [4] J.H. Dinitz, A.C.H. Ling, and D.R. Stinson, Fault tolerant routings with minimum optical index, *Networks* 48 (2006), 47–54.
- [5] A. Gupta, J. Mañuch, and L. Stacho, Fault tolerant forwarding and optical indexes: a design theory approach, *Journal of Combinatorial Designs* 14 (2006), 25–40.
- [6] C.C. Lindner, and C.A. Rodger, *Design theory*, CRC Press, Inc. Boca Raton, FL, 1997.

- [7] J. Maňuch, and L. Stacho, On  $f$ -wise arc forwarding index and wavelength allocations in faulty all-optical hypercubes, Theoretical Informatics and Applications 37 (2003), 255–270.
- [8] T.W. Tillson, A Hamiltonian decomposition of  $K_{2m}^*$ ,  $2m \geq 8$ , J. Combin. Theory Ser. B 2 (1980), 68–74.
- [9] V.G. Vizing, On an estimate of the chromatic class of a  $p$ -graph, Diskret. Analiz 3 (1964), 25–30.

## Appendix

### A Routing in $K_{1,1}^*$

We are in the case  $n \equiv 1 \pmod{3}$ , but the general decomposition is valid for  $n \geq 3$ . However, here, there is just one value of  $f$  to study:  $f = 0$  and the direct routing immediately gives  $w_0(K_{1,1}^*) = \pi_0(K_{1,1}^*) = 1$ .

### B Routing in $K_{2,2}^*$

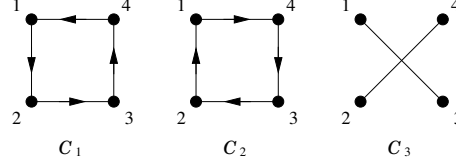
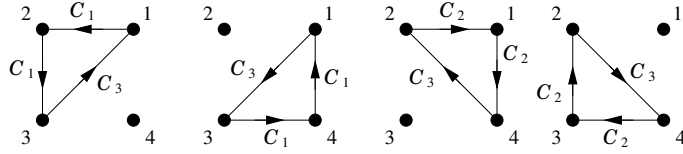
For  $n = 2$ , there is no idempotent Latin square of order 2, so we provide precisely the routing.

- For  $f = 0$ ,  $\mathcal{D}_0$  forms a color class, and we choose the set of pairwise arc-disjoint paths of length 2 given by  $\mathcal{S}_0 = \{x_1y_1x_2, x_2y_2x_1, y_1x_1y_2, y_2x_2y_1\}$  to form the second color class. So, we obtain  $w_0(K_{2,2}^*) = \pi_0(K_{2,2}^*) = 2$ .
- For  $f = 1$ , we define  $\mathcal{S}_1 = \{x_1y_2x_2, x_2y_1x_1, y_1x_2y_2, y_2x_1y_1\}$ , which contains pairwise arc-disjoint paths, and  $\mathcal{D}_1 = \{x_iy_{j+1}x_{i+1}y_j, y_ix_{j+1}y_{i+1}x_j : 1 \leq i \leq 2, 1 \leq j \leq 2\}$  (which corresponds to usual definition of  $\mathcal{D}_1$ ). The color classes are defined by:  $c_1 = \mathcal{S}_0$ ,  $c_2 = \mathcal{S}_1$  and  $c_{(i,j)} = \mathcal{D}_0\{x_iy_j, y_jx_i\} \cup \mathcal{D}_1\{x_iy_{j+1}x_{i+1}y_j, y_jx_{i+1}y_{j+1}x_i\}$  for  $1 \leq i \leq 2$  and  $1 \leq j \leq 2$ . Finally, we obtain  $w_1(K_{2,2}^*) = \pi_1(K_{2,2}^*) = 6$ .

### C Routing in $K_{4,4}^*$

The case  $n = 4$  is the first exception to Tillson's Theorem of decomposition, so we define in Figure 9 the partition of the arcs of  $K_4^*$ . Then, for the different values of  $f$ , the routings are defined as previously, in Section 4.1.

- For  $f = 0$ ,  $f = 1$ , and  $f = 2$ , since  $C_1$  and  $C_2$  are disjoint Hamiltonian circuits of  $K_4^*$ , the routings are defined exactly as in the general case  $n \equiv 1 \pmod{3}$ . So, in these cases, we obtain  $w_f(K_4^*) = \pi_f(K_4^*)$ .
- For  $f = 3$ , we need to pack the paths of  $\mathcal{R}_3$ . We differently organize the arcs of  $K_4^*$  in order to obtain a suitable decomposition of the paths. The scheme of the routing is given Figure 10. The arcs are labeled with the name of the circuit  $C_i$  they belong to.

Figure 9: The chosen decomposition of the arcs of  $K_4^*$ .Figure 10: New decomposition of the arcs of  $K_4^*$  in case  $f = 3$ .

Now, using these circuits, we route almost as in general case  $n \equiv 1 \pmod{3}$ . Using paths of some  $\mathcal{D}_k$  and  $\overline{\mathcal{S}_k[Y, Y]}$ , each one of these new circuits gives three color classes. The remaining paths, mainly from  $\mathcal{S}_k[X, X]$ , are packed together to form the four remaining colors.

More precisely, the first circuit provides the following classes:

$$c_1 = \mathcal{D}_1[Y, X]\{\star x_1 \star x_2\} \cup \mathcal{D}_1[X, Y]\{x_2 \star x_3 \star\} \cup \overline{\mathcal{S}_1[Y, Y]}\{\star x_4 \star\}$$

$$c_2 = \mathcal{D}_1[Y, X]\{\star x_2 \star x_3\} \cup \mathcal{D}_3[X, Y]\{x_3 \star x_1 \star\} \cup \overline{\mathcal{S}_2[Y, Y]}\{\star x_4 \star\}$$

$$c_3 = \mathcal{D}_3[Y, X]\{\star x_3 \star x_1\} \cup \mathcal{D}_1[X, Y]\{x_1 \star x_2 \star\} \cup \overline{\mathcal{S}_3[Y, Y]}\{\star x_4 \star\}$$

From the second circuit, we obtain:

$$c_4 = \mathcal{D}_3[Y, X]\{\star x_1 \star x_3\} \cup \mathcal{D}_1[X, Y]\{x_3 \star x_4 \star\} \cup \overline{\mathcal{S}_1[Y, Y]}\{\star x_2 \star\}$$

$$c_5 = \mathcal{D}_1[Y, X]\{\star x_3 \star x_4\} \cup \mathcal{D}_1[X, Y]\{x_4 \star x_1 \star\} \cup \overline{\mathcal{S}_2[Y, Y]}\{\star x_2 \star\}$$

$$c_6 = \mathcal{D}_1[Y, X]\{\star x_4 \star x_1\} \cup \mathcal{D}_3[X, Y]\{x_1 \star x_3 \star\} \cup \overline{\mathcal{S}_3[Y, Y]}\{\star x_2 \star\}$$

From the third circuit, we obtain:

$$c_7 = \mathcal{D}_2[Y, X]\{\star x_1 \star x_4\} \cup \mathcal{D}_3[X, Y]\{x_4 \star x_2 \star\} \cup \overline{\mathcal{S}_1[Y, Y]}\{\star x_3 \star\}$$

$$c_8 = \mathcal{D}_3[Y, X]\{\star x_4 \star x_2\} \cup \mathcal{D}_2[X, Y]\{x_2 \star x_1 \star\} \cup \overline{\mathcal{S}_2[Y, Y]}\{\star x_3 \star\}$$

$$c_9 = \mathcal{D}_2[Y, X]\{\star x_2 \star x_1\} \cup \mathcal{D}_2[X, Y]\{x_1 \star x_4 \star\} \cup \overline{\mathcal{S}_3[Y, Y]}\{\star x_3 \star\}$$

Finally, we obtain from the fourth circuit:

$$c_{10} = \mathcal{D}_3[Y, X]\{\star x_2 \star x_4\} \cup \mathcal{D}_2[X, Y]\{x_4 \star x_3 \star\} \cup \overline{\mathcal{S}_1[Y, Y]}\{\star x_1 \star\}$$

$$c_{11} = \mathcal{D}_2[Y, X]\{\star x_4 \star x_3\} \cup \mathcal{D}_2[X, Y]\{x_3 \star x_2 \star\} \cup \overline{\mathcal{S}_2[Y, Y]}\{\star x_1 \star\}$$

$$c_{12} = \mathcal{D}_2[Y, X]\{\star x_3 \star x_2\} \cup \mathcal{D}_3[X, Y]\{x_2 \star x_4 \star\} \cup \overline{\mathcal{S}_3[Y, Y]}\{\star x_1 \star\}$$

The next class uses  $\mathcal{S}_0[Y, Y]$  and the last matching of  $\mathcal{D}_0$ :

$$c_{13} = \overline{\mathcal{S}_0[Y, Y]}$$

To conclude, as previously in the case  $n \equiv 0 \pmod{3}$ , we pack the paths of  $\bigcup_{k=0}^3 \mathcal{S}_k[X, X]$  in three color classes:

$$c_{14} = \bigcup_{j=1}^4 \bigcup_{i=0}^3 \mathcal{S}_i[X, X]\{x_j \star x_{j+1}\}$$

$$c_{15} = \bigcup_{j=1}^4 \bigcup_{i=0}^3 \mathcal{S}_i[X, X]\{x_j \star x_{j+2}\}$$

$$c_{16} = \bigcup_{j=1}^4 \bigcup_{i=0}^3 \mathcal{S}_i[X, X]\{x_j \star x_{j+3}\}$$

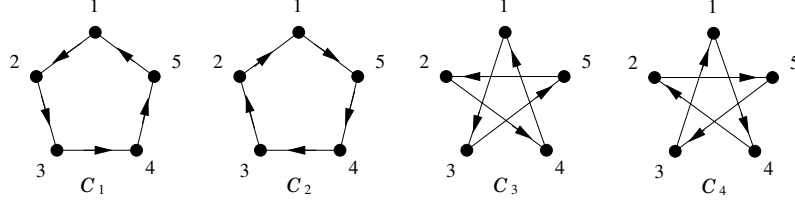
Finally, we obtain  $w_3(K_{4,4}^\star) = \pi_3(K_{4,4}^\star) = 16$ .

## D Routing in $K_{5,5}^\star$

We are in the case  $n \equiv 2 \pmod{3}$ , but the general decomposition is only valid for  $n \geq 6$ . Indeed, for every  $k$ , we need six sets of  $\overline{\mathcal{S}_k[Y, Y]}$  centered on six different vertices to complete the three color classes constructed from the paths of  $\mathcal{D}_k$ . In the case  $n = 5$ , we need to use paths from different sets  $\overline{\mathcal{S}_k[Y, Y]}$  to complete the color classes constructed from a single set  $\mathcal{D}_k$ .

To simplify the notation, we fix, in Figure 11, a decomposition of the arcs of  $K_5^\star$ . Using this decomposition we define the routing sets as previously, in Section 4.1.

Now, we detail how to pack the paths for all values of  $f$ .

Figure 11: An explicit decomposition of the arcs of  $K_{5,5}^*$ .

- For  $f = 0$ , as usual, we set  $c_0^1 = \mathcal{D}_0$ ,  $c_0^2 = \mathcal{S}_0[X, X]$  and  $c_0^3 = \mathcal{S}_0[Y, Y]$  and have  $w_0(K_{5,5}^*) = \pi_0(K_{5,5}^*) = 3$ .
- For  $f = 1$ , we use paths from  $\overline{\mathcal{S}_0[Y, Y]}$  and  $\overline{\mathcal{S}_1[Y, Y]}$  to complete the colors obtained with paths from  $\mathcal{D}_1$ . More precisely, we define for  $k = 1, \dots, 5$ :

$$c_1^k = \mathcal{D}_1[Y, X]\{\star x_k \star x_{k+1}\} \cup \mathcal{D}_1[X, Y]\{x_{k+1} \star x_{k+2} \star\} \cup \\ \overline{\mathcal{S}_0[Y, Y]}\{\star x_{k+3} \star\} \cup \overline{\mathcal{S}_1[Y, Y]}\{\star x_{k+4} \star\}$$

The remaining paths form the three color classes  $\mathcal{S}_0[X, X]$ ,  $\mathcal{S}_1[X, X]$  and  $\mathcal{D}_0\{M_0, M_1, M_2\}$  ( $\mathcal{D}_0\{M_3\}$  and  $\mathcal{D}_0\{M_4\}$  are respectively contained in  $\overline{\mathcal{S}_0[Y, Y]}$  and  $\overline{\mathcal{S}_1[Y, Y]}$ ). So, we obtain  $w_1(K_{5,5}^*) = \pi_1(K_{5,5}^*) = 8$ .

- For  $f = 2$ , the paths from  $\mathcal{D}_1$  are packed as previously: for  $k = 1, \dots, 5$  we use  $c_1^k$ . The paths from  $\mathcal{D}_2$  form color classes with paths from  $\overline{\mathcal{S}_0[X, X]}$  and  $\overline{\mathcal{S}_1[X, X]}$ . We define, for  $k = 1, \dots, 5$ :

$$c_2^k = \mathcal{D}_2[X, Y]\{\star y_k \star y_{k-1}\} \cup \mathcal{D}_2[Y, X]\{y_{k-1} \star y_{k-2} \star\} \cup \\ \overline{\mathcal{S}_0[X, X]}\{\star y_{k-3} \star\} \cup \overline{\mathcal{S}_1[X, X]}\{\star y_{k-4} \star\}$$

The remaining paths form the two color classes  $\overline{\mathcal{S}_2[X, X]}$  and  $\mathcal{S}_2[Y, Y]$  ( $\mathcal{D}_0\{M_0\}$ ,  $\mathcal{D}_0\{M_1\}$  and  $\mathcal{D}_0\{M_2\}$  are respectively contained in  $\overline{\mathcal{S}_0[X, X]}$ ,  $\overline{\mathcal{S}_1[X, X]}$  and  $\overline{\mathcal{S}_2[X, X]}$ ). So, we obtain  $w_2(K_{5,5}^*) = \pi_2(K_{5,5}^*) = 12$ .

- For  $f = 3$ , we use the same method to pack the paths from  $\mathcal{D}_3$  with paths from  $\mathcal{S}_2[Y, Y]$  and  $\mathcal{S}_3[Y, Y]$ . For the paths from  $\mathcal{D}_1$  and  $\mathcal{D}_2$ , we use  $c_1^1, \dots, c_1^5$  and  $c_2^1, \dots, c_2^5$  and, for

$k = 1, \dots, 5$ , we define:

$$c_3^k = \mathcal{D}_3[Y, X]\{\star x_k \star x_{k+2}\} \cup \mathcal{D}_3[X, Y]\{x_{k+2} \star x_{k+4} \star\} \cup \\ \mathcal{S}_2[Y, Y]\{\star x_{k+1} \star\} \cup \mathcal{S}_3[Y, Y]\{\star x_{k+3} \star\}$$

The remaining paths form the two color classes  $\overline{\mathcal{S}_2[X, X]}$  and  $\mathcal{S}_3[X, X]$ . So, we obtain  $w_3(K_{5,5}^\star) = \pi_3(K_{5,5}^\star) = 17$ .

- Finally, for  $f = 4$ , we have to change color classes. Indeed, this time, paths from  $\mathcal{D}_i$  are packed with sets of type  $(\bigcup_{p=0}^4 \mathcal{S}_p[X, X]\{x_l \star x_m\}) \cup (\bigcup_{p=0}^4 \mathcal{S}_p[X, X]\{x_m \star x_l\})$  which saturate exactly the in and out-neighborhood of the vertices  $x_l$  and  $x_m$ . Precisely, we define, for  $k = 1, \dots, 5$ :

$$c_1^k = \mathcal{D}_1[Y, X]\{\star x_k \star x_{k+1}\} \cup \mathcal{D}_1[X, Y]\{x_{k+1} \star x_{k+2} \star\} \cup \\ (\bigcup_{p=0}^4 \mathcal{S}_p[X, X]\{x_{k+3} \star x_{k+4}\}) \cup (\bigcup_{p=0}^4 \mathcal{S}_p[X, X]\{x_{k+4} \star x_{k+3}\})$$

$$c_2^k = \mathcal{D}_2[X, Y]\{\star y_k \star y_{k-1}\} \cup \mathcal{D}_2[Y, X]\{y_{k-1} \star y_{k-2} \star\} \cup \\ (\bigcup_{p=0}^4 \mathcal{S}_p[Y, Y]\{y_{k-3} \star y_{k-4}\}) \cup (\bigcup_{p=0}^4 \mathcal{S}_p[Y, Y]\{y_{k-4} \star y_{k-3}\})$$

$$c_3^k = \mathcal{D}_3[Y, X]\{\star x_k \star x_{k+2}\} \cup \mathcal{D}_3[X, Y]\{x_{k+2} \star x_{k+4} \star\} \cup \\ (\bigcup_{p=0}^4 \mathcal{S}_p[X, X]\{x_{k+1} \star x_{k+3}\}) \cup (\bigcup_{p=0}^4 \mathcal{S}_p[X, X]\{x_{k+3} \star x_{k+1}\})$$

$$c_4^k = \mathcal{D}_4[X, Y]\{\star y_k \star y_{k-2}\} \cup \mathcal{D}_4[Y, X]\{y_{k-2} \star y_{k-4} \star\} \cup \\ (\bigcup_{p=0}^4 \mathcal{S}_p[Y, Y]\{y_{k-1} \star y_{k-3}\}) \cup (\bigcup_{p=0}^4 \mathcal{S}_p[Y, Y]\{y_{k-3} \star y_{k-1}\})$$

The remaining paths are exactly the direct paths,  $\mathcal{D}_0$ . So, we obtain  $w_4(K_{5,5}^\star) = \pi_4(K_{5,5}^\star) = 21$ .



### E Routing in $K_{6,6}^*$

The case  $n = 6$  is the second exception to Tillson's Theorem of decomposition, so we define in Figure 12 the partition of the arcs of  $K_6^*$ . Then, for the different values of  $f$ , the routings are defined as previously, in Section 4.1.

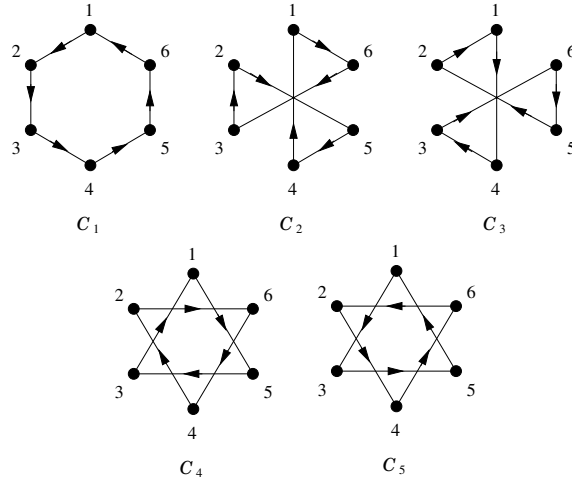


Figure 12: The chosen decomposition of the arcs of  $K_6^*$ .

As all the circuits involved in this decomposition have length 3 or 6, the computation of  $w_f$  works exactly as in general case  $n \equiv 0 \pmod{3}$ . Finally, we obtain, for all  $f = 0, \dots, 5$ ,  $w_f(K_{6,6}^*) = \pi_f(K_{6,6}^*)$ , which completes the proof of Theorem 4.

## 2.6 A proof of Lehel's Conjecture

### Partitioning a graph into a cycle and an anticyle, a proof of Lehel's conjecture

Stéphane Bessy and Stéphan Thomassé\*

#### Abstract

We prove that every graph  $G$  has a vertex partition into a cycle and an anticyle (a cycle in the complement of  $G$ ). Emptyset, singletons and edges are considered as cycles. This problem was posed by Lehel and shown to be true for very large graphs by Łuczak, Rödl and Szemerédi [7], and more recently for large graphs by Allen [1].

Many questions deal with the existence of monochromatic paths and cycles in edge-colored complete graphs. Erdős, Gyárfás and Pyber asked for instance in [3] if every coloring with  $k$  colors of the edges of a complete graph admits a vertex partition into  $k$  monochromatic cycles. In a recent paper, Gyárfás, Ruszinkó, Sárközy and Szemerédi [5] proved that  $O(k \log k)$  cycles suffice to partition the vertices. This question was also studied for other structures like complete bipartite graphs by Haxell [6]. One case which received a particular attention was the case  $k = 2$ , where one would like to cover a complete graph which edges are colored blue and red by two monochromatic cycles. A conjecture of Lehel, first cited in [2], asserts that a blue and a red cycle partition the vertices, where emptyset, singletons and edges are considered as cycles. This was proved for sufficiently large  $n$  by Łuczak, Rödl and Szemerédi [7], and more recently by Allen [1] with a better bound. Our goal is to completely answer Lehel's conjecture.

Our starting point is the proof of Gyárfás of the existence of two such cycles covering the vertices and intersecting on at most one vertex (see [4]). For this, he considered a longest path consisting of a red path followed by a blue path. The nice fact is that such a path  $P$  is hamiltonian. Indeed, if a vertex  $v$  is not covered, it must be joined in blue to the origin  $a$  of  $P$  and in red to the end  $b$  of  $P$ . But then, one can cover the vertices of  $P$  and  $v$  using the edge  $ab$ . Consequently, there exists a hamiltonian cycle consisting of two monochromatic paths. Hence, there exists a monochromatic cycle  $C$ , of size at least two, and a monochromatic path  $P$  with different colors partitioning the vertex set. This is the key-structure for the proof of our main result:

**Theorem 1** *Every complete graph with red and blue edges has a vertex partition into a red cycle and a blue cycle.*

**Proof.** Assume that  $C$  and  $P$  are chosen as above in such a way that  $C$  has maximum size and has color, say, blue. We will show that we can either increase the length of  $C$  or prove the existence of our two cycles. If  $P$  has less than three vertices, we are done. We denote by  $x$  and  $y$  the endvertices of  $P$ . Note that if  $x$  and  $y$  are joined by a red edge, we have our two cycles. We then assume that  $xy$  is a blue edge. A vertex of  $C$  is *red* if it is joined to both  $x$  and  $y$  by red edges. The other vertices of  $C$  are *blue*. Observe that  $C$  cannot have two consecutive blue vertices, otherwise we would extend  $C$ . Moreover, the two neighbors in  $C$  of a red vertex  $v$  cannot be joined by a blue edge, since we could add  $v$  to the path  $P$  to form a red cycle. Similarly, if  $C$  has two or three vertices, one of them is red and could be added to  $P$  to form a red cycle. So, in particular,  $C$  has at least four vertices. Observe also that  $|C| > |P|$  since we

---

\*Université Montpellier 2 - CNRS, LIRMM, 161 rue Ada, 34392 Montpellier Cedex 5, France, [bessy,thomasse@lirmm.fr](mailto:bessy,thomasse@lirmm.fr)

could give a red vertex to  $P$  to form a better partition into a cycle and a path. In this proof, removing a vertex  $x$  from a path or a cycle  $Q$  is denoted by  $Q \setminus x$ , whereas removing an edge  $xy$  is denoted by  $Q - xy$ .

**Claim 1** *There are no successive blue, red, blue, red, blue vertices in  $C$ .*

**Proof.** Assume that  $b, r, b', r', b''$  is such a sequence (with possibly  $b = b''$ ) and that, say,  $xb'$  is a blue edge. Let  $x'$  be the successor of  $x$  on the red path  $P$ . We claim that  $x'r$  is a blue edge. Otherwise, either  $bx$  is a blue edge in which case  $(C \setminus r) \cup x$  is a blue cycle and  $(P \setminus x) \cup r$  is a red cycle, or  $by$  is a blue edge and then  $(P \setminus \{x, y\}) \cup r$  is a red path and  $(C \setminus r) \cup \{x, y\}$  is a blue cycle longer than  $C$ , a contradiction. Similarly,  $x'r'$  is a blue edge. Since  $b''$  is blue, there exists a blue path  $P' = b'xb''$  or  $P' = b'xyb''$ . Replacing the path  $brb'r'b''$  in  $C$  by  $bx'r'b'P'b''$  would increase the length of  $C$ , a contradiction. ■

When  $acb$  are consecutive vertices of  $C$  and  $c$  is a red vertex, we call  $ab$  a *special edge*. Observe that special edges are red. We denote by  $G_s$  the graph on the same vertex set as  $C$  whose edges are the special edges. Observe that the maximum degree of  $G_s$  is two. It appears that the proof is easier if we have several blue vertices in  $C$ . Let us prove for a start that there exists at least one.

**Claim 2** *There exists a blue vertex in  $C$ .*

**Proof.** If not,  $G_s$  is either a cycle or the union of two cycles, depending if  $C$  has an odd or an even number of vertices. If  $C$  contains a red hamiltonian path, we can form, with  $P$ , a hamiltonian red cycle of the whole graph. Therefore  $G_s$  is the union of two red cycles  $W$  and  $Z$ , alternating along  $C$ , with the same cardinality and no red edge between them. We denote by  $x'$  and  $y'$  the respective neighbors of  $x$  and  $y$  in  $P$ , with possibly  $x' = y'$  if  $|P| = 3$ . There is no red edge from  $x'$  to  $W$ , otherwise,  $(P \setminus x)WxZ$  forms a hamiltonian red cycle. Similarly, there is no red edge from  $x'$  to  $Z$  or from  $y'$  to  $W \cup Z$ .

- If  $|P| = 3$ , then  $W \cup x \cup Z \cup y$  is spanned by a red cycle and  $x'$  forms a blue one.
- If  $|P| = 4$ , then pick a vertex  $w$  of  $W$  and a vertex  $z$  of  $Z$ , consecutive along  $C$ , and form a red cycle  $wxyz$ . To conclude, partition the remaining blue path of  $C$  into two subpaths  $P'$  and  $P''$  and form the blue cycle  $x'P'y'P''$ .

Now we assume that  $P$  has at least five vertices. We denote by  $x''$  and  $y''$  the respective neighbors of  $x'$  and  $y'$  on  $P$ . There is no red edge from  $x''$  to  $W$ , otherwise  $(P \setminus \{x', x''\})WxZ$  forms a red cycle and  $x'$  forms a blue one. Similarly, there are all blue edges from  $x''$  to  $Z$  and from  $y''$  to  $W \cup Z$ . Observe that  $xx''$  is a blue edge, otherwise  $P \setminus x'$  forms a red path and  $C \cup x'$  is spanned by a blue cycle longer than  $C$ , a contradiction. Similarly,  $yy''$  is a blue edge.

- Assume that  $|P| = 5$ , in particular  $y'' = x''$  and  $|C| \geq |P| + 1 = 6$ . Pick a vertex  $w$  in  $W$  and a vertex  $z$  in  $Z$ . Form the red cycle  $wxyz$ , a blue cycle covering the blue bipartite graph  $(W \setminus w) \cup (Z \setminus z)$  and finally insert in this blue cycle, of length more than three, the vertices  $x', y'$  and  $x''$ .
- If  $|P| \geq 6$ , we insert the three blue paths  $x', y'$  and  $x''xyy''$  in  $C$  to form a blue cycle longer than  $C$ , a contradiction. ■

Now, fix an orientation of the cycle  $C$ . We define the set  $L$  of *left* vertices as the vertices which are left neighbors in  $C$  of some blue vertex. We define similarly the set  $R$  of *right* vertices. Note that  $L$  and  $R$  are not empty, may intersect and contain only red vertices.

**Claim 3** *The set of left (resp. right) vertices spans a red clique.*

**Proof.** Assume for contradiction that there exists a blue edge joining two left red vertices  $u$  and  $v$ . We denote by  $u'$  and  $v'$  their respective right blue neighbors in  $C$ . There exists a path  $Q$  from  $u'$  to  $v'$  in  $\{u', x, y, v'\}$  with length at least two. Now  $(C - \{uu', vv'\}) \cup Q \cup uv$  is a blue cycle which is longer than  $C$ , a contradiction. ■

Every connected component of  $G_s$  which is a path has an endvertex in  $L$  and the other endvertex in  $R$ . Furthermore, if  $G_s$  has a cycle  $Z$ , it is unique and it contains all the blue vertices of  $C$ . In this case,  $Z$  cannot contain all the vertices of  $C$ , for instance, the neighbor of a blue vertex in  $C$  does not belong to  $Z$ . Indeed, if it exists,  $Z$  is simply obtained by taking all the vertices of  $C$  at even distance of some blue vertex, so  $Z$  contains every other vertex on  $C$ , and  $|C|$  is even. Hence, the vertices of the whole graph are partitioned into a red clique  $L$ , a red clique  $R$ , a set  $S = C \setminus (R \cup L \cup Z)$  which is covered by a set  $\mathcal{S}$  of  $|R|$  disjoint  $RL$  paths of red edges, the original path  $P$ , and (possibly) the cycle  $Z$ .

**Claim 4** *There exists a red path which spans  $S \cup R \cup L$ . More precisely:*

- If  $|\mathcal{S}|$  is even, for all distinct vertices of  $R$  (resp.  $L$ )  $x$  and  $y$  there is a red path from  $x$  to  $y$  which spans  $S \cup R \cup L$ .
- If  $|\mathcal{S}|$  is odd, then for all  $x \in R$  and  $y \in L$  such that  $x$  and  $y$  are not the endvertices of a same path of  $\mathcal{S}$ , there is a red path from  $x$  to  $y$  which spans  $S \cup R \cup L$ .

**Proof.** We give a constructive proof. Denote by  $P_x$  (resp.  $P_y$ ) the path of  $\mathcal{S}$  which contains  $x$  (resp.  $y$ ). Starting from  $x$ , we follow the path  $P_x$  until its end. At the end of a path of  $\mathcal{S}$ ,  $R$  and  $L$  being red cliques, we go to the beginning of a unvisited path of  $\mathcal{S}$ , which is not  $P_y$ , and follow it. When the process stops, using a red edge of  $R$  or  $L$ , we go to the endvertices of  $P_y$ , which is not  $y$  (because of the parity of  $|\mathcal{S}|$ , and terminate the spanning path on  $y$ . ■

A direct corollary of Claim 4 is that if  $Z$  does not exist, one can cover  $C$ , whose vertices are exactly  $S \cup R \cup L$ , by a red path  $P'$  ending in two red vertices. Hence  $P \cup P'$  forms a red hamiltonian cycle. Thus we can assume that  $Z$  exists. Observe that by Claim 1, every blue vertex of  $Z$  is the neighbor in  $Z$  of a red vertex.

**Claim 5** *Every blue vertex is joined in blue to  $R \cup L$ .*

**Proof.** Indeed, assume for contradiction that  $br$  is a red edge where  $b$  is a blue vertex and  $r$  belongs to  $R$ . Let  $z$  be a red vertex which is consecutive to  $b$  in  $Z$ . By Claim 4, there exists a red path  $P'$  starting at  $r$ , covering  $S \cup R \cup L$ , and terminating on a red vertex of  $C$ . Now,  $(Z - zb) \cup P \cup P'$  forms a hamiltonian red cycle. ■

**Claim 6** *There is a red cycle  $W$  spanning  $S \cup R \cup L$ .*

**Proof.** If  $|\mathcal{S}|$  is even, then Claim 4 directly gives the result. So, we assume that  $|\mathcal{S}|$  is odd. If there is a unique blue vertex  $b$  in  $C$ , the graph  $G_s$  consists of the union of  $Z$  and a unique path  $P'$  ( $|\mathcal{S}| = 1$ ) whose endvertices  $u$  and  $v$  are the neighbors of  $b$  in  $C$ . If  $uv$  is a red edge, we are done. So, assume that  $uv$  is a blue edge, in particular  $|C| > 4$  otherwise  $uv$  would be a (red) edge of  $G_s$ . Denote by  $u'$  the second neighbor of  $u$  in  $C$  and by  $u''$  the second neighbor of  $u'$  in  $C$  (and thus the successor of  $u$  in  $P'$ ). If  $bu''$  is a blue edge, then replacing in  $C$  the path  $vbuu'u''$  by  $vubu''$  forms a blue cycle and  $P \cup u'$  forms a red cycle. Thus  $bu''$  is a red edge, in which case we form a red cycle  $(Z - bu') \cup (P' \setminus u) \cup P$  and the singleton  $u$  as a blue cycle.

Assume now that  $C$  has at least two blue vertices. As  $|\mathcal{S}|$  is odd, by Claim 4, we just have to prove that there exists a red edge between a vertex of  $R$  and a vertex of  $L$  which are not the endvertices of the same path of  $\mathcal{S}$ . For this, we consider a subpath  $I$  of  $C$  containing two blue vertices forming the

endvertices of  $I$ . By Claim 1 and the fact that there exists at least two blue vertices, there is such an  $I = br_1 \dots r_k b'$ , where  $k > 1$  and  $r_1, \dots, r_k$  are red vertices. By Claim 5,  $br_k$  and  $b'r_1$  are blue edges, we can replace  $I$  by  $br_k r_{k-1} \dots r_1 b'$ . Hence,  $r_1$  becomes a left vertex. Thus by Claim 3,  $r_1$  is joined in red to all the vertices of  $R \cup L$ , except possibly  $r_k$ . ■

Now, if a blue vertex is joined in red to any vertex of  $S \cup R \cup L$ , we can conclude as in Claim 5.

**Claim 7** *There is no red edge between  $W$  and  $Z$ .*

**Proof.** Assume that  $zw$  is a red edge with  $z \in Z$  and  $w \in W$ . Let  $z'$  be the first vertex to the right of  $z$  in  $Z$  which is joined to  $W$  with at least a red edge (here  $z'$  can be  $z$ ). By the above remark,  $z'$  is a red vertex. Let  $A$  be the set of vertices between  $z$  and  $z'$  in  $Z$ . Let  $B$  be the set of  $|A|$  consecutive vertices to the right of  $w$  on the cycle  $W$  (recall that  $Z$  and  $W$  have the same size, hence  $B$  does not contain  $w$ ). Now,  $A \cup B$  is a complete bipartite blue graph hence it has a blue spanning cycle. Moreover,  $(Z \setminus A) \cup (W \setminus B)$  is spanned by a path  $P'$  starting at  $z'$  and ending in  $W$ . Both endvertices of  $P'$  are red, thus  $P \cup P'$  forms a red cycle. ■

We now achieve the proof of the theorem. Let  $W$  be the red cycle  $w_1 \dots w_k$  and  $Z$  be the red cycle  $z_1 \dots z_k$ , where  $z_1$  is a blue vertex such that, say, the edge  $xz_1$  is blue. Denote by  $y'$  the neighbor of  $y$  on  $P$ . There is no red edge between  $y'$  and a vertex of  $Z$ . Otherwise, letting  $z$  be this vertex and  $Q$  be a minimal path in  $Z$  from  $z$  to a red vertex of  $Z$  ( $Q$  has length zero or one). Denote by  $Q'$  a red path of  $W$  with same length as  $Q$ . Then,  $P \cup Q \cup Q'$  is spanned by a red cycle (by inserting  $Q$  between  $y'$  and  $y$  and inserting  $Q'$  between  $x$  and  $y$ ) and  $C \setminus (Q \cup Q')$  is spanned by a blue one. Similarly, there is no red edge between  $y'$  and  $W$ , otherwise denote by  $w$  a red neighbor of  $y'$  on  $W$  and by  $z$  a red vertex on  $Z$ . Then,  $P \cup w \cup z$  is spanned by a red cycle and  $(Z \setminus z) \cup (W \setminus w)$  is spanned by a blue one. Hence  $y'$  is linked in blue to  $W \cup Z$ .

If  $|P| = 3$ , we choose two red vertices  $z$  and  $w$  respectively in  $Z$  and in  $W$ . Now,  $xzyw$  is a red cycle, and  $((W \cup Z) \setminus \{w, z\}) \cup y'$  is spanned by a blue cycle. Finally, if  $|P| \geq 4$ , we denote by  $y''$  the second neighbor of  $y'$  on  $P$ . The edge  $yy''$  is a blue one, otherwise,  $P \setminus y'$  is a red path and  $C \cup y'$  is spanned by a blue cycle longer than  $C$ . The edge  $y''w_1$  is a blue one, otherwise for any red vertex  $z$  of  $Z$ , we would span  $(P \setminus y') \cup w_1 \cup z$  by a red cycle and  $(C \setminus \{w_1, z\}) \cup y'$  by a blue one. Now, starting with any blue cycle covering  $W \cup Z$  which contains the subpath  $z_1, w_1, z_2$ , we replace this path by  $z_1, x, y, y'', w_1, y', z_2$ , a contradiction to the maximality of  $C$ . ■

We would like to thank P. Allen and A. Gyárfás for stimulating discussions.

## References

- [1] P. Allen, Covering two-edge-coloured complete graphs with two disjoint monochromatic cycles, *Combinatorics, Probability and Computing*, **17** (2008), 471–486.
- [2] J. Ayel, Sur l'existence de deux cycles supplémentaires unicolores, disjoints et de couleurs différentes dans un graphe complet bicolore, *Thèse de l'université de Grenoble* (1979).
- [3] P. Erdős, A. Gyárfás and L. Pyber, Vertex coverings by monochromatic cycles and trees, *J. Combin. Theory Ser. B*, **51** (1991), 90–95.
- [4] A. Gyárfás, Vertex coverings by monochromatic paths and cycles, *Journal of Graph Theory*, **7** (1983), 131–135.

- [5] A. Gyárfás, M. Ruszinkó, G.N. Sárközy and E. Szemerédi, An improved bound for the monochromatic cycle partition number, *J. Combin. Theory Ser. B*, **96** (2006), 855–873.
- [6] P.E. Haxell, Partitioning complete bipartite graphs by monochromatic cycles, *J. Combin. Theory Ser. B*, **69** (1997), 210–218.
- [7] T. Łuczak, V. Rödl and E. Szemerédi, Partitioning two-colored complete graphs into two monochromatic cycles, *Combinatorics, Probability and Computing*, **7** (1998), 423–436.

## 2.7 A poly-kernel for 3-leaf power graph modification

### Polynomial kernels for 3-leaf power graph modification problems<sup>☆</sup>

Stéphane Bessy<sup>b</sup>, Christophe Paul<sup>a</sup>, Anthony Perez<sup>b</sup>

<sup>a</sup>*CNRS - LIRMM, Université Montpellier 2, 161 rue Ada  
34932 Montpellier, France*

<sup>b</sup>*LIRMM - Université Montpellier 2, 161 rue Ada  
34932 Montpellier, France*

---

#### Abstract

A graph  $G = (V, E)$  is a 3-leaf power iff there exists a tree  $T$  the leaf set of which is  $V$  and such that  $uv \in E$  iff  $u$  and  $v$  are at distance at most 3 in  $T$ . The 3-leaf power graph edge modification problems, *i.e.* edition (also known as the CLOSEST 3-LEAF POWER), completion and edge-deletion are FPT when parameterized by the size of the edge set modification. However, polynomial kernels were known for none of these three problems. For each of them, we provide kernels with  $O(k^3)$  vertices that can be computed in linear time. We thereby answer an open problem first mentioned by Dom, Guo, Hüffner and Niedermeier [9].

*Key words:* Algorithms, data-structures, FPT, kernel, graph modification problems, leaf power

---

#### Introduction

The combinatorial analysis of experimental data-sets naturally leads to graph modification problems. For example, extracting a threshold graph from a dissimilarity on a set is a classical technique used in clustering and data analysis to move from a numerical to a combinatorial data-set [1, 16]. The edge set of the threshold graph aims at representing the pairs of elements which are close to each another. As the dissimilarity reflects some experimental measures, the edge set of the threshold graph may reflect some false positive or negative errors. So for the sake of cluster identification, the edge set of the threshold graph has to be edited in order to obtain a disjoint union of cliques. This problem, known as CLUSTER EDITING, is fixed-parameter tractable (see e.g. [12, 13, 25]) and efficient parameterized algorithms have been proposed to solve biological instances with about 1000 vertices and several thousand edge modifications [2, 7]. So, motivated

---

<sup>☆</sup>Work supported by the French research grant ANR-06-BLAN-0148-01 "Graph Decomposition and Algorithms - GRAAL".

*Email addresses:* `bessy@lirmm.fr` (Stéphane Bessy), `paul@lirmm.fr` (Christophe Paul), `perez@lirmm.fr` (Anthony Perez)

by the identification of some hidden combinatorial structures on experimental data-sets, edge-modification problems cover a broad range of classical graph optimization problems, among which *completion* problems, *edition* problems and *edge-deletion* problems (see [19] for a recent survey). Precisely, for a given graph  $G = (V, E)$ , in a completion problems, the set  $F$  of modified edges is constrained to be disjoint from  $E$ , whereas in an edge deletion problems  $F$  has to be a subset of  $E$ . If no restriction applies to  $F$ , then we obtain an edition problem. Though most of the edge-modification problems turn out to be NP-hard problems, efficient algorithms can be obtained to solve the natural parameterized version of some of them. Indeed, as long as the number  $k$  of errors generated by the experimental process is not too large, one can afford a time complexity exponential in  $k$ . A problem is *fixed parameterized tractable* (FPT for short) with respect to parameter  $k$  whenever it can be solved in time  $f(k) \cdot n^{O(1)}$ , where  $f(k)$  is an arbitrary computable function. Here, the natural parameterization is the number  $k = |F|$  of modified edges. The generic question is thereby whether for fixed  $k$ , a given edge modification problem is tractable. More formally:

PARAMETERIZED  $\Pi$ -MODIFICATION PROBLEM

**Input:** An undirected graph  $G = (V, E)$ .

**Parameter:** An integer  $k \geq 0$ .

**Question:** Is there a subset  $F \subseteq V \times V$  with  $|F| \leq k$  such that the graph  $G + F = (V, E \triangle F)$  satisfies  $\Pi$ .

This paper studies the parameterized version of edge modification problems and more precisely the existence of a polynomial *kernel*. A problem is kernelizable if every instance  $(G, k)$  can be reduced in polynomial time (using *reduction rules*) into an instance  $(G', k')$  such that  $k' \leq k$  and the size of  $G'$  is bounded by a function of  $k$ . The membership to the FPT complexity class is equivalent to the property of having a kernel (see [20] for example). Having a kernel of small size is clearly highly desirable [15]. Indeed, preprocessing the input in order to reduce its size while preserving the existence of a solution is an important issue in the context of various applications ([15]). However, the equivalence mentioned above only provides an exponential bound on the kernel size. For a parameterized problem, the challenge is then to know whether it admits or not a polynomial - or even linear (in  $k$ ) - kernel (see e.g. [20]). The  $k$ -VERTEX COVER problem is the classical example of a problem with a linear kernel. Recently, parameterized problems (among which the  $k$ -TREEWIDTH problem) have been shown to not have polynomial kernels [3] (unless some collapse occurs in the computational complexity hierarchy).

This paper follows this line of research and studies the kernelization of edge-modification problems related to the family of *leaf powers*, graphs arising from a phylogenetic reconstruction context [17, 18, 21]. The goal is to extract, from a threshold graph  $G$  on a set  $S$  of species, a tree  $T$ , whose leaf set is  $S$  and such that the distance between two species is at most  $p$  in  $T$  iff they are adjacent in  $G$  ( $p$  being the value used to extract  $G$  from dissimilarity information). If such a tree  $T$  exists, then  $G$  is a  $p$ -leaf power and  $T$  is its  $p$ -leaf root. For  $p \leq 5$ , the  $p$ -leaf power recognition is polynomial time solvable [5, 6], whereas the question is still open for  $p$  strictly larger than 5. Parameterized  $p$ -leaf power



edge modification problems have been studied so far for  $p \leq 4$ . The edition problem for  $p = 2$  is known as the CLUSTER EDITING problem for which the kernel size bound has been successively improved in a series of recent papers [12, 13, 24], culminating in [14] with a kernel with  $4k$  vertices. For larger values of  $p$ , the edition problem is known as the CLOSEST  $p$ -LEAF POWER problem. For  $p = 3$  and 4, the CLOSEST  $p$ -LEAF POWER problem is known to be FPT [10, 9], while its fixed parameterized tractability is still open for larger values of  $p$ . However, the existence of a polynomial kernel for  $p \neq 2$  remained an open question [8, 11]. Though the completion and edge-deletion problems are FPT for  $p \leq 4$  [9, 11], no polynomial kernel is known for  $p \neq 2$  [14].

**Our results.** We prove that the CLOSEST 3-LEAF POWER, the 3-LEAF POWER COMPLETION and the 3-LEAF POWER EDGE-DELETION admit a kernel with  $O(k^3)$  vertices. We thereby answer positively to the open question of Dom, Guo, Hüffner and Niedermeier [11, 9].

**Outlines.** First section is dedicated to some known and new structural results of 3-leaf powers and their related critical clique graphs. Section 2 describes the data-reduction rules for the CLOSEST 3-LEAF POWER problem. The kernels for the other two variants, the 3-LEAF POWER COMPLETION and the 3-LEAF POWER EDGE-DELETION problems, are presented in Section 3.

## 1. Preliminaries

The graphs we consider in this paper are undirected and loopless. The vertex set of a graph  $G$  is denoted by  $V(G)$ , with  $|V(G)| = n$ , and its edge set by  $E(G)$ , with  $|E(G)| = m$  (or  $V$  and  $E$  when the context is clear). The *open* neighborhood of a vertex  $x$  is denoted by  $N_G(x)$  (or  $N(x)$ ), and its *closed* neighborhood (i.e.  $N_G(x) \cup \{x\}$ ) by  $N[x]$ . Two vertices  $x$  and  $y$  of  $G$  are *true twins* if they are adjacent and  $N(x) = N(y)$ . A subset  $S$  of vertices is a *module* if for every distinct vertices  $x$  and  $y$  of  $S$ ,  $N(x) \setminus S = N(y) \setminus S$ . Given a subset  $S$  of vertices,  $G[S]$  denotes the subgraph of  $G$  induced by  $S$ . If  $H$  is a subgraph of  $G$ ,  $G \setminus H$  stands for  $G[V(G) \setminus V(H)]$ . We write  $d_G(u, v)$  the distance between two vertices  $u$  and  $v$  in  $G$ . For a subset  $S \subseteq V$ ,  $d_S(u, v)$  denotes the distance between  $u$  and  $v$  within  $G[S]$ , and is set to  $\infty$  if  $u$  and  $v$  are not connected in  $G[S]$ . A graph family  $\mathcal{F}$  is *hereditary* if for every graph  $G \in \mathcal{F}$ , every induced subgraph  $H$  of  $G$  also belongs to  $\mathcal{F}$ . For a set  $\mathcal{S}$  of graphs, we say that  $G$  is  $\mathcal{S}$ -free if none of the graphs of  $\mathcal{S}$  is an induced subgraph of  $G$ .

As the paper deals with undirected graphs, we abusively denote by  $X \times Y$  the set of unordered pairs containing one element of  $X$  and one of  $Y$ . Let  $G = (V, E)$  be a graph and  $F$  be a subset of  $V \times V$ , we denote by  $G + F$  the graph on vertex set  $V$ , the edge set of which is  $E \Delta F$  (the symmetric difference between  $E$  and  $F$ ). Such a set  $F$  is called an *edition* of  $G$  (we may also abusively say that  $G + F$  is an edition). We improperly speak about edges of  $F$ , even if the elements of  $F$  are not all edges of  $G$ . A vertex  $v \in V$  is *affected* by an edition  $F$  whenever  $F$  contains an edge incident to  $v$ . Given a graph family  $\mathcal{F}$  and given a graph  $G = (V, E)$ , a subset  $F \subseteq V \times V$  is an *optimal  $\mathcal{F}$ -edition* of  $G$  if  $F$  is a set of minimum cardinality such that  $G + F \in \mathcal{F}$ . If we constrain  $F$  to be

disjoint from  $E$ , we say that  $F$  is a *completion*, whereas if  $F$  is asked to be a subset of  $E$ , then  $F$  is an *edge deletion*.

### 1.1. Critical cliques

The notions of critical clique and critical clique graph have been introduced in [18] in the context of phylogenetic. More recently, they have been successfully used in various modification problems such as CLUSTER EDITING [14] and BICLUSTER EDITING [24].

**Definition 1.1.** A critical clique of a graph  $G$  is a clique  $K$  which is a module and is maximal under this property.

It follows from definition that the set  $\mathcal{K}(G)$  of critical cliques of a graph  $G$  defines a partition of its vertex set  $V$ .

**Definition 1.2.** Given a graph  $G = (V, E)$ , its critical clique graph  $\mathcal{C}(G)$  has vertex set  $\mathcal{K}(G)$  and edge set  $E(\mathcal{C}(G))$  with

$$KK' \in E(\mathcal{C}(G)) \Leftrightarrow \forall v \in K, v' \in K', vv' \in E(G)$$

Let us note that given a graph  $G$ , its critical clique graph  $\mathcal{C}(G)$  can be computed in linear time with modular decomposition algorithm (see [26] for example).

The following lemma was used in the construction of polynomial kernels for CLUSTER EDITING and BICLUSTER EDITING problems in [24].

**Lemma 1.3.** Let  $G = (V, E)$  be a graph. If  $H$  is the graph  $G + \{(u, v)\}$  with  $(u, v) \in V \times V$ , then  $|\mathcal{K}(H)| \leq |\mathcal{K}(G)| + 4$ .

The next lemma shows that for a range of graph families, critical cliques are robust under optimal edition.

**Lemma 1.4.** Let  $\mathcal{F}$  be an hereditary graph family closed under true twin addition. For every graph  $G = (V, E)$ , there exists an optimal  $\mathcal{F}$ -edition (resp.  $\mathcal{F}$ -deletion,  $\mathcal{F}$ -completion)  $F$  such that every critical clique of  $G + F$  is the disjoint union of a subset of critical cliques of  $G$ .

**Proof.** We prove the statement for the edition problem. Same arguments apply for edge deletion and edge completion problems.

Let  $F$  be an optimal  $\mathcal{F}$ -edition of  $G$  such that the number  $i$  of critical cliques of  $G$  which are clique modules in  $H = G + F$  is maximum. Denote  $\mathcal{K}(G) = \{K_1, \dots, K_c\}$  and assume that  $i < c$  (i.e.  $K_1, \dots, K_i$  are clique modules in  $H$  and  $K_{i+1}, \dots, K_c$  are no longer clique modules in  $H$ ). Let  $x$  be a vertex of  $K_{i+1}$  such that the number of edges of  $F$  incident to  $x$  is minimum among the vertices of  $K_{i+1}$ . Roughly speaking, we will modify  $F$  by editing all vertices of  $K_{i+1}$  like  $x$ . Let  $H_x$  be the subgraph  $H \setminus (K_{i+1} \setminus \{x\})$ . As  $\mathcal{F}$  is hereditary,  $H_x$  belongs to  $\mathcal{F}$  and, as  $\mathcal{F}$  is closed under true twin addition, reinserting  $|K_{i+1}| - 1$  true twins of  $x$  in  $H_x$  results in a graph  $H'$  belonging to  $\mathcal{F}$ . It follows that

$F' = E(G) \Delta E(H')$  is an  $\mathcal{F}$ -edition of  $G$ . By the choice of  $x$ , we have  $|F'| \leq |F|$ . Finally let us remark that, now,  $K_1, \dots, K_i$  and  $K_{i+1}$  are clique modules of  $H'$ , thus proving the lemma.  $\square$

In other words, for an hereditary graph family  $\mathcal{F}$  which is closed under true twin addition and for every graph  $G$ , there always exists an optimal  $\mathcal{F}$ -edition  $F$  (resp.  $\mathcal{F}$ -deletion,  $\mathcal{F}$ -completion) such that :

- 1) every edge between two vertices of a critical clique of  $G$  is an edge of  $G + F$ , and
- 2) between two distinct critical cliques  $K, K' \in \mathcal{K}(G)$ , either  $V(K) \times V(K') \subseteq E(G + F)$  or  $(V(K) \times V(K')) \cap E(G + F) = \emptyset$ .

From now on, every considered optimal edition (resp. deletion, completion) is supposed to verify these two properties.

### 1.2. Leaf powers

**Definition 1.5.** Let  $T$  be an unrooted tree whose leaves are one-to-one mapped to the elements of a set  $V$ . The  $k$ -leaf power of  $T$  is the graph  $T^k$ , with  $T^k = (V, E)$  where  $E = \{uv \mid u, v \in V \text{ and } d_T(u, v) \leq k\}$ . We call  $T$  a  $k$ -leaf root of  $T^k$ .

It is easy to see that for every  $k$ , the  $k$ -leaf power family of graphs satisfies the conditions of Lemma 1.4. In this paper we focus on the class of 3-leaf powers for which several characterizations are known.

**Theorem 1.6.** [4, 10] For a graph  $G$ , the following conditions are equivalent:

1.  $G$  is a 3-leaf power.
2.  $G$  is  $\{\text{bull, dart, gem, } C_{\geq 4}\}$ -free,  $C_{\geq 4}$  being the cycles of length at least 4. (see Figure 1).
3. The critical clique graph  $\mathcal{C}(G)$  is a forest.

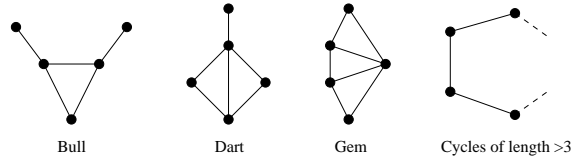


Figure 1: Forbidden induced subgraphs of a 3-leaf power.

The parameterized 3-LEAF POWER EDITION problem, with respect to parameter  $k$  being the size of the edited set, is tractable. An  $O((2k + 8)^k \cdot nm)$  algorithm is proposed in [10]. The proofs of our kernel for the 3-LEAF POWER EDITION problem rely on the critical clique graph characterization and on the following new one which is based on the

join composition of graphs.

**Join Composition.** Let  $G_1 = (V_1, E_1)$  and  $G_2 = (V_2, E_2)$  be two disjoint graphs and let  $S_1 \subseteq V_1$  and  $S_2 \subseteq V_2$  be two non empty subsets of vertices. The *join composition* of  $G_1$  and  $G_2$  on  $S_1$  and  $S_2$ , denoted  $(G_1, S_1) \otimes (G_2, S_2)$ , results in the graph  $H = (V_1 \cup V_2, E_1 \cup E_2 \cup (V(S_1) \times V(S_2)))$  (see Figure 2).

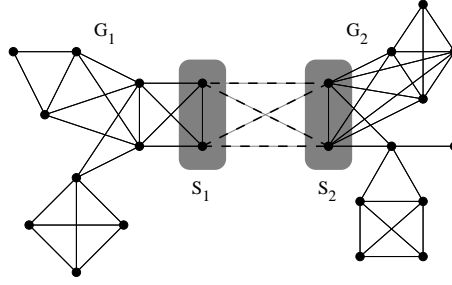


Figure 2: The join composition  $H = (G_1, S_1) \otimes (G_2, S_2)$  creates the dotted edges. As  $G_1$  and  $G_2$  are two 3-leaf powers and as the subsets  $S_1$  and  $S_2$  of vertices are critical cliques of respectively  $G_1$  and  $G_2$ , by Theorem 1.7,  $H$  is also a 3-leaf power.

**Theorem 1.7.** *Let  $G_1 = (V_1, E_1)$  and  $G_2 = (V_2, E_2)$  be two connected 3-leaf powers. The graph  $H = (G_1, S_1) \otimes (G_2, S_2)$ , with  $S_1 \subseteq V_1$  and  $S_2 \subseteq V_2$ , is a 3-leaf power if and only if one of the following conditions holds:*

1.  $S_1$  and  $S_2$  are two cliques of  $G_1$  and  $G_2$  respectively, and if  $S_1$  (resp.  $S_2$ ) is not critical, then  $G_1$  (resp.  $G_2$ ) is a clique or,
2. there exists a vertex  $v \in V_1$  such that  $S_1 = N[v]$  and  $S_2 = V_2$  is a clique.

**Proof.**

$\Leftarrow$  If condition (2) holds, then  $H$  is obtained from  $G_1$  by adding true twins to  $v$ , hence  $H$  is a 3-leaf power.. Assume  $S_1$  and  $S_2$  are two cliques. If  $S_1$  and  $S_2$  are both critical cliques of respectively  $G_1$  and  $G_2$ , then the critical clique graph  $\mathcal{C}(H)$  is clearly the tree obtained from  $\mathcal{C}(G_1)$  and  $\mathcal{C}(G_2)$  by adding the edges between  $S_1$  and  $S_2$ . By Theorem 1.6,  $H$  is a 3-leaf power. For  $i = 1$  or  $2$ , if  $G_i$  is a clique and  $S_i \subset V(G_i)$ , then  $S_i$  and  $V(G_i) \setminus S_i$  are critical cliques in  $H$ . Again, it is easy to see that  $\mathcal{C}(H)$  is a tree.

$\Rightarrow$  First, let us notice that if  $S_1$  and  $S_2$  are not cliques, then  $H$  contains a  $C_4$ , which is forbidden. So let us assume that  $S_1$  is not a clique but  $S_2$  is. Then  $S_1$  contains two non-adjacent vertices  $x$  and  $y$ . First of all, if  $d_{S_1}(x, y) = \infty$  (i.e. if  $x$  and  $y$  are not connected in  $G_1[S_1]$ ), we consider  $\pi_{G_1}$  a shortest path in  $G - 1$  between

the connected component of  $G_1[S_1]$  containing  $x$  and the one containing  $y$  (such a path exists because  $G_1$  is connected). It is easy to see that  $v$  together with  $\pi_{G_1}$  forms an induced (chordless) cycle in  $H$ , which is forbidden. Now, if  $d_{S_1}(x, y) > 2$ , then  $H$  contains a *gem*. To see this, let  $\pi_{S_1}$  be a shortest  $x, y$ -path in  $S_1$ . Together with any vertex  $v \in S_2$ , the vertices of  $\pi_{S_1}$  induce a cycle at length at least 5 in  $H$ . By construction the only possible chords are incident to  $v$ . So any 4 consecutive vertices on  $\pi_{S_1}$  plus the vertex  $v$  induce a *gem*. It follows that there exists in  $S_1$  a vertex  $u$  which dominates  $x$  and  $y$ . Now if there exists a vertex in  $V(G_2) \setminus S_2$ , as  $G_2$  is connected, there exists two adjacent vertices,  $v \in S_2$  and  $w \in V(G_2) \setminus S_2$ . But,  $\{w, u, x, y, v\}$ , induce a *dart* in  $H$ , which is also forbidden. So,  $S_2 = V(G_2)$  and  $G_2$  is a clique. Finally, assume by contradiction again that  $u$  has a neighbor  $w \in V(G_1) \setminus S_1$ . Considering a vertex  $v$  of  $S_2$ , the set of vertices  $\{w, x, y, u, v\}$  induces an obstruction in  $H$ , whatever the adjacency between  $w$  and  $\{x, y\}$  is. So,  $N[u] \subset S_1$ . Conversely, if  $S_1$  contains a vertex  $w \notin N(u)$ ,  $\{w, x, y, u, v\}$  induces an obstruction in  $H$ . So,  $S_1 = N[u]$ , as expected in condition (2).

Assume now that both  $S_1$  and  $S_2$  are cliques. If  $S_1$  and  $S_2$  are not modules in respectively  $G_1$  and  $G_2$ , then we can find a *bull* in  $H$ . Assume that only  $S_1$  is not a module, *i.e.* there exist  $x, y \in S_1$  and  $u \in V(G_1) \setminus S_1$  such that *w.l.o.g.*  $ux \in E(G_1)$  and  $uy \notin E(G_1)$ . If  $S_2 \neq V(G_2)$ , then again  $H$  has a *bull* induced by  $\{u, x, y, v, w\}$  with  $v \in S_2$  and  $w \in V(G_2) \setminus S_2$ ,  $w$  neighbor of  $v$ . Otherwise, either condition (2) holds or  $y$  has a neighbor  $w$  in  $V(G_1) \setminus S_1$ . The latter case is impossible since we find in  $H$  an obstruction induced by  $\{u, x, y, v, w\}$  whatever the adjacency between  $w$  and  $\{u, x\}$  is. Finally assume that  $S_1$  and  $S_2$  are modules, but consider the case where  $S_1$  is not critical (the case  $S_2$  is not critical is symmetric). Then there exists a critical clique  $C_1 \in \mathcal{K}(G_1)$  containing  $S_1$ . Denote by  $x$  a vertex of  $S_1$  and by  $y$  a vertex of  $C_1 \setminus S_1$ . If  $V(G_1) \neq C_1$ , then  $G_1$  contains two non-adjacent vertices, say  $u$  and  $u'$ . If  $u = x$  and  $u' \notin C_1$ , then as  $G_1$  is connected, we can choose  $u'$  and  $w \notin C_1$  such that  $\{u', w, x, y, v\}$  with  $v \in S_2$  is a *bull* in  $H$ . Otherwise we can choose  $u$  and  $u'$  both adjacent to the vertices of  $C_1$ , and then  $\{u, u', x, y, v\}$  would induce a *dart* in  $H$ . It follows that if  $S_1$  is not critical, then condition (1) holds.

□

The following observation will be helpful to apply Theorem 1.7 in the safeness' proofs of the reduction rules.

**Observation 1.8.** *Let  $C$  be a critical clique of a 3-leaf power  $G = (V, E)$ . For every  $S \subseteq V$ , if the clique  $C \setminus S$  is not critical in  $G[V \setminus S]$ , then the connected component of  $G[V \setminus S]$  containing  $C \setminus S$  is a clique.*

**Proof.** Assume that  $C \setminus S$  is not a critical clique of  $G[V \setminus S]$ , *i.e.* though  $C \setminus S$  is a clique module in  $G[V \setminus S]$ , it is not maximal. Let  $x \notin S$  be a vertex such that  $C \cup \{x\}$  is a clique module of  $G[V \setminus S]$ . Then  $x$  belongs to a critical clique  $C'$  of  $G$  adjacent to  $C$  in  $\mathcal{C}(G)$ . It follows that  $S$  has to contain the union of all the critical cliques of  $G$  adjacent

to  $C$  in  $\mathcal{C}(G)$  but  $C'$  (otherwise  $C \cup \{x\}$  could not be a module of  $G[V \setminus S]$ ), and all the critical cliques of  $G$  adjacent to  $C'$  in  $\mathcal{C}(G)$  but  $C$  (for the same reason). This means that the connected component containing  $C \setminus S$  in  $G[V \setminus S]$  is a subset of  $C \cup C'$  which is a clique.  $\square$

Finally, let us conclude this preliminary study of 3-leaf powers by a technical lemma required in the proof of the last reduction rule.

**Lemma 1.9.** *Let  $G = (V, E)$  be a 3-leaf power. Every cycle  $C$  of length at least 5 in  $G$  contains four distinct vertices  $a, b, c, d$  (appearing in this order along  $C$ ) with  $ab$  and  $cd$  edges of  $C$  such that  $ad \in E$ ,  $ac \in E$  and  $bd \in E$ .*

**Proof.** As the 3-leaf power graphs form an hereditary family, the subgraph  $H$  of  $G$  induced by the vertices of the cycle  $C$  is a 3-leaf power with at least 5 vertices. As  $H$  is not a tree, it contains a critical clique  $K$  of size at least 2. Let  $a$  and  $d$  be two distinct vertices of  $K$ . As  $|C| \geq 5$ , observe that there exist two distinct vertices  $b$  and  $c$ , distinct from  $a$  and  $d$ , such that  $a, b, c$  and  $d$  appear in this order along  $C$  and such that  $ab$  and  $cd$  are edges of  $C$ . As  $K$  is a clique module, any vertex adjacent to some vertex in  $K$  neighbors all the vertices of  $K$ . It follows that  $ad \in E$ ,  $ac \in E$  and  $bd \in E$ .  $\square$

## 2. A kernel for 3-leaf power edition

In this section, we present five preprocessing reduction rules the application of which leads to a kernel with  $O(k^3)$  vertices for the 3-LEAF POWER EDITION problem. The first rule gets rid of connected components of the input graph that are already 3-leaf powers. Rule 2.1 is trivially safe.

**Rule 2.1.** *If  $G$  has a connected component  $C$  such that  $G[C]$  is 3-leaf power, then remove  $C$  from  $G$ .*

The next rule was already used to obtain a kernel with  $O(k^2)$  vertices for the parameterized CLUSTER EDITING problem [24]. It bounds the size of every critical clique in a reduced instance by  $k + 1$ .

**Rule 2.2.** *If  $G$  has a critical clique  $K$  such that  $|K| > k + 1$ , then remove  $|K| - k - 1$  vertices of  $K$  from  $V(G)$ .*

**Proof.** By Lemma 1.4, we know that there always exists an optimal 3-leaf power edition that contains none or every edge incident to a critical clique  $K$ . Thus, if  $|K| \geq k + 1$ , this means that there is no optimal 3-leaf power edition that contains an edge incident to  $K$ . As this is still true if  $|K| = k + 1$ , it is safe to remove  $|K| - (k + 1)$  vertices of  $K$  from  $V(G)$  (meaning that every optimal 3-leaf power edition in the reduced graph will also be an optimal 3-leaf power edition in the input graph).  $\square$

### 2.1. Branch reduction rules

We now assume that the input graph  $G$  is reduced under Rule 2.1 (*i.e.* none of the connected component is a 3-leaf power) and Rule 2.2 (*i.e.* critical cliques of  $G$  have size at most  $k + 1$ ). The next three reduction rules use the fact that the critical clique graph of a 3-leaf power is a forest. The idea is to identify induced subgraphs of  $G$ , called *branches*, which correspond to subtrees of  $\mathcal{C}(G)$ . That is, a branch of  $G$  is an induced subgraph which is already a 3-leaf power. More precisely:

**Definition 2.1.** Let  $G = (V, E)$  be a graph. An induced subgraph  $G[S]$ , with  $S \subseteq V$ , is a *branch* if  $S$  is the disjoint union of critical cliques  $K_1, \dots, K_r \in \mathcal{K}(G)$  such that the subgraph of  $\mathcal{C}(G)$  induced by  $\{K_1, \dots, K_r\}$  is a tree.

Let  $B = G[S]$  be a branch of a graph  $G$  and let  $K_1, \dots, K_r$  be the critical cliques of  $G$  contained in  $S$ . We say that  $K_i$  ( $1 \leq i \leq r$ ) is an *attachment point* of the branch  $B$  if it contains a vertex  $x$  such that  $N_G(x)$  intersects  $V(G) \setminus S$ . A branch  $B$  is a  *$l$ -branch* if it has exactly  $l$  attachment points. Our next three rules deal with 1-branches and 2-branches.

In the following, we denote by  $B^R$  the subgraph of  $B$  in which the vertices of the attachment points have been removed. If  $P$  is an attachment point of  $B$ , then the set of neighbors of vertices of  $P$  in  $B$  is denoted  $N_B(P)$ .

**Lemma 2.2.** Let  $G = (V, E)$  be a graph and  $B$  be a 1-branch of  $G$  with attachment point  $P$ . There exists an optimal 3-leaf power edition  $F$  of  $G$  such that :

1. the set of affected vertices of  $B$  is a subset of  $P \cup N_B(P)$  and
2. in  $G + F$ , the vertices of  $N_B(P)$  are all adjacent to the same vertices of  $V(G) \setminus V(B^R)$ .

**Proof.** Let  $F$  be an arbitrary optimal 3-leaf power edition of  $G$ . We construct from  $F$  another optimal 3-leaf power edition which satisfies the two conditions above.

Let  $C$  be the critical clique of  $H = G + F$  that contains  $P$  and set  $C' = C \setminus B^R$ . By Lemma 1.4, the set of critical cliques of  $G$  whose vertices belong to  $N_B(P)$  contains two kind of cliques:  $K_1, \dots, K_c$ , whose vertices are in  $C'$  or adjacent to the vertices of  $C'$  in  $H$ , and  $K_{c+1}, \dots, K_h$  whose vertices are not adjacent to the vertices of  $C'$  in  $H$ . For  $i \in \{1, \dots, h\}$ , let  $C_i$  be the connected component of  $B^R$  containing  $K_i$ .

Let us consider the three following induced subgraphs :  $G_1$  the subgraph of  $G$  induced by the disjoint union of  $C_1, \dots, C_c$ ;  $G_2$  the subgraph of  $G$  induced by the disjoint union of  $C_{c+1}, \dots, C_h$ ; and finally  $G'$ , the subgraph of  $H$  induced by  $V(G) \setminus V(B^R)$ . Let us notice that these three graphs are 3-leaf powers.

By Observation 1.8, if  $C'$  is not a critical clique of  $G'$ , then the connected component of  $G'$  containing  $C'$  is a clique. Similarly, if  $K_i$ , for every  $1 \leq i \leq c$ , is not a critical clique of  $G_1$ , then the connected component of  $G_1$  containing  $K_i$  is a clique. Thus, by Theorem 1.7, the disjoint union  $H'$  of  $G_2$  and  $(G', C') \otimes (G_1, \{K_1, \dots, K_c\})$  is a 3-leaf power. By construction, the edge edition set  $F'$  such that  $H' = G + F'$  is a subset of  $F$  and thus  $|F'| \leq |F|$ . Moreover the vertices of  $B$  affected by  $F'$  all belong to  $P \cup N_B(P)$ , which proves the first point.

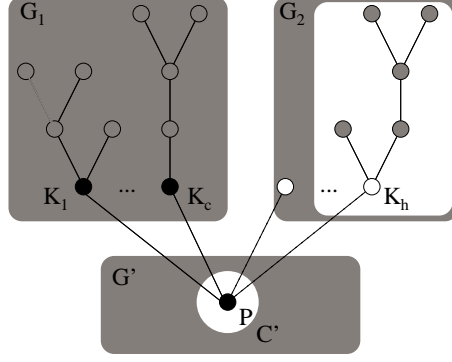


Figure 3: A 1-branch of  $G$  where all editions adjacent to vertices in  $V(B^R) \setminus N_B(P)$  have been removed. This shows in particular the appearance of the three subgraphs we consider.

To state the second point, we focus on the relationship between the critical cliques  $K_i$  and  $C'$  in  $H' = G + F'$ . If some  $K_i$  is linked to  $C'$  in  $H'$  (i.e.  $c > 1$ ), it means that the cost of adding the missing edges between  $K_i$  and  $C'$  (which, by Theorem 1.7, would also result in a 3-leaf power) is lower than the cost of removing the existing edges between  $K_i$  and  $C'$ :  $|K_i| \cdot |C' \setminus P| \leq |K_i| \cdot |P|$ . On the other hand, if some  $K_j$  is not linked to  $C'$  in  $H'$  (i.e.  $c < h$ ), we conclude that  $|P| \leq |C' \setminus P|$ . Finally, if both cases occur, we have  $|P| = |C' \setminus P|$ , and we can choose to add all or none of the edges between  $K_i$  and  $C'$ . In all cases, we provide an optimal edition of  $G$  into a 3-leaf power in which the vertices of  $N_B(P)$  are all adjacent to the same vertices of  $V(G) \setminus V(B^R)$ .  $\square$

We can now state the first 1-branch reduction rule whose safeness follows from Lemma 2.2.

**Rule 2.3.** *If  $G$  contains a 1-branch  $B$  with attachment point  $P$ , then remove from  $G$  the vertices of  $B^R$  and add a new critical clique of size  $\min\{|N_B(P)|, k + 1\}$  adjacent to  $P$ .*

Our second 1-branch reduction rule considers the case where several 1-branches are attached to the rest of the graph by a join. The following lemma shows that under certain cardinality conditions, the vertices of such 1-branches are not affected by an optimal 3-leaf power edition.

**Lemma 2.3.** *Let  $G = (V, E)$  be a graph for which a 3-leaf power edition of size at most  $k$  exists. Let  $B_1, \dots, B_l$  be 1-branches, the attachment points  $P_1, \dots, P_l$  of which all have the same neighborhood  $N$  in  $V \setminus \cup_{i=1}^l V(B_i)$ . If  $\sum_{i=1}^l |P_i| > 2k + 1$ , then, in every 3-leaf power optimal edition  $F$  of  $G$ ,  $N$  has to be a critical clique of  $H = G + F$  and none of the vertices of  $\cup_{i=1}^l V(B_i)$  is affected.*

**Proof.** We just show that every optimal 3-leaf power edition  $F$  of  $G$  has to transform  $N$  into a critical clique, which directly implies the second part of the result. Notice that



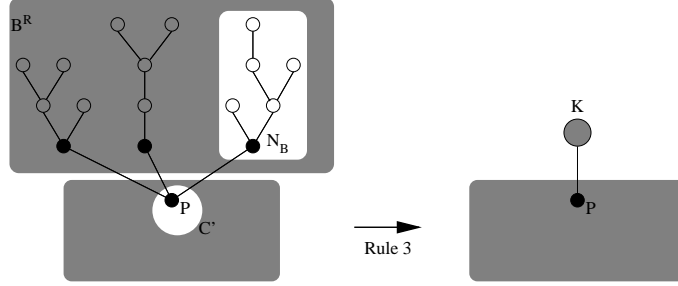


Figure 4: On the left, a 1-branch  $B$  with attachment point  $P$ . On the right, the effect of Rule 2.3 which replace  $B^R$  by a clique  $K$  of size  $\min\{|N_B(P)|, k+1\}$ .

since  $G$  is reduced under Rule 2.2, every attachment point  $P_i$  satisfies  $|P_i| \leq k+1$ , thus implying that  $l \geq 2$ .

First, assume that  $F$  does not edit  $N$  into a clique: *i.e.* there are two vertices  $a$  and  $b$  of  $N$  such that  $ab \notin E(H)$ . For every pair of vertices  $u_i \in P_i$  and  $u_j \in P_j$  with  $i \neq j$ , the set  $\{a, b, u_i, u_j\}$  cannot induce a chordless cycle in  $H$ , which implies that the vertices of  $P_i$  or those of  $P_j$  must be affected. It follows that among the attachment points, the vertices of at most one  $P_i$  are not affected by  $F$ . As  $|P_i| \leq k+1$  for every  $i$ , the size of  $F$  has to be at least  $k+1$ : contradicting the assumptions. So  $N$  is a clique in  $H$ .

Now, assume that  $N$  is not a module of  $H$  : *i.e.* there exists  $w \notin N$  such that for some  $x, y \in N$  we have w.l.o.g.  $xw \in E(H)$  and  $yw \notin E(H)$ . As  $|F| \leq k$ , there exist two vertices  $u_i \in P_i$  and  $u_j \in P_j$ , non affected by  $F$  and such that  $u_i u_j \notin E(H)$ . Together with  $x, y$  and  $w$ ,  $u_i$  and  $u_j$  induce a *dart* in  $H$ , contradicting Theorem 1.6. So, in  $H$ , the set of vertices  $N$  has to be a clique module.

Finally, let us notice that  $N$  has to be critical in  $H$ , otherwise it would imply that there exists a vertex  $v \notin N$  that has been made adjacent to at least  $k+1$  vertices of  $\cup_{i=1}^l B_i$ , implying that  $|F| > k$ : contradiction.  $\square$

By Lemma 2.3, if there exists a 3-leaf power edition  $F$  of  $G$  such that  $|F| \leq k$ , then the 1-branches  $B_1, \dots, B_l$  can be safely replaced by two critical cliques of size  $k+1$ . This gives us the second 1-branch reduction rule.

**Rule 2.4.** *If  $G$  has several 1-branches  $B_1, \dots, B_l$ , the attachment points  $P_1, \dots, P_l$  of which all have the same neighborhood  $N$  in  $V \setminus \cup_{i=1}^l V(B_i)$  and if  $\sum_{i=1}^l |P_i| > 2k+1$ , then remove from  $G$  the vertices of  $\cup_{i=1}^l V(B_i)$  and add two new critical cliques of size  $k+1$  neighboring exactly  $N$ .*

## 2.2. The 2-branch reduction rule

Let us consider a 2-branch  $B$  of a graph  $G = (V, E)$  with attachment points  $P_1$  and  $P_2$ . The subgraph of  $G$  induced by the critical cliques of the unique path from  $P_1$  to  $P_2$  in

$\mathcal{C}(B)$  is called the *main path* of  $B$  and denoted  $path(B)$ . A min-cut of  $path(B)$  is a set  $F$  of edges of  $B$  such that  $B - F$  does not contain any path from  $P_1$  to  $P_2$  and such that  $F$  has minimal cardinality for this property. We say that  $B$  is *clean* if  $P_1$  and  $P_2$  are leaves of  $\mathcal{C}(B)$ , in which case we denote by  $Q_1$  and  $Q_2$  the critical cliques which respectively neighbor  $P_1$  and  $P_2$  in  $B$ .

**Lemma 2.4.** *Let  $B$  be a clean 2-branch of a graph  $G = (V, E)$  with attachment points  $P_1$  and  $P_2$  such that  $path(B)$  contains at least 5 critical cliques. There exists an optimal 3-leaf power edition  $F$  of  $G$  such that :*

1. *if  $path(B)$  is a disconnected subgraph of  $G + F$ , then  $F$  may contain a min-cut of  $path(B)$ ;*
2. *and in each case, the other affected vertices of  $B$  belong to  $V(P_1 \cup Q_1 \cup P_2 \cup Q_2)$ .*

**Proof.** Let  $F$  be an arbitrary optimal 3-leaf power edition of  $G$ . We call  $C_1$  and  $C_2$  the critical cliques of  $G + F$  that respectively contain  $P_1$  and  $P_2$  (possibly,  $C_1$  and  $C_2$  could be the same), and denote  $C_1 \setminus B^R$  and  $C_2 \setminus B^R$  respectively by  $C'_1$  and  $C'_2$  see Figure 6). We will construct from  $F$  another optimal 3-leaf power edition  $F'$  of  $G$  satisfying the statement.

- *Assume that  $F$  disconnects  $path(B)$ .* First of all, it is clear that for every subset  $F_1$  of  $F$ , if  $F_2$  is an optimal edition of  $H_1 = G + F_1$ , then  $F' = F_1 \cup F_2$  is an optimal edition of  $G$ . We use this fact in the following different cases. Assume that  $F$  contains the edges  $F_1 := V(P_1) \times V(Q_1)$  and consider the graph  $H_1 := G + F_1$ . We call  $B_1$  the 1-branch  $B \setminus P_1$  of  $H_1$  whose attachment point is  $P_2$ . Then, Lemma 2.2 applies to  $B_1$  and provides from  $F$  an optimal 3-leaf power edition of  $H_1$ , say  $F_2$ , where the edited vertices of  $B_1$  are contained in  $V(P_2 \cup Q_2)$ . By the previous observation, it follows that  $F_1 \cup F_2$  is an optimal edition for  $G$  that respects conditions (1) and (2). We proceed similarly if  $F$  contains the edges  $V(P_2) \times V(Q_2)$ . Now, assume that  $F$  does not contain  $V(P_1) \times V(Q_1)$  nor  $V(P_2) \times V(Q_2)$ . In that case, there exists  $F_1 \subset F$  which is a minimal cut of  $path(B)$  disjoint from  $V(P_1) \times V(Q_1)$  and  $V(P_2) \times V(Q_2)$ . Then, there are two connected components in  $B + F_1$ , the one containing  $P_1$ , say  $B_1$ , and the one containing  $P_2$ , say  $B_2$ . The subgraphs  $B_1$  and  $B_2$  of  $H_1 := G + F_1$  are 1-branches with respectively  $P_1$  and  $P_2$  as attachment points. So, Lemma 2.2 applies to  $B_1$  and  $B_2$ , and provides from  $F$  an optimal 3-leaf power edition of  $H_1$ , say  $F_2$ , where the edited vertices of  $B_1$  and  $B_2$  are contained in  $V(P_1 \cup P_2 \cup Q_1 \cup Q_2)$ . To conclude, remark that if  $F_1$  is not a minimum (for cardinality) cut of  $path(B)$ , we could replace  $F_1$  by such a minimum cut, and perform a similar 3-leaf power edition for  $G$  with size strictly lower than  $|F|$ , thus contradicting the choice of  $F$ . It follows that  $F_1 \cup F_2$  is an optimal edition for  $G$  that respects conditions (1) and (2).

- *Assume that  $F$  does not disconnect  $path(B)$ .* Let  $X_1$  (resp.  $X_2$ ) be the connected component of  $(G + F) \setminus B^R$  containing  $P_1$  (resp.  $P_2$ ).

We first consider the case where  $X_1$  and  $X_2$  are two distinct connected components. By definition,  $B^R$  is a 3-leaf power and  $Q_1$  and  $Q_2$  are two of its critical cliques (since  $\text{path}(B)$  contains at least 5 critical cliques). Moreover, the subgraph  $X_1$  (resp.  $X_2$ ) is also a 3-leaf power which is a clique if  $C'_1$  (resp.  $C'_2$ ) is not a critical clique (Observation 1.8). By Theorem 1.7, it follows that the composition of these three subgraphs :  $H' = (X_1, C'_1) \otimes (B^R, Q_1)$  and  $(H', Q_2) \otimes (X_2, C'_2)$  yields a 3-leaf power. Thus, if  $F$  affects some vertices of  $V(B^R) \setminus V(Q_1 \cup Q_2)$ , then a smaller edition could be found by removing from  $F$  the edges in  $V(B^R) \times V(B^R)$ . This would contradict the optimality of  $F$ .

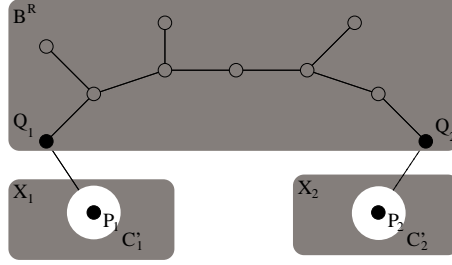


Figure 5: A clean 2-branch of  $G$ . The bold edge represent the first join composition  $H' := (X_1, C'_1) \otimes (B^R, Q_1)$  while the dotted bold edge represent the join composition which is done in a second place, namely  $(H', Q_2) \otimes (X_2, C'_2)$

So, assume that  $P_1$  and  $P_2$  belong to the same connected component  $X$  of  $(G + F) \setminus B^R$ . Let  $y_1$  and  $y_2$  be respectively vertices of  $P_1$  and  $P_2$  (in the case  $C_1 = C_2$ , choose  $y_1 = y_2$ ). Let  $\pi_B$  and  $\pi_X$  be two distinct paths between  $y_1$  and  $y_2$  defined as follows :  $\pi_B$  is obtained by picking one vertex  $b_i$  in each critical clique  $H_i$  of  $\text{path}(B)$  ( $H_1 = P_1$  and  $H_q = P_2$ , with  $q \geq 5$ ) and  $\pi_X$  is a chordless path in  $X$  (thereby its vertices  $x_1, \dots, x_p$ , with  $x_1 = y_1$  and  $x_p = y_2$  belong to distinct critical cliques of  $G + F$ , say  $K_1, \dots, K_p$ , with  $K_1 = C'_1$  and  $K_p = C'_2$ ). The union of these two paths results in a cycle  $C$  of length at least 5. By Lemma 1.9, there are two disjoint edges  $e = ab$  and  $f = cd$  in  $C$  such that the edges  $(a, c)$  and  $(b, d)$  belong to  $E \triangle F$ . By construction of  $C$ , at most one of the edges  $e$  and  $f$  belongs to  $\pi_X$  (otherwise  $\pi_X$  would not be chordless). We now study the different cases :

*Edges  $e$  and  $f$  belong to  $\pi_B$ .* W.l.o.g assume that  $a = b_i$ ,  $b = b_{i+1}$  and  $c = b_j$ ,  $d = b_{j+1}$  ( $i+1 < j$ ). By Lemma 1.4,  $F$  contains the set of edges  $(V(H_i) \times V(H_j)) \cup (V(H_{i+1}) \times V(H_{j+1}))$ . Notice that  $\min\{|H_i| \cdot |H_{i+1}|, |H_j| \cdot |H_{j+1}|\} < |H_i| \cdot |H_j| + |H_{i+1}| \cdot |H_{j+1}|$ . W.l.o.g., assume that  $\min\{|H_i| \cdot |H_{i+1}|, |H_j| \cdot |H_{j+1}|\} = |H_i| \cdot |H_{i+1}|$ . We will 'cut' the edges between  $H_i$  and  $H_{i+1}$ . Consider the set :

$$F' = (F \setminus (V \times V(B^R))) \cup (V(H_i) \times V(H_{i+1}))$$

Moreover, if  $H_i \neq P_1$ , add to  $F'$  the edges  $F_{C_1} := V(C'_1 \setminus P_1) \times V(Q_1)$  (which belong to  $F$ ) and, if  $H_{j+1} \neq P_2$ , add to  $F'$  the edges  $F_{C_2} := V(C'_2 \setminus P_2) \times V(Q_2)$  (which belong to  $F$ ). In all cases, we have  $|F'| < |F|$ . As in the case where  $X_1$  and  $X_2$  were distinct connected components, by Theorem 1.7, the graph  $G + F'$  is a 3-leaf power : contradicting the optimality of  $F$ .

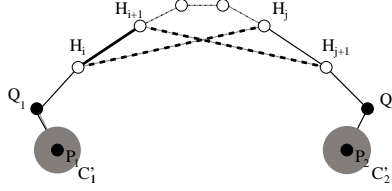


Figure 6: An illustration of the first case, with  $H_i \neq P_1$  and  $H_{j+1} \neq P_2$ . The bold edges represent the edges we put in  $F'$  while the dotted bold edges denote the edges we removed from  $F$  to obtain  $F'$ .

Edge  $e$  belongs to  $\pi_B$  and  $f$  to  $\pi_X$ . W.l.o.g., assume that  $a = b_i$  and  $b = b_{i+1}$ , and  $c = x_{j+1}$ ,  $d = x_j$ . As above, by Lemma 1.4,  $F$  contains  $(V(H_i) \times V(K_{j+1})) \cup (V(H_{i+1}) \times V(K_j))$ . Notice that  $\min\{|H_i| \cdot |H_{i+1}|, |K_j| \cdot |K_{j+1}|\} < |H_i| \cdot |K_{j+1}| + |H_{i+1}| \cdot |K_j|$ . If  $\min\{|H_i| \cdot |H_{i+1}|, |K_j| \cdot |K_{j+1}|\} = |H_i| \cdot |H_{i+1}|$ , then we consider the set :

$$F' = (F \setminus (V \times V(B^R))) \cup (V(H_i) \times V(H_{i+1}))$$

Here again, if  $H_i \neq P_1$ , add to  $F'$  the edges  $F_{C_1}$  (which belong to  $F$ ) and, if  $H_{j+1} \neq P_2$ , add to  $F'$  the edges  $F_{C_2}$  (which belong to  $F$ ). As previously,  $|F'|$  is smaller than  $|F|$  and by Theorem 1.7, we can prove that  $G + F'$  is a 3-leaf power. Finally, if  $\min\{|H_i| \cdot |H_{i+1}|, |K_j| \cdot |K_{j+1}|\} = |K_j| \cdot |K_{j+1}|$ , then we consider the set

$$F' = (F \setminus (V \times V(B^R))) \cup (V(K_j) \times V(K_{j+1})) \cup F_{C_1} \cup F_{C_2}$$

Again  $|F'|$  is smaller than  $|F|$  and by Theorem 1.7, we can prove that  $G + F'$  is a 3-leaf power. In each case, we found a better 3-leaf power edition  $F'$ , contradicting the optimality of  $F$ .

□

**Rule 2.5.** Let  $G$  be a graph having a clean 2-branch  $B$  such that  $\text{path}(B)$  is composed by at least 5 critical cliques. Remove from  $G$  all the vertices of  $V(B)$  except those of  $V(P_1 \cup Q_1 \cup P_2 \cup Q_2)$  and add four new critical cliques :

- $K_1$  (resp.  $K_2$ ) of size  $k + 1$  adjacent to  $Q_1$  (resp.  $Q_2$ );
- $K'_1$  (resp.  $K'_2$ ) adjacent to  $K_1$  (resp.  $K_2$ ) and such that  $K'_1$  and  $K'_2$  are adjacent and  $|K'_1| \cdot |K'_2|$  equals the min-cut of  $\text{path}(B)$ .

**Proof.** Let  $B'$  be the 2-branch replacing  $B$  after the application of the rule. It is easy to see that by construction the min-cut of  $B'$  equals the min-cut of  $\text{path}(B)$ . Moreover the attachment points  $P_1$  and  $P_2$  and their respective neighbors  $Q_1$  and  $Q_2$  are unchanged. It follows by Lemma 2.4 that every optimal edition  $F$  of  $G$  corresponds to an optimal edition  $F'$  of  $G'$ , the graph reduced by Rule 2.5, such that  $|F| = |F'|$ . □

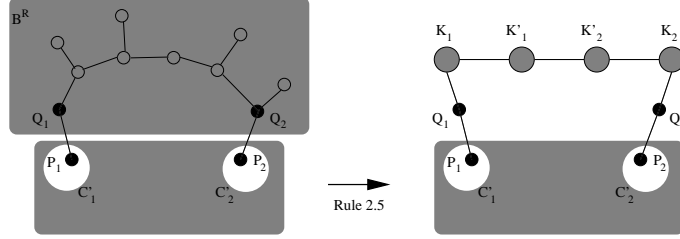


Figure 7: A 2-branch  $B$  on the left (only pendant critical cliques are hanging on  $path(B)$  since we can assume that the graph is reduced by the previous rules). On the right, the way Rule 2.5 reduces  $B$ .

### 2.3. Kernel size and time complexity

Let us discuss the time complexity of the reduction rules. The 3-leaf power recognition problem can be solved in  $O(n + m)$  [4]. It follows that Rule 2.1 requires linear time. To implement the other reduction rules, we first need to compute the critical clique graph  $\mathcal{C}(G)$ . As noticed in [24],  $\mathcal{C}(G)$  can be built in  $O(n + m)$ . For instance, to do so, we can compute in linear time the modular decomposition tree of  $G$ , which is a classical and well-studied problem in algorithmic graph theory (see [26] for a recent paper). Then, a critical clique is a serie-node of the decomposition tree with only leaves below it. Given  $\mathcal{K}(G)$ , which is linear in the size of  $G$ , it is easy to detect the critical cliques of size at least  $k + 1$ . So, Rule 2.2 requires linear time. A search on  $\mathcal{C}(G)$  can identify the 1-branches. It follows that the two 1-branches reduction rules (Rule 2.3 and Rule 2.4) can also be applied in  $O(n + m)$  time. Let us now notice that in a graph reduced by the first four reduction rules, a 2-branch is a path to which pendant vertices are possibly attached. It follows that to detect a 2-branch  $B$ , such that  $path(B)$  contains at least 5 critical cliques, we first prune the pendant vertices, and then identify in  $\mathcal{C}(G)$  the paths containing only vertices of degree 2, and at least 5 of them. To do this, we compute the connected components of the graph induced on vertices of degree 2 in  $\mathcal{C}(G)$ . This shows that Rule 2.5 can be carried in linear time.

**Theorem 2.5.** *The parameterized 3-LEAF POWER EDITION problem admits a kernel with  $O(k^3)$  vertices. Given a graph  $G$ , a reduced instance can be computed in linear time.*

**Proof.** The discussion above established the time complexity to compute a kernel. Let us determine the kernel size. Let  $G = (V, E)$  be a reduced graph (*i.e.* none of the reduction rules applies to  $G$ ) which can be edited into a 3-leaf power with a set  $F \subseteq V \times V$  such that  $|F| \leq k$ . Let us denote  $H = G + F$  the edited graph. We first show that  $\mathcal{C}(H)$  has  $O(k^2)$  vertices (*i.e.*  $|\mathcal{C}(H)| \in O(k^2)$ ), and then Lemma 1.3 enables us to conclude.

We say that a critical clique is affected if it contains an affected vertex and denote by  $A$  the set of the affected critical cliques. As each edge of  $F$  affects two vertices, we have that  $|A| \leq 2k$ . Since  $H$  is a 3-leaf power, its critical clique graph  $\mathcal{C}(H)$  is a tree. Let  $T$  be

the minimal subtree of  $\mathcal{C}(H)$  that spans the affected critical cliques. Let us observe that if  $B$  is a maximal subtree of  $\mathcal{C}(H) - T$ , then none of the critical cliques in  $B$  contains an affected vertex and thus  $B$  was the critical clique graph of a 1-branch of  $G$ , which has been reduced by Rule 2.3 or Rule 2.4. Let  $A' \subset \mathcal{K}(H)$  be the critical cliques of degree at least 3 in  $T$ . As  $|A| \leq 2k$ , we also have  $|A'| \leq 2k$ . The connected components resulting from the removal of  $A$  and  $A'$  in  $T$  are paths. There are at most  $4k$  such paths. Each of these paths is composed by non-affected critical cliques. It follows that each of them corresponds to  $path(B)$  for some 2-branch  $B$  of  $G$ , which has been reduced by Rule 2.5.

From these observations, we can now estimate the size of the reduced graph. Attached to each of the critical cliques of  $T \setminus A$ , we can have 1 pendant critical clique resulting from the application of Rule 2.3. Remark that any 2-branch reduced by Rule 2.5 has no such pendant clique and that  $path(B)$  contains 5 critical cliques. So, a considered 2-branch in  $\mathcal{C}(H)$  is made of at most 8 critical cliques. Finally, attached to each critical clique of  $A$ , we can have at most  $(4k + 2)$  extra critical cliques resulting from the application of Rule 2.4. See Figure 8 for an illustration of the shape of  $\mathcal{C}(H)$ . Summing up everything, we obtain that  $\mathcal{C}(H)$  contains at most  $4k \cdot 8 + 2k \cdot 2 + 2k \cdot (4k + 3) = 8k^2 + 42k$  vertices.

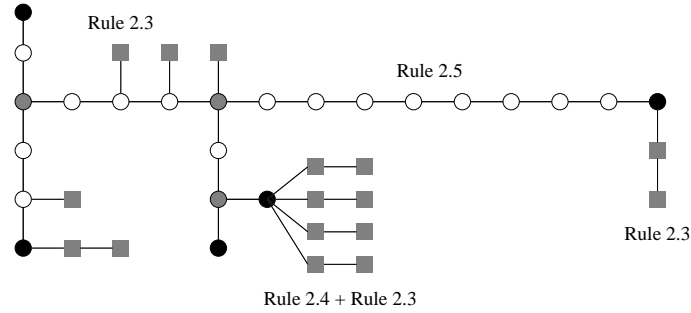


Figure 8: The black circles are the critical cliques of  $A$ , the grey ones belong to  $A'$ , and the squares are the critical cliques not in  $T$ . On the figure, we can observe a 2-branch of size 8 reduced by Rule 2.5. There cannot be pendant critical cliques attached to its nodes. Application of Rule 2.3, may let a path of two critical cliques pendant to the elements of  $A \cup A'$  and a single critical clique pendant to the elements of the small 2-branches. Finally, Rule 2.4 can only affect critical cliques of  $A$ .

By Lemma 1.3, we know that for each edited edge in a graph, the number of critical cliques increase by at most 4. It follows that  $\mathcal{K}(G)$  contains at most  $8k^2 + 46k$  critical cliques. By Rule 2.2, each critical clique of the reduced graph has size at most  $k + 1$ . This implies that the reduced graph contains at most  $8k^3 + 54k^2 + 46k$  vertices, proving the  $O(k^3)$  kernel size.  $\square$

We should notice that some small modifications of the branch reduction rules and a more precise analysis would improve the constants involved in the kernel size. However, the cubic bound would not change.

### 3. Kernels for edge completion and edge deletion

We now prove and adapt the previous rules to the cases where only insertions or only deletions of edges are allowed. First, observe that Rules 2.1 and 2.2 are also safe in 3-LEAF POWER COMPLETION and 3-LEAF POWER EDGE-DELETION (Rule 2.2 directly follows from Lemma 1.4). We have a similar result for the 1-branches reduction rules.

**Lemma 3.1.** *Rule 2.3 is safe for both 3-LEAF POWER COMPLETION and 3-LEAF POWER DELETION.*

**Proof.** In the following, we consider an optimal solution  $F$  such that  $H := G + F$  is a 3-leaf power, denote by  $C$  the critical clique containing  $P$  in  $H$  and set  $C' = C \setminus B^R$ .

- **3-LEAF POWER COMPLETION.** To show the safeness of Rule 2.3 in this case, we will build from  $F$  an optimal 3-leaf power completion that respects conditions of Lemma 2.2. By Lemma 1.4, we know that the set of critical cliques  $\{K_1, \dots, K_h\}$  of  $G$  whose vertices belong to  $N_B(P)$  are in  $C$  or adjacent to the vertices of  $C$  in  $H$ . For  $i \in \{1, \dots, h\}$ , let  $C_i$  be the connected component of  $B^R$  containing  $K_i$ . As previously, we consider  $G_1$  the subgraph of  $G$  induced by the disjoint union of  $C_1, \dots, C_h$  and  $G'$  the subgraph of  $H$  induced by  $V(H) \setminus V(B^R)$ . By Observation 1.8, if  $C'$  is not a critical clique of  $G'$ , then  $G'$  is a clique. Similarly, if  $K_i$ , for every  $1 \leq i \leq h$ , is not a critical clique of  $G_1$ , then  $C_i$  is a clique. By Theorem 1.7, it follows that the graph  $H' := (G', C') \otimes (G_1, \{K_1, \dots, K_h\})$  is a 3-leaf power. By construction, the edge completion set  $F'$ , such that  $H' = G + F'$ , is a subset of  $F$  and the vertices of  $B$  affected by  $F$  all belong to  $P \cup N_B(P)$ . Finally, as every  $K_i$  is connected to  $C'$  in  $H'$ , the vertices of  $N_B(P)$  are all adjacent to the same vertices of  $V(G) \setminus V(B^R)$ .
- **3-LEAF POWER EDGE-DELETION.** In the case where only edges deletion are allowed, we will build from  $F$  an optimal 3-leaf power deletion respecting the conditions of Lemma 2.2 by studying the behavior of  $P$  in  $H$ . First of all, notice that if  $P$  forms a larger critical clique in  $H$  with some vertex  $x \in V(G) \setminus V(B^R)$ , this means that  $F$  contains  $P \times N_B(P)$ . Thus, there is no need to do extra deletions in  $B^R$  and then we are done.

Now, consider the cases where  $P$  is critical in  $H$  or form a larger critical clique with some  $K_i$ . In both cases, we have  $C' = P$ . By Theorem 1.7, the graph  $H' := (G', C') \otimes (G_1, \{K_1, \dots, K_c\})$  is a 3-leaf power, and the edge set  $F'$  used to transform  $G$  into  $H'$  is a subset of  $F$  (all the edges between  $C'$  and  $\{K_1, \dots, K_c\}$  are present in  $G$ ), and then we are done.

□

**Lemma 3.2.** *Rule 2.4 is safe for both 3-LEAF POWER COMPLETION and 3-LEAF POWER DELETION.*



**Proof.** As in Lemma 2.3, we consider  $B_1, \dots, B_l$  1-branches of  $G$ , the attachment points  $P_1, \dots, P_l$  of which all have the same neighborhood  $N$  and satisfy  $\sum_{i=1}^l |P_i| > 2k + 1$ . Again, as every critical clique of  $G$  has at most  $k + 1$  vertices, we have  $l \geq 2$ .

- **3-LEAF POWER COMPLETION.** In this case, same arguments as the ones used in the proof of Lemma 2.3 hold. We briefly detail them. First, assume that  $N$  is not transformed into a clique by an optimal 3-leaf power completion  $F$ . To get rid of all the  $C_4$ 's involving two non-adjacent vertices of  $N$  and  $P_i, P_j, i \neq j$ , the only possibility is to transform  $\cup_{i=1}^l P_i$  into a clique, which requires more than  $k + 1$  edge insertions. Thus  $N$  has to be a clique. Moreover,  $N$  must also become a module, otherwise we would find *darts* that would imply to transform  $\cup_{i=1}^l P_i$  into a clique, which is impossible. Finally,  $N$  must be critical (otherwise, at least one insertion for each vertex of  $\cup_{i=1}^l P_i$  must be done), thus implying that no vertex in  $\cup_{i=1}^l P_i$  is affected by an optimal edition.
- **3-LEAF POWER EDGE-DELETION.** Firstly, observe that if  $N$  is not a clique, then every optimal 3-leaf power deletion in that case would have to destroy at least  $k + 1$  edge disjoint  $C_4$ 's with edges deletion only, which is impossible. The arguments used previously hold again in this case to conclude that  $N$  must become a critical clique in the modified graph.

□

Now, observe that the 2-branch reduction rule can be applied directly to 3-LEAF POWER EDGE-DELETION, but will not be safe for 3-LEAF POWER COMPLETION. Indeed, in the proof of Lemma 2.4, if we look at the cycle  $C$  of  $G$  containing vertices of  $B$ , we may have to delete edges between two consecutive critical cliques along  $C$  to transform  $\mathcal{C}(C)$  into a tree. Nevertheless, it is possible to bound the number of vertices of  $\text{path}(B)$  in the case of 3-LEAF POWER COMPLETION by looking at the edges completions required to make a cycle chordal (see Lemma 3.4).

**Lemma 3.3.** *Rule 2.5 is safe for 3-LEAF POWER EDGE-DELETION.*

**Proof.** Let  $F$  be an arbitrary optimal 3-leaf power deletion of  $G$ . We call  $C_1$  and  $C_2$  the critical cliques of  $H := G + F$  that respectively contain  $P_1$  and  $P_2$ , and set  $C'_1 := C_1 \setminus B^R$  and  $C'_2 := C_2 \setminus B^R$ . We will construct from  $F$  another optimal 3-leaf power deletion  $F'$  of  $G$  satisfying the conditions of Lemma 2.4.

We have two cases to consider : 1) either  $\text{path}(B)$  is disconnected in  $H$  or 2)  $\text{path}(B)$  is still connected in  $H$ . Case 1) works exactly as the first case studied in the proof of Lemma 2.4, and thus there exists an optimal 3-leaf power deletion on which conditions of Lemma 2.4 holds.

If case 2) holds, *i.e.* if  $\text{path}(B)$  is still connected in  $H$ , then  $P_1$  and  $P_2$  must belong to distinct connected components of  $H \setminus B^R$ , say  $X_1$  and  $X_2$  (otherwise  $H$  would admit a chordless cycle as induced subgraph). Furthermore, notice that we must have  $C_1 = P_1$  and

$C_2 = P_2$  in  $H$ . Indeed, if  $P_1$  forms a critical clique with some vertex  $x \in V(G) \setminus V(B^R)$ , this means  $F$  must contain  $V(P_1) \times V(Q_1)$ , which contradicts the hypothesis. Similarly, if  $P_1$  forms a critical clique with  $Q_1$ , then  $F$  must contain edges between  $Q_1$  and  $N_{B^R}(Q_1)$  which cannot be (the cases for  $P_2$  are symmetric). By definition,  $B^R$  is a 3-leaf power, and so are  $X_1$  and  $X_2$ . By Theorem 1.7, it follows that the composition of these three subgraphs:  $H' = (X_1, P_1) \otimes (B^R, Q_1)$  and  $H'' := (H', Q_2) \otimes (X_2, P_2)$  yields a 3-leaf power. The edge set  $F'$  used to obtain  $H'$  from  $G$  is a subset of  $F$  that respects conditions of Lemma 2.4, thus implying the lemma.  $\square$

The next lemma is useful to conclude on the size of the kernel in the 3-LEAF POWER COMPLETION problem.

**Lemma 3.4.** *Let  $G$  be a graph admitting a clean 2-branch  $B$  such that  $\text{path}(B)$  is composed by at least  $k+4$  critical cliques. If  $P_1$  and  $P_2$  belong to the same connected component in  $G \setminus B^R$ , then there is no 3-leaf power completion of size at most  $k$ .*

**Proof.** Let  $G$  be a graph with a clean 2-branch  $B$  on which conditions of the Lemma 3.4 apply, and let  $F$  be an optimal 3-leaf power completion of  $G$ . As  $P_1$  and  $P_2$  belong to the same connected component in  $G \setminus B^R$ , we have a cycle  $C$  of size at least  $k+4$  in  $\mathcal{C}(G)$ . Consider the subgraph of  $\mathcal{C}(G)$  induced by the critical cliques of  $C$ . By Lemma 1.4 we know that there exists  $F' \subseteq F$  such that  $\mathcal{C}(C + F')$  is a tree. It is known that  $F'$  is a triangulation of  $C$  [10]. Moreover, every triangulation of a  $n$ -cycle needs at least  $n-3$  chords, thus implying that  $|F'| > k$ , which is impossible.  $\square$

This result allows us to obtain a 2-branch reduction rule for the 3-LEAF POWER COMPLETION problem as well.

**Rule 3.1.** *Let  $G$  be a graph having a clean 2-branch  $B$  with attachment points  $P_1$  and  $P_2$  such that  $\text{path}(B)$  is composed by at least  $k+4$  critical cliques.*

- *if  $P_1$  and  $P_2$  belong to the same connected component in  $G \setminus B^R$ , then there is no completion of size at most  $k$ .*
- *otherwise, remove from  $G$  all the vertices of  $V(B^R)$  except those of  $V(Q_1 \cup Q_2)$  and add all possible edges between  $Q_1$  and  $Q_2$ .*

**Proof.** The first point follows directly from Lemma 3.4. To see the second point, we need to show that if  $P_1$  and  $P_2$  belong to different connected components in  $G \setminus B^R$ , then there exists an optimal 3-leaf power completion that affects no vertex in  $V(B^R) \setminus V(Q_1 \cup Q_2)$ . To show this, assume that  $F$  is an optimal 3-leaf power completion of  $G$  and let  $C_1, C_2$  be the critical cliques containing  $P_1, P_2$  in  $H := G + F$ . Notice that  $P_1$  and  $P_2$  belong to different connected components in  $H \setminus B^R$ , otherwise we show that, as in the proof of Lemma 3.4,  $F$  has to triangulate a cycle of length at least  $k+4$ , thus contradicting the assumption  $|F| \leq k$ . Now, consider the subgraphs  $H_1$  being the connected component of  $H \setminus B^R$  containing  $P_1$ , and  $H_2$  being the one containing  $P_2$ . By Theorem 1.7 and

Observation 1.8,  $H' := (H_1, C'_1) \otimes (B^R, Q_1)$  where  $C'_1 := C_1 \setminus B^R$  is a 3-leaf power. With a similar argument, we can show that  $H'' := (H_2, C'_2) \otimes (H', Q_2)$ , where  $C'_2 := C_2 \setminus B^R$ , is a 3-leaf power. The completion used to obtain  $H''$  from  $G$  is a subset of  $F$  respecting conditions of Rule 3.1, thus implying the result.  $\square$

**Theorem 3.5.** *The parameterized 3-LEAF POWER COMPLETION and 3-LEAF POWER EDGE-DELETION problem admit kernels with  $O(k^3)$  vertices. Given a graph  $G$  a reduced instance can be computed in linear time.*

**Proof.** We detail separately completion and deletion.

- **3-LEAF POWER COMPLETION.** As in the proof of Theorem 2.5, we consider  $H := G + F$  with  $G$  being reduced and  $F$  being an optimal completion and we denote by  $T$  the minimal subtree of  $\mathcal{C}(H)$  spanning the set of affected critical cliques  $A$ . As noticed before, we have  $|A| \leq 2k$ .

First, remark that the only difference between this case and 3-LEAF POWER EDITION concerns the 2-branch reduction rule. This means that the only difference will occur in the number of vertices of the paths resulting from the removal of  $A$  and  $A'$  in  $T$  ( $A'$  being critical cliques of degree at least 3 in  $T$ ). Due to Lemma 3.4 and Rule 3.1 we know that a 2-branch in  $\mathcal{C}(H)$  is made of at most  $2k + 6$  critical cliques, corresponding to a path of at most  $k + 4$  critical cliques, each one (excepted the terminal ones) having a pendant critical clique (Rule 2.3). This means that  $\mathcal{C}(H)$  contains at most  $4k \cdot (2k + 6) + 2k \cdot 2 + 2k \cdot (4k + 3) = 16k^2 + 34k$  critical cliques. By Lemma 1.3, we know that each edited edge creates at most 4 new critical cliques. It follows that  $\mathcal{C}(G)$  contains at most  $16k^2 + 38k$  vertices. By Rule 2.2, each critical clique of the reduced graph has size at most  $k + 1$ , thus implying that the reduced graph contains at most  $16k^3 + 54k^2 + 38k$  vertices, proving the  $O(k^3)$  kernel size.

- **3-LEAF POWER EDGE-DELETION.** The rules used for the 3-LEAF POWER EDGE-DELETION problem are exactly the same than the one used to obtain a cubic kernel for 3-LEAF POWER EDITION. Thus, the size of a reduced instance of 3-LEAF POWER EDGE-DELETION will be exactly the same as one of a reduced instance of 3-LEAF POWER EDITION.

$\square$

#### 4. Conclusion

By proving the existence of a kernel with  $O(k^3)$  vertices for the 3-LEAF POWER EDITION problem, we positively answered an open problem [11, 9]. The natural question is now whether the cubic bound could be improved. A strategy could be, as for the quadratic kernel of 3-HITTING SET [22] which is based on the linear kernel of VERTEX COVER [20], to consider the following subproblem:

PARAMETERIZED FAT STAR EDITION PROBLEM

**Input:** An undirected graph  $G = (V, E)$ .

**Parameter:** An integer  $k \geq 0$ .

**Question:** Is there a subset  $F \subseteq V \times V$  with  $|F| \leq k$  such that the graph  $G + F = (V, E \triangle F)$  is a 3-leaf power and its critical clique graph  $\mathcal{C}(G + F)$  is a star (we say that  $G + F$  is a *fat star*).

It can be shown that small modifications of the Rule 2.1, 2.2 and 2.4 yield a kernel with  $O(k^2)$  vertices for the FAT STAR EDITION problem [23]. A linear bound may be helpful to improve the kernel of the 3-LEAF POWER EDITION since it can be shown that the neighborhood of each big enough critical clique of the input graph as to be edited into a fat star.

#### References

- [1] J-P. Barthélémy and A. Guénoche. *Trees and proximity representations*. John Wiley & Sons, 1991.
- [2] S. Böcker, S. Briesemeister, and G. W. Klau. Exact algorithms for cluster editing: evaluation and experiments. In *International Workshop on Experimental Algorithms (WEA)*, volume 5038 of *Lecture Notes in Computer Science*, pages 289–302, 2008.
- [3] H. Bodlaender, R.G. Downey, M.R. Fellows, and D. Hermelin. On problems without polynomial kernels (extended abstract). In *Automata, Languages and Programming, 35th International Colloquium (ICALP)*, volume 5125 of *Lecture Notes in Computer Science*, pages 563–574, 2008.
- [4] A. Brandstädt and V.B. Le. Structure and linear time recognition of 3-leaf powers. *Information Processing Letters*, 98(4):133–138, 2006.
- [5] A. Brandstädt, V.B. Le, and R. Sritharan. Structure and linear time recognition of 4-leaf powers. *ACM Transactions on Algorithms (TALG)*, 5(1):1–22, 2008.
- [6] M-S. Chang and M-T. Ko. The 3-steiner root problem. In *WG*, pages 109–120, 2007.
- [7] F. Dehne, M. Langston, X. Luo, S. Pitre, P. Shaw, and Y. Zhang. The cluster editing problem: implementations and experiments. In *International Workshop on Parameterized and Exact Computation (IWPEC)*, volume 4169 of *Lecture Notes in Computer Science*, pages 13–24, 2006.

- [8] M. Dom, J. Guo, F. Hüffner, and R. Niedermeier. Error compensation in leaf root problems. In *International Symposium on Algorithms and Computation (ISAAC)*, number 3341 in Lecture Notes in Computer Science, pages 389–401, 2004.
- [9] M. Dom, J. Guo, F. Hüffner, and R. Niedermeier. Extending the tractability border for closest leaf powers. In *International Workshop on Graph Theoretical Concepts in Computer Science (WG)*, number 3787 in Lecture Notes in Computer Science, pages 397–408, 2005.
- [10] M. Dom, J. Guo, F. Hüffner, and R. Niedermeier. Error compensation in leaf root problems. *Algorithmica*, 44:363–381, 2006.
- [11] M. Dom, J. Guo, F. Hüffner, and R. Niedermeier. Closest 4-leaf power is fixed-parameter tractable. *Discrete Applied Mathematics*, 156(18):3345–3361, 2008.
- [12] M.R. Fellows, M. Langston, F. Rosamond, and P. Shaw. Efficient parameterized preprocessing for cluster editing. In *International Symposium on Fundamentals of Computation Theory (FCT)*, number 4639 in Lecture Notes in Computer Science, pages 312–321, 2007.
- [13] J. Gramm, J. Guo, F. Hüffner, and R. Niedermeier. Graph-modeled data clustering: Exact algorithms for clique generation. *Theory Comput. Syst.*, 38(4):373–392, 2005.
- [14] J. Guo. A more effective linear kernelization for cluster editing. In *International Symposium Combinatorics, Algorithms, Probabilistic and Experimental Methodologies (ESCAPE)*, volume 4614 of *Lecture Notes in Computer Science*, pages 36–47, 2007.
- [15] J. Guo and R. Niedermeier. Invitation to data reduction and problem kernelization. *SIGACT News*, 38(1):31–45, 2007.
- [16] P. Hansen and B. Jaumard. Cluster analysis and mathematical programming. *Math. Program.*, 79(1-3):191–215, 1997.
- [17] P. Kearney and D. Corneil. Tree powers. *Journal of Algorithms*, 29(1):111–131, 1998.
- [18] G.H. Lin, P.E. Kearney, and T. Jiang. Phylogenetic k-root and steiner k-root. In *International Symposium on Algorithms and Computation (ISAAC)*, number 1969 in Lecture Notes in Computer Science, pages 539–551, 2000.
- [19] A. Ntazon, R. Shamir, and R. Sharan. Complexity classification of some edge modification problems. *Discrete Applied Mathematics*, 113:109–128, 2001.
- [20] R. Niedermeier. *Invitation to fixed parameter algorithms*, volume 31 of *Oxford Lectures Series in Mathematics and its Applications*. Oxford University Press, 2006.
- [21] N. Nishimura, P. Ragde, and D. Thilikos. On graph powers for leaf-labeled trees. *Journal of Algorithms*, 42(1):69–108, 2002.

- [22] N. Nishimura, P. Ragde, and D. Thilikos. Smaller kernels for hitting set problems of constant arity. In *International Workshop on Parameterized and Exact Computation (IWPEC)*, number 3162 in Lecture Notes in Computer Science, pages 121–126, 2004.
- [23] A. Perez. Complexité paramétrique et recherche de noyaux. Master’s thesis, Université de Montpellier 2, France, 2008.
- [24] F. Protti, M. Dantas da Silva, and J.L. Szwarcfiter. Applying modular decomposition to parameterized cluster editing problems. *Theory of Computing Systems*, 2007.
- [25] R. Shamir, R. Sharan, and D. Tsur. Cluster graph modification problems. *Discrete Applied Mathematics*, 144(1-2):173–182, 2004.
- [26] M. Tedder, D. Corneil, M. Habib, and C. Paul. Simpler linear-time modular decomposition via recursive factorizing permutations. In *Automata, Languages and Programming (ICALP)*, volume 5115 of *Lecture Notes in Computer Science*, pages 634–645, 2008.

## 2.8 A poly-kernel for proper interval completion

### Polynomial kernels for PROPER INTERVAL COMPLETION and related problems \*

Stéphane Bessy    Anthony Perez  
 bessy@lirmm.fr    perez@lirmm.fr  
 LIRMM – Université Montpellier II - FRANCE

October 30, 2011

#### Abstract

Given a graph  $G = (V, E)$  and a positive integer  $k$ , the PROPER INTERVAL COMPLETION problem asks whether there exists a set  $F$  of at most  $k$  pairs of  $(V \times V) \setminus E$  such that the graph  $H = (V, E \cup F)$  is a proper interval graph. The PROPER INTERVAL COMPLETION problem finds applications in molecular biology and genomic research [16, 24]. First announced by Kaplan, Tarjan and Shamir in FOCS '94, this problem is known to be FPT [16], but no polynomial kernel was known to exist. We settle this question by proving that PROPER INTERVAL COMPLETION admits a kernel with at most  $O(k^3)$  vertices. Moreover, we prove that a related problem, the so-called BIPARTITE CHAIN DELETION problem, admits a kernel with at most  $O(k^2)$  vertices, completing a previous result of Guo [13].

#### Introduction

The aim of a graph modification problem is to transform a given graph in order to get a certain property  $\Pi$  satisfied. Several types of transformations can be considered: for instance, in *vertex deletion* problems, we are only allowed to delete vertices from the input graph, while in *edge modification problems* the only allowed operation is to modify the edge set of the input graph. The optimization version of such problems consists in finding a *minimum* set of edges (or vertices) whose modification makes the graph satisfy the given property  $\Pi$ . Graph modification problems cover a broad range of NP-Complete problems and have been extensively studied in the literature [20, 23, 24]. Well-known examples include the VERTEX COVER [8], FEEDBACK VERTEX SET [26], or CLUSTER EDITING [5] problems. These problems find applications in various domains, such as computational biology [16, 24], image processing [23] or relational databases [25].

A natural approach to deal with such problems is to measure their difficulty with respect to some parameter such as, for instance, the number of allowed modifications. *Parameterized complexity* provides a useful theoretical framework to that aim [10, 21]. A problem *parameterized* by some integer  $k$  is said to be *fixed-parameter tractable* (FPT for short) whenever it can be solved in time  $f(k) \cdot n^c$  for some constant  $c > 0$ , where  $n$  is the size of the instance (for problems on graphs, usually,  $n$  is the number of vertices of the input graph). A natural parameterization for graph modification problems thereby consists in the number of allowed transformations. As one of the most powerful technique to design fixed-parameter algorithms, *kernelization algorithms* have

---

\*Research supported by the AGAPE project (ANR-09-BLAN-0159).

been extensively studied in the last decade (see [2] for a survey). A *kernelization algorithm* is a polynomial-time algorithm (called *reduction rules*) that given an instance  $(I, k)$  of a parameterized problem  $P$  computes an instance  $(I', k')$  of  $P$  such that (i)  $(I, k)$  is a YES-instance if and only if  $(I', k')$  is a YES-instance and (ii)  $|I'| \leq h(k)$  for some computable function  $h()$  and  $k' \leq k$ . The instance  $(I', k')$  is called the *kernel* of  $P$ . We say that  $(I', k')$  is a *polynomial kernel* if the function  $h()$  is a polynomial. It is well-known that a decidable parameterized problem is FPT if and only if it has a kernelization algorithm [21]. But this equivalence only yields kernels of super-polynomial size. To design efficient fixed-parameter algorithms, a kernel of small size - polynomial (or even linear) in  $k$  - is highly desirable [22]. However, recent results give evidence that not every parameterized problem admits a polynomial kernel, unless  $NP \subseteq coNP/poly$  [3]. On the positive side, notable kernelization results include a less-than- $2k$  kernel for VERTEX COVER [8], a  $4k^2$  kernel for FEEDBACK VERTEX SET [26] and a  $2k$  kernel for CLUSTER EDITING [5].

We follow this line of research with respect to graph modification problems. It has been shown that a graph modification problem is FPT whenever  $\Pi$  is hereditary and can be characterized by a finite set of forbidden induced subgraphs [4]. However, recent results proved that several graph modification problems do not admit a polynomial kernel even for such properties  $\Pi$  [12, 18]. In this paper, we are in particular interested in *completion* problems, where the only allowed operation is to add edges to the input graph. We consider the property  $\Pi$  as being the class of *proper interval graphs*. This class is a well-studied class of graphs, and several characterizations are known to exist [19, 30]. In particular, there exists an *infinite* set of forbidden induced subgraphs that characterizes proper interval graphs [30] (see Figure 1). More formally, we consider the following problem:

PROPER INTERVAL COMPLETION:

**Input:** A graph  $G = (V, E)$  and a positive integer  $k$ .

**Parameter:**  $k$ .

**Output:** A set  $F$  of at most  $k$  pairs of  $(V \times V) \setminus E$  such that the graph  $H = (V, E \cup F)$  is a proper interval graph.

Interval completion problems find applications in molecular biology and genomic research [15, 16], and in particular in *physical mapping* of DNA. In this case, one is given a set of long contiguous intervals (called *clones*) together with experimental information on their pairwise overlaps, and the goal is to reconstruct the relative position of the clones along the target DNA molecule. We focus here on the particular case where all intervals have equal length, which is a biologically important case (e.g. for cosmid clones [15]). In the presence of (a small number of) unidentified overlaps, the problem becomes equivalent to the PROPER INTERVAL COMPLETION problem. It is known to be NP-Complete for a long time [11], but fixed-parameter tractable due to a result of Kaplan, Tarjan and Shamir in FOCS '94 [16, 17]<sup>1</sup>. The fixed-parameter tractability of the PROPER INTERVAL COMPLETION can also be seen as a corollary of a characterization of Wegner [30] combined with Cai's result [4]. Nevertheless, it was not known whether this problem admits a polynomial kernel or not.

**Our results** We prove that the PROPER INTERVAL COMPLETION problem admits a kernel with at most  $O(k^3)$  vertices. To that aim, we identify *nice* parts of the graph that induce proper interval graphs and can hence be safely reduced. Moreover, we apply our techniques to the so-called BIPARTITE CHAIN DELETION problem, closely related to the PROPER INTERVAL COMPLETION problem where one is given a graph  $G = (V, E)$  and seeks a set of at most  $k$  edges whose deletion

<sup>1</sup>Notice also that the *vertex deletion* of the problem is fixed-parameter tractable [28].



from  $E$  results in a bipartite chain graph (a graph that can be partitioned into two independent sets connected by a join). We obtain a kernel with  $O(k^2)$  vertices for this problem. This result completes a previous result of Guo [13] who proved that the BIPARTITE CHAIN DELETION WITH FIXED BIPARTITION problem admits a kernel with  $O(k^2)$  vertices.

**Outline** We begin with some definitions and notations regarding proper interval graphs. Next, we give the reduction rules the application of which leads to a kernelization algorithm for the PROPER INTERVAL COMPLETION problem. These reduction rules allow us to obtain a kernel with at most  $O(k^3)$  vertices. Finally, we prove that our techniques can be applied to BIPARTITE CHAIN DELETION to obtain a quadratic-vertex kernel.

## 1 Preliminaries

### 1.1 Proper interval graphs

We consider simple, loopless, undirected graphs  $G = (V(G), E(G))$  where  $V(G)$  denotes the vertex set of  $G$  and  $E(G)$  its edge set<sup>2</sup>. Given a vertex  $v \in V$ , we use  $N_G(v)$  to denote the *open neighborhood* of  $v$  and  $N_G[v] = N_G(v) \cup \{v\}$  for its *closed neighborhood*. Two vertices  $u$  and  $v$  are *true twins* if  $N[u] = N[v]$ . If  $u$  and  $v$  are not true twins but  $uv \in E$ , we say that a vertex of  $N[u] \Delta N[v]$  *distinguishes*  $u$  and  $v$ . Given a subset of vertices  $S \subseteq V$ ,  $N_S(v)$  denotes the set  $N_G(v) \cap S$  and  $N_G(S)$  denotes the set  $(\cup_{s \in S} N_G(s)) \setminus S$ . Moreover,  $G[S]$  denotes the subgraph *induced* by  $S$ , i.e.  $G[S] = (S, E_S)$  where  $E_S = \{uv \in E : u, v \in S\}$ . A *join* in a graph  $G = (V, E)$  is a bipartition  $(X, Y)$  of  $G$  and an order  $x_1, \dots, x_{|X|}$  on  $X$  such that for all  $i = 1, \dots, |X| - 1$ ,  $N_Y(x_i) \subseteq N_Y(x_{i+1})$ . The edges between  $X$  and  $Y$  are called the *edges of the join*, and a subset  $F \subseteq E$  is said to *form a join* if  $F$  corresponds to the edges of a join of  $G$ . Finally, a graph is an *interval graph* if it admits a representation on the real line such that: (i) the vertices of  $G$  are in bijection with intervals of the real line and (ii)  $uv \in E$  if and only if  $I_u \cap I_v \neq \emptyset$ , where  $I_u$  and  $I_v$  denote the intervals associated to  $u$  and  $v$ , respectively. Such a graph is said to admit an *interval representation*. A graph is a *proper interval graph* if it admits an interval representation such that  $I_u \not\subset I_v$  for every  $u, v \in V$ . In other words, no interval strictly contains another interval.

We will make use of the two following characterizations of proper interval graphs to design our kernelization algorithm.

**Theorem 1.1** (Forbidden subgraphs [30]). *A graph is a proper interval graph if and only if it does not contain any  $\{\text{hole}, \text{claw}, \text{net}, 3\text{-sun}\}$  as an induced subgraph (see Figure 1).*

The claw graph is the bipartite graph  $K_{1,3}$ . Denoting its bipartition by  $(\{c\}, \{l_1, l_2, l_3\})$ , we call  $c$  the *center* and  $\{l_1, l_2, l_3\}$  the *leaves* of the claw.

**Theorem 1.2** (Umbrella property [19]). *A graph is a proper interval graph if and only if its vertices admit an ordering  $\sigma$  (called umbrella ordering) satisfying the following property: given  $v_i v_j \in E$  with  $i < j$  then  $v_i v_l, v_l v_j \in E$  for every  $i < l < j$  (see Figure 2).*

In the following, we associate an umbrella ordering  $\sigma_G$  to any proper interval graph  $G = (V, E)$ . There are several things to remark. First, note that in an umbrella ordering  $\sigma_G$  of a graph  $G$ , every maximal set of true twins of  $G$  is consecutive. Moreover, it is known [9] that  $\sigma_G$  is unique up to permutation of true twins of  $G$  or by reversal of the ordering induced on a connected component of

<sup>2</sup>In all our notations, we forget the mention to the graph  $G$  whenever the context is clear.

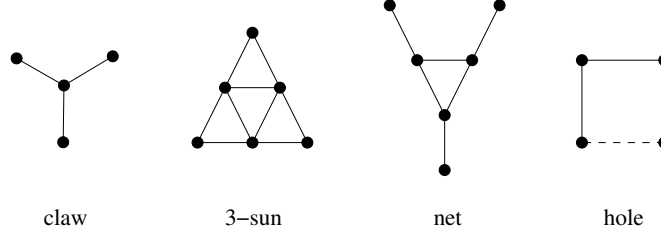


Figure 1: The forbidden induced subgraphs of proper interval graphs. A *hole* is an induced cycle of length at least 4.

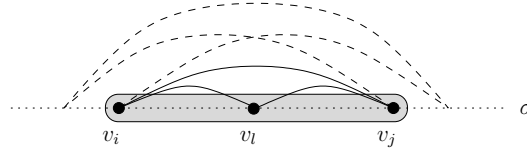


Figure 2: Illustration of the umbrella property. The edge  $v_i v_j$  is extremal.<sup>3</sup>

$G$ . Remark also that for any edge  $uv$  with  $u <_{\sigma_G} v$ , the set  $\{w \in V : u \leq_{\sigma_G} w \leq_{\sigma_G} v\}$  is a clique of  $G$ , and for every  $i$  with  $1 \leq i < l$ ,  $(\{v_1, \dots, v_i\}, \{v_{i+1}, \dots, v_n\})$  is a join of  $G$ .

According to this ordering, we say that an edge  $uv$  is *extremal* if there does not exist any edge  $u'v'$  different from  $uv$  such that  $u' \leq_{\sigma_G} u$  and  $v \leq_{\sigma_G} v'$  (see Figure 2).

Let  $G = (V, E)$  be an instance of PROPER INTERVAL COMPLETION. A *completion* of  $G$  is a set  $F \subseteq (V \times V) \setminus E$  such that the graph  $H = (V, E \cup F)$  is a proper interval graph. In a slight abuse of notation, we use  $G + F$  to denote the graph  $H$ . A *k-completion* of  $G$  is a completion such that  $|F| \leq k$ , and an *optimal completion*  $F$  is such that  $|F|$  is minimum. We say that  $G = (V, E)$  is a *positive* instance of PROPER INTERVAL COMPLETION whenever it admits a  $k$ -completion. We state a simple observation that will be very useful for our kernelization algorithm.

**Observation 1.3.** *Let  $G = (V, E)$  be a graph and  $F$  be an optimal completion of  $G$ . Given an umbrella ordering  $\sigma$  of  $G + F$ , any extremal edge of  $\sigma$  is an edge of  $G$ .*

*Proof.* Assume that there exists an extremal edge  $e$  in  $\sigma$  that belongs to  $F$ . By definition,  $\sigma$  is still an umbrella ordering if we remove the edge  $e$  from  $F$ , contradicting the optimality of  $F$ .  $\square$

## 1.2 Branches

We now give the main definitions of this Section. The branches that we will define correspond to some parts of the graph that already behave like proper interval graphs. They are the parts of the graph that we will reduce in order to obtain a kernelization algorithm.

**Definition 1.4** (1-branch). *Let  $B \subseteq V$ . We say that  $B$  is a 1-branch if the following properties hold (see Figure 3):*

<sup>3</sup>In all the figures, (non-)edges between blocks stand for all the possible (non-)edges between the vertices that lie in these blocks, and the vertices within a gray box form a clique of the graph.

- (i) The graph  $G[B]$  is a connected proper interval graph admitting an umbrella ordering  $\sigma_B = b_1, \dots, b_{|B|}$  and,
- (ii) The vertex set  $V \setminus B$  can be partitioned into two sets  $R$  and  $C$  with: no edges between  $B$  and  $C$ , every vertex in  $R$  has a neighbor in  $B$ , no edges between  $\{b_1, \dots, b_{l-1}\}$  and  $R$  where  $b_l$  is the neighbor of  $b_{|B|}$  with minimal index in  $\sigma_B$ , and for every  $l \leq i < |B|$ , we have  $N_R(b_i) \subseteq N_R(b_{i+1})$ .

We denote by  $B_1$  the set of vertices  $\{v \in V : b_l \leq_{\sigma_B} v \leq_{\sigma_B} b_{|B|}\}$ , which is a clique (because  $b_l$  is a neighbor of  $b_{|B|}$ ). This set is exactly the neighborhood of  $b_{|B|}$  in  $B$ . We call  $B_1$  the *attachment clique* of  $B$ , and use  $B^R$  to denote  $B \setminus B_1$ .

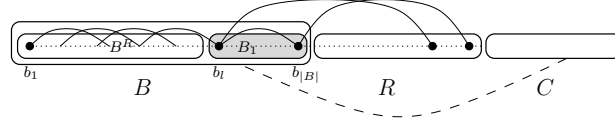


Figure 3: A 1-branch of a graph  $G = (V, E)$ . The vertices of  $B$  are ordered according to the umbrella ordering  $\sigma_B$ .

**Definition 1.5** (2-branch). Let  $B \subseteq V$ . We say that  $B$  is a 2-branch if the following properties hold (see Figure 4):

- (i) The graph  $G[B]$  is a connected proper interval graph admitting an umbrella ordering  $\sigma_B = b_1, \dots, b_{|B|}$  and,
- (ii) The vertex set  $V \setminus B$  can be partitioned into sets  $L, R$  and  $C$  with:
- no edges between  $B$  and  $C$ ,
  - every vertex in  $L$  (resp.  $R$ ) has a neighbor in  $B$ ,
  - no edges between  $\{b_1, \dots, b_{l-1}\}$  and  $R$  where  $b_l$  is the neighbor of  $b_{|B|}$  with minimal index in  $\sigma_B$ ,
  - no edges between  $\{b_{l'+1}, \dots, b_{|B|}\}$  and  $L$  where  $b_{l'}$  is the neighbor of  $b_1$  with maximal index in  $\sigma_B$  and,
  - $N_R(b_i) \subseteq N_R(b_{i+1})$  for every  $l \leq i < |B|$  and  $N_L(b_{i+1}) \subseteq N_L(b_i)$  for every  $1 \leq i < l'$ .

Again, we denote by  $B_1$  (resp.  $B_2$ ) the set of vertices  $\{v \in V : b_1 \leq_{\sigma_B} v \leq_{\sigma_B} b_{l'}\}$  (resp.  $\{v \in V : b_l \leq_{\sigma_B} v \leq_{\sigma_B} b_{|B|}\}$ ). We call  $B_1$  and  $B_2$  the *attachment cliques* of  $B$ , and use  $B^R$  to denote  $B \setminus (B_1 \cup B_2)$ . We assume that  $L \neq \emptyset$  and  $R \neq \emptyset$ , otherwise  $B$  is a 1-branch. Finally, when  $B^R = \emptyset$ , it is possible that a vertex of  $L$  or  $R$  is adjacent to all the vertices of  $B$ . In this case, we will denote by  $N$  the set of vertices that are adjacent to every vertex of  $B$ , remove them from  $R$  and  $L$  and abusively still denote by  $L$  (resp.  $R$ ) the set  $L \setminus N$  (resp.  $R \setminus N$ ). We will precise when we need to use the set  $N$ .

In both cases, in a 1- or 2-branch, whenever the proper interval graph  $G[B]$  is a *clique*, we say that  $B$  is a  $K$ -join. Observe that, in a 1- or 2-branch  $B$ , for any extremal edge  $uv$  in  $\sigma_B$ , the set of vertices  $\{w \in V : u \leq_{\sigma_B} w \leq_{\sigma_B} v\}$  defines a  $K$ -join. In particular, this means that a branch can

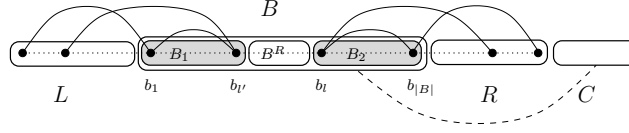


Figure 4: A 2-branch of a graph  $G = (V, E)$ . The vertices of  $B$  are ordered according to the umbrella ordering  $\sigma_B$ .

be decomposed into a sequence of  $K$ -joins. Observe however that the decomposition is not unique: for instance, the  $K$ -joins corresponding to all the extremal edges of  $\sigma_B$  are not disjoint. We will precise in Section 2.1.5, when we will reduce the size of 2-branches, how to fix a decomposition. Finally, we say that a  $K$ -join is *clean* whenever its vertices are not contained in any claw or 4-cycle. Remark that a subset of a  $K$ -join (resp. clean  $K$ -join) is also a  $K$ -join (resp. clean  $K$ -join).

## 2 Kernel for PROPER INTERVAL COMPLETION

The basic idea of our kernelization algorithm is to detect the large enough branches and then to reduce them. This section details the rules we use for that.

### 2.1 Reduction rules

#### 2.1.1 Basic rules

We say that a rule is *safe* if when it is applied to an instance  $(G, k)$  of the problem,  $(G, k)$  admits a  $k$ -completion if, and only if, the instance  $(G', k')$  reduced by the rule admits a  $k'$ -completion.

The first reduction rule gets rid of connected components that are already proper interval graphs. This rule is trivially safe and can be applied in  $O(n + m)$  time using any recognition algorithm for proper interval graphs [6].

**Rule 2.1** (Connected components). *Remove any connected component of  $G$  that is a proper interval graph.*

The following reduction rule can be applied since proper interval graphs are closed under true twin addition and induced subgraphs. For a class of graphs satisfying these two properties, we know that this rule is safe [1] (roughly speaking, we edit all the large set of true twins in the same way). Furthermore, it is possible to compute every set of pairwise true twins using a modular decomposition algorithm or more easily, partition refinement (see [14] for example).

**Rule 2.2** (True twins [1]). *Let  $T$  be a set of true twins in  $G$  such that  $|T| > k$ . Remove  $|T| - (k + 1)$  arbitrary vertices from  $T$ .*

We also use the classical *sunflower* rule, allowing to identify a set of edges that must be added in any optimal completion.

**Rule 2.3** (Sunflower). *Let  $\mathcal{S} = \{C_1, \dots, C_m\}$ ,  $m > k$  be a set of claws having two leaves  $u, v$  in common but distinct third leaves. Add  $uv$  to  $F$  and decrease  $k$  by 1.*

*Let  $\mathcal{S} = \{C_1, \dots, C_m\}$ ,  $m > k$  be a set of distinct 4-cycles having a non-edge  $uv$  in common. Add  $uv$  to  $F$  and decrease  $k$  by 1.*

**Lemma 2.1.** *Rule 2.3 is safe and can be carried out in polynomial time. More precisely, it is possible to detect all the 4-cycles and claws of  $G$  in time  $O(n^2m)$ .*

*Proof.* We only prove the first rule. The second rule can be proved similarly. Let  $F$  be a  $k$ -completion of  $G$  and assume that  $F$  does not contain  $(u, v)$ . Since any two claws in  $\mathcal{S}$  only share  $(u, v)$  as a common non-edge,  $F$  must contain one edge for every  $C_i$ ,  $1 \leq i \leq m$ . Since  $m > k$ , we have  $|F| > k$ , which cannot be. Now, we briefly indicate how to compute all claws and the 4-cycles of  $G$ . For every edge  $xy$  of  $G$ , in time  $O(n)$ , we compute the sets  $N_x = N_G(x) \setminus N_G[y]$  and  $N_y = N_G(y) \setminus N_G[x]$ . Each edge  $uv$  between  $N_x$  and  $N_y$  correspond to the 4-cycle  $xyvu$ . So, in time  $O(m \cdot (n + m))$  (less than  $O(n^2m)$ ), we enumerate all the 4-cycles of  $G$ . On the other hand, for every vertex  $x$  of  $G$ , we compute all the three cycle in  $H_x$ , the complementary of  $G[N_G(x)]$ , what can be done in time  $O(n(H_x)m(H_x))$  (for instance, by computing for every vertex  $y$  of  $H_x$ , a breadth search tree rooted on  $y$ ). This gives all the claws with center  $x$ . And, in all, we enumerate all the claws of  $G$  in time  $O(n^2m)$ . Finally, sparsing the claws and the 4-cycles, it is then easy to detect the sunflowers.  $\square$

### 2.1.2 Number of vertices in claws or 4-cycles

The general idea of our process is to reduce the size of the branches. However, we realized that is not always possible, even for  $K$ -join. We will see that this problem is due to the presence of claws or 4-cycles intersecting the branches. So, in this part, we give a bound of the number of vertices belonging to these obstructions in a positive instance of PROPER INTERVAL COMPLETION.

**Lemma 2.2.** *Let  $G = (V, E)$  be a positive instance of PROPER INTERVAL COMPLETION on which Rule 2.3 has been applied. There are at most  $k^2$  claws with distinct sets of leaves, and at most  $k^2 + 2k$  vertices of  $G$  are leaves of claw. Furthermore, there are at most  $2k^2 + 2k$  vertices of  $G$  that are vertices of a 4-cycle.*

*Proof.* As  $G$  is a positive instance of PROPER INTERVAL COMPLETION, every claw or 4-cycle of  $G$  has a non-edge that will be completed and then is an edge of  $F$ . Let  $xy$  be an edge of  $F$ . As we have applied Rule 2.3 on  $G$ , there are at most  $k$  vertices in  $G$  that form the three leaves of a claw with  $x$  and  $y$ . So, at most  $(k + 2)k$  vertices of  $G$  are leaves of claws. Similarly, there are at most  $k$  non-edges of  $G$ , implying at most  $2k$  vertices, that form a 4-cycle with  $x$  and  $y$ . So, at most  $(2k + 2)k$  vertices of  $G$  are in a 4-cycle.  $\square$

**Lemma 2.3.** *Let  $G = (V, E)$  be a positive instance of PROPER INTERVAL COMPLETION on which Rule 2.2 and Rule 2.3 have been applied. There are at most  $4k^3 + 15k^2 + 16k$  vertices of  $G$  that belong to a claw or a 4-cycle.*

*Proof.* As  $G$  is a positive instance of PROPER INTERVAL COMPLETION, there exists a set  $F$  of at most  $k$  edges such that  $G + F$  is a proper interval graph and admits an umbrella ordering  $\sigma$ . We contract all the set of true twins of  $G$  and denote by  $G'$  the obtained graph. Remark that, as Rule 2.2 has been applied on  $G$ , every contracted set has size at most  $k + 1$ . As  $G'$  is also an induced subgraph of  $G$ , we denote by  $\sigma'$  the order induced by  $\sigma$  on  $G'$ . Now, we define  $C$  to be the vertices of  $G'$  which are center of a claw in  $G'$ , not incident to any edge of  $F$ , are not contained in a 4-cycle neither a leaf of a claw. We sort this set according to  $\sigma'$  and denote by  $c_1, \dots, c_l$  its vertices in this order. As the vertices of  $C$  are not incident with edges of  $F$ , the edges incident with vertices of  $C$  respect the umbrella property. We look for distinct vertices which distinguish the pairs of consecutive vertices of  $C$ . Remark that

it is possible that two consecutive vertices of  $C$ ,  $c_i$  and  $c_{i+1}$  are twins, but not true twins. In this case, we can identify all the neighbors of  $c_i$  and  $c_{i+1}$ . Indeed, assume that  $c_i$  and  $c_{i+1}$  are not linked but that they have same neighborhood. Then,  $c_i$  has no neighbor  $x$  with  $x <_{\sigma'} c_i$ , otherwise  $x$  is also a neighbor of  $c_{i+1}$  and  $c_i$  and  $c_{i+1}$  would be neighbors, by the umbrella property. As  $c_i$  is not an isolated vertex, it has at least one neighbor. So, let  $x$  be the neighbor of  $c_i$  with maximal index in  $\sigma'$ . As  $c_i$  and  $c_{i+1}$  are not linked, then  $x <_{\sigma'} c_{i+1}$ . So, let  $Y$  denotes the set  $\{y \in G' : x <_{\sigma'} y <_{\sigma'} c_{i+1}\}$ . If  $Y \neq \emptyset$ , as  $x$  and  $c_{i+1}$  are linked by an edge, then  $Y$  is a set of neighbors of  $c_{i+1}$  and then a set of neighbors of  $c_i$  also, what contradicts the choice of  $x$ . So,  $Y = \emptyset$ , and  $c_{i+1}$  is the first non-neighbor of  $c_i$  after  $c_i$  according to  $\sigma'$ . Similarly,  $c_i$  is the last non-neighbor of  $c_{i+1}$  before  $c_{i+1}$  according to  $\sigma'$ , and we conclude that  $N_{G'}(c_i) = N_{G'}(c_{i+1}) = \{x \in G' : c_i <_{\sigma'} x <_{\sigma'} c_{i+1}\}$ . So,  $c_i$  and  $c_{i-1}$  cannot be twins, so it is for  $c_{i+1}$  and  $c_{i+2}$ . It means that we can remove at most half of  $c_i$  and obtain  $C' = \{c'_1, \dots, c'_p\}$  (with  $p \geq l/2$ ), a subset of  $C$ , sorted according to  $\sigma'$ , in which every pair of consecutive vertices is not made of twins.

Now, let  $x$  be a vertex of  $G'$ . As no vertex of  $C'$  are incident to an edge of  $F$ , it means that the neighborhood of  $x$  in  $C'$  is consecutive according to the order  $c'_1, \dots, c'_p$ . Then,  $x$  distinguishes at most two pairs  $\{c'_i, c'_{i+1}\}$ , for  $1 \leq i \leq p-1$ . So, for  $1 \leq i \leq p-1$ , we choose  $d_i$  a vertex of  $G'$  which distinguishes  $c'_i$  from  $c'_{i+1}$ . If, amongst all the vertices of  $G'$  which distinguishes  $c'_i$  from  $c'_{i+1}$ , one is the leaf of a claw, we preferably choose it for  $d_i$ . As seen previously, it is possible that a vertex has been chosen twice to be a vertex  $d_i$ , but no more than two times. So, the set  $\{d_1, \dots, d_{p-1}\}$  contains at least  $(p-1)/2$  distinct vertices which we denote by  $d'_1, \dots, d'_q$  sorted according to  $\sigma'$ , and with  $q \geq (p-1)/2 \geq l/4 - 1$ .

Now, for every  $i = 1, \dots, q$ , we will find a claw containing  $d'_i$  as leaf. Assume that such a claw does not exist, we will derive a contradiction. Without loss of generality, we can assume that we have  $d'_i c'_j \notin E(G')$  and  $d'_i c'_{j+1} \in E(G')$ , for some  $j$  with  $1 \leq j \leq p-1$ . By hypothesis,  $c'_{j+1}$  is the center of a claw in  $G'$ . We denote by  $x, y$  and  $z$  the leaves of this claw. As  $d'_i$  is not the leaf of a claw, it is disjoint from  $\{x, y, z\}$ , and by the choice of  $d'_i$ , no one of these vertices distinguishes  $c'_j$  from  $c'_{j+1}$ . It means that  $c'_j$  is linked to all vertices of  $\{x, y, z\}$ . If two elements of this set, say  $x$  and  $y$ , are adjacent to  $d'_i$ , then  $\{x, d'_i, y, c'_j\}$  forms a 4-cycle that contains  $c'_j$ , which is not possible. So, at least two elements among  $\{x, y, z\}$ , say  $x$  and  $y$ , are not adjacent to  $d'_i$  and then, we find the claw  $\{c'_{j+1}, x, y, d'_i\}$  of center  $c'_{j+1}$  that contains  $d'_i$ , which is also not possible, by assumption.

Finally, for  $1 \leq i \leq q$  every  $d'_i$  is the leaf of a claw. So, by Lemma 2.2, we have  $q \leq k^2 + 2k$ . Then, we conclude that  $l \leq 4(k^2 + 2k + 1)$  and that  $G$  contains at most  $4(k^2 + 2k + 1)(k + 1)$  vertices which are center of a claw. Finally, using Lemma 2.2  $G$  contains at most  $4(k^2 + 2k + 1)(k + 1) + k^2 + 2k + 2k^2 + 2k$  vertices belonging to a claw or a 4-cycle.  $\square$

Remark that, using Lemma 2.1, it is possible to detect all the vertices of  $G$  which belongs to a claws or a 4-cycle in time  $O(n^2m)$ .

### 2.1.3 Bounding the size of the clean $K$ -joins

Now, we set a rule that will bound the number of vertices in a clean  $K$ -join, once applied. Although quite technical to prove, this rule is the core tool of our process of kernelization. Remark that, if we remove the vertices contained in a claw or a 4-cycle from a (general)  $K$ -join, we obtain a clean  $K$ -join. So, by the result of the previous subsection, providing a bound on the size of the clean  $K$ -joins will give a bound on the size of  $K$ -joins.

**Rule 2.4** ( $K$ -join). *Let  $B$  be a clean  $K$ -join of size at least  $2k + 2$ , provided with an umbrella ordering  $\sigma_B$ . Let  $B_L$  be the  $k + 1$  first vertices of  $B$  (according to  $\sigma_B$ ),  $B_R$  be its  $k + 1$  last vertices*

(according to  $\sigma_B$ ) and  $M = B \setminus (B_R \cup B_L)$ . Remove the set of vertices  $M$  from  $G$ .

**Lemma 2.4.** *Rule 2.4 is safe.*

*Proof.* Let  $G' = G \setminus M$ . Observe that the restriction to  $G'$  of any  $k$ -completion of  $G$  is a  $k$ -completion of  $G'$ , since proper interval graphs are closed under induced subgraphs. So, let  $F$  be a  $k$ -completion for  $G'$ . We denote by  $H$  the resulting proper interval graph  $G' + F$  and by  $\sigma_H = h_1, \dots, h_{|H|}$  an umbrella ordering of  $H$ . We prove that we can insert the vertices of  $M$  into  $\sigma_H$  and modify it if necessary, to obtain an umbrella ordering for  $G$  without adding any edge (in fact, some edges of  $F$  might even be deleted during the process). This will imply that  $G$  admits a  $k$ -completion as well. To see this, we need the following structural description of  $G$ . As explained before, we denote by  $N$  the set  $\cap_{b \in B} N_G(b) \setminus B$ , and abusively still denote by  $L$  (resp.  $R$ ) the set  $L \setminus N$  (resp.  $R \setminus N$ ) (see Figure 5). We also denote by  $b_1, \dots, b_{|B|}$  the umbrella ordering  $\sigma_B$  of  $B$ .

**Claim 2.5.** *The sets  $L$  and  $R$  are cliques of  $G$ .*

*Proof.* We prove that  $R$  is a clique in  $G$ . The proof for  $L$  uses similar arguments. No vertex of  $R$  is a neighbor of  $b_1$ , otherwise such a vertex must be adjacent to every vertex of  $B$  and then stands in  $N$ . So, if  $R$  contains two vertices  $u, v$  such that  $uv \notin E$ , we form the claw  $\{b_{|B|}, b_1, u, v\}$  with center  $b_{|B|}$ , contradicting the fact that  $B$  is clean.  $\diamond$

The following observation comes from the definition of a  $K$ -join.

**Observation 2.6.** *Given any vertex  $r \in R$ , if  $N_B(r) \cap B_L \neq \emptyset$  holds then  $M \subseteq N_B(r)$ . Similarly, given any vertex  $l \in L$ , if  $N_B(l) \cap B_R \neq \emptyset$  holds then  $M \subseteq N_B(l)$ .*

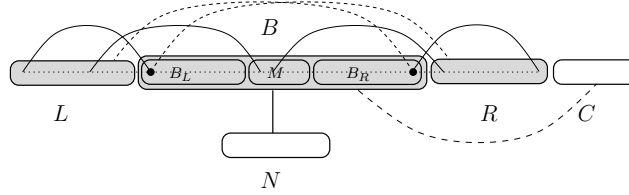


Figure 5: The structure of the  $K$ -join  $B$ .

We use these facts to prove that an umbrella ordering can be obtained for  $G$  by inserting the vertices of  $M$  into  $\sigma_H$ . Let  $h_f$  and  $h_l$  be respectively the first and last vertex of  $B \setminus M$  appearing in  $\sigma_H$ . We let  $B_H$  denote the set  $\{u \in V(H) : h_f \leq_{\sigma_H} u \leq_{\sigma_H} h_l\}$ . Observe that  $B_H$  is a clique in  $H$  since  $h_f h_l \in E(G)$  and that  $B \setminus M \subseteq B_H$ . Now, we modify  $\sigma_H$  by ordering the true twins in  $H$  according to their neighborhood in  $M$ : if  $x$  and  $y$  are true twins in  $H$ , are consecutive in  $\sigma_H$ , verify  $x <_{\sigma_H} y <_{\sigma_H} h_f$  and  $N_M(y) \subset N_M(x)$ , then we exchange  $x$  and  $y$  in  $\sigma_H$ . This process stops when the considered true twins are ordered following the join between  $\{u \in V(H) : u <_{\sigma_H} h_f\}$  and  $M$ . We proceed similarly on the right of  $B_H$ , i.e. for  $x$  and  $y$  consecutive twins with  $h_l <_{\sigma_H} x <_{\sigma_H} y$  and  $N_M(x) \subset N_M(y)$ . The obtained order is clearly an umbrella ordering too (in fact, we just re-labeled some vertices in  $\sigma_H$ ), and we abusively still denote it by  $\sigma_H$ .

**Claim 2.7.** *The set  $B_H \cup \{m\}$  is a clique of  $G$  for any  $m \in M$ , and consequently  $B_H \cup M$  is a clique of  $G$ .*

*Proof.* Let  $u$  be any vertex of  $B_H$ . We claim that  $um \in E(G)$ . Observe that if  $u \in B$  then the claim trivially holds. So assume  $u \notin B$ . Recall that  $B_H$  is a clique in  $H$ . It follows that  $u$  is adjacent to every vertex of  $B \setminus M$  in  $H$ . Since  $B_L$  and  $B_R$  both contain  $k+1$  vertices, we have  $N_G(u) \cap B_L \neq \emptyset$  and  $N_G(u) \cap B_R \neq \emptyset$ . Hence,  $u$  belongs to  $L \cup N \cup R$  and  $um \in E(G)$  by Observation 2.6.  $\diamond$

**Claim 2.8.** *Let  $m$  be any vertex of  $M$  and  $\sigma'_H$  be the ordering obtained from  $\sigma_H$  by removing  $B_H$  and inserting  $m$  to the position of  $B_H$ . The ordering  $\sigma'_H$  respects the umbrella property.*

*Proof.* Assume that  $\sigma'_H$  does not respect the umbrella property, i.e. that there exist (w.l.o.g.) two vertices  $u$  and  $v$  of  $H \setminus B_H$  such that either (1)  $u <_{\sigma'_H} v <_{\sigma'_H} m$ ,  $um \in E(H)$  and  $uv \notin E(H)$  or (2)  $u <_{\sigma'_H} m <_{\sigma'_H} v$ ,  $um \notin E(H)$  and  $uv \in E(H)$  or (3)  $u <_{\sigma'_H} v <_{\sigma'_H} m$ ,  $um \in E(H)$  and  $vm \notin E(H)$ . First, assume that (1) holds. Since  $uv \notin E(H)$  and  $\sigma_H$  is an umbrella ordering,  $uw \notin E(H)$  for any  $w \in B_H$ , and hence  $uw \notin E(G)$ . This means that  $B_L \cap N_G(u) = \emptyset$  and  $B_R \cap N_G(u) = \emptyset$ , which is impossible since  $um \in E(G)$ . Then, assume that (2) holds. Since  $uv \in E(H)$  and  $\sigma_H$  is an umbrella ordering,  $B_H \subseteq N_H(u)$ , and in particular  $B_L$  and  $B_R$  are included in  $N_H(u)$ . As  $|B_L| = |B_R| = k+1$ , we know that  $N_G(u) \cap B_L \neq \emptyset$  and  $N_G(u) \cap B_R \neq \emptyset$ , but then, Observation 2.6 implies that  $um \in E(G)$ . So, (3) holds, and we choose the first  $u$  satisfying this property according to the order given by  $\sigma'_H$ . So we have  $vm \notin E(G)$  for any  $v <_{\sigma'_H} u$ . Similarly, we choose  $v$  to be the first vertex after  $u$  satisfying  $vm \notin E(G)$ . Since  $um \in E(G)$ , we know that  $u$  belongs to  $L \cup N \cup R$ . Moreover, since  $vm \notin E(G)$ ,  $v \in C \cup L \cup R$ . There are several cases to consider:

- (i)  $u \in N$ : in this case we know that  $B \subseteq N_G(u)$ , and in particular that  $uh_l \in E(G)$ . Since  $\sigma_H$  is an umbrella ordering for  $H$ , it follows that  $vh_l \in E(H)$  and  $B_H \subseteq N_H(v)$ . Since  $|B_L| = |B_R| = k+1$ , we know that  $N_G(v) \cap B_L \neq \emptyset$  and  $N_G(v) \cap B_R \neq \emptyset$ . But, then Observation 2.6 implies that  $vm \in E(G)$ .
- (ii)  $u \in R$ ,  $v \notin R$ : since  $um \in E(G)$ ,  $B_R \subseteq N_G(u)$ . Let  $b \in B_R$  be the vertex such that  $B_R \subseteq \{w \in V : u <_{\sigma_H} w \leq_{\sigma_H} b\}$ . Since  $ub \in E(G)$ , this means that  $B_R \subseteq N_H(v)$ . Now, since  $|B_R| = k+1$ , it follows that  $N_G(v) \cap B_R \neq \emptyset$ . Observation 2.6 allows us to conclude that  $vm \in E(G)$ .
- (iii)  $u, v \in R$ : in this case,  $uv \in E(G)$  by Claim 2.7 but  $u$  and  $v$  are not true twins in  $H$  (otherwise  $v$  would be placed before  $u$  in  $\sigma_H$  due to the modification we have applied to  $\sigma_H$ ). This means that there exists a vertex  $w \in V(H)$  that distinguishes  $u$  from  $v$  in  $H$ . Assume first that  $w <_{\sigma_H} u$  and  $uw \in E(H)$ ,  $vw \notin E(H)$ . We choose the first  $w$  satisfying this according to the order given by  $\sigma_H$ . There are two cases to consider. First, if  $uw \in E(G)$ , then since  $vm \notin E(G)$  for any  $w <_{\sigma_H} u$  by the choice of  $u$ ,  $\{u, v, w, m\}$  is a claw in  $G$  containing a vertex of  $B$  (see Figure 6 (a) ignoring the vertex  $u'$ ), which cannot be. So assume  $uw \in F$ . By Observation 1.3,  $uw$  is not an extremal edge of  $\sigma_H$ . By the choice of  $w$  and since  $vw \notin E(H)$ , there exists  $u'$  with  $u <_{\sigma_H} u' <_{\sigma_H} v$  such that  $wu'$  is an extremal edge of  $\sigma_H$  (and hence belongs to  $E(G)$ , see Figure 6 (a)). Now, by the choice of  $v$  we have  $u'm \in E(G)$  and hence  $u' \in N \cup R \cup L$ . Observe that  $u'v \notin E(G)$ : otherwise  $\{u', v, w, m\}$  would form a claw in  $G$ . Since  $R$  is a clique of  $G$ , it follows that  $u' \in L \cup N$ . Moreover, since  $u'm \in E(G)$ ,  $B_L \subseteq N_G(u')$ . We conclude like in configuration (ii) that  $v$  should be adjacent to a vertex of  $B_L$  and hence to  $m$ .

Hence we can assume that all the vertices that distinguish  $u$  and  $v$  are after  $u$  in  $\sigma_H$  and that  $uw'' \in E(H)$  implies  $vw'' \in E(H)$  for any  $w'' <_{\sigma_H} u$ . Now, suppose that there exists  $w \in H$



such that  $h_l <_{\sigma_H} w$  and  $uw \notin E(H)$ ,  $vw \in E(H)$ . In particular, this means that  $B_L \subseteq N_H(v)$ . Since  $|B_L| = k+1$  we have  $N_G(v) \cap B_L \neq \emptyset$ , implying  $vm \in E(G)$  by Observation 2.6. Assume now that there exists a vertex  $w$  which distinguishes  $u$  and  $v$  with  $v <_{\sigma_H} w <_{\sigma_H} h_f$ . In this case, since  $uw \notin E(H)$ ,  $B \cap N_H(u) = \emptyset$  holds and hence  $B \cap N_G(u) = \emptyset$ , which cannot be since  $u \in R$ . Finally, assume that there is  $w \in B_H$  with  $wu \notin E(H)$  and  $wv \in E(H)$ . Recall that  $wm \in E(G)$  as  $B_H \cup \{m\}$  is a clique by Claim 2.7. We choose  $w$  in  $B_H$  distinguishing  $u$  and  $v$  to be the last according to the order given by  $\sigma_H$  (i.e.  $vw' \notin E(H)$  for any  $w <_{\sigma_H} w'$ , see Figure 6 (b), ignoring the vertex  $u'$ ).

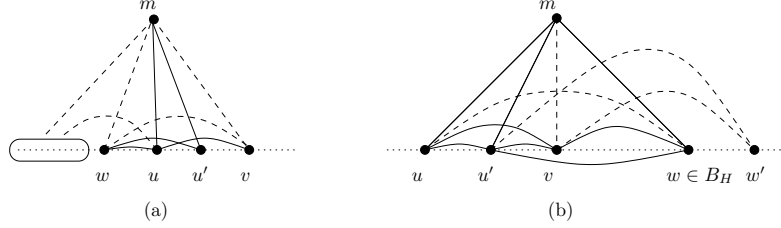


Figure 6: (a)  $u$  and  $v$  are distinguished by some vertex  $w <_{\sigma_H} u$ ; (b)  $u$  and  $v$  are distinguished by a vertex  $w \in B_H$ .

If  $vw \in E(G)$  then  $\{u, m, w, v\}$  is a 4-cycle in  $G$  containing a vertex of  $B$ , which cannot be. Hence  $vw \in F$  and by the choice of  $w$ , there exists  $u' \in V(H)$  such that  $u <_{\sigma_H} u' <_{\sigma_H} v$  and  $u'w$  is an extremal edge of  $\sigma_H$  (and then belongs to  $E(G)$ ). By the choice of  $v$  we know that  $u'm \in E(G)$ . Moreover, by the choice of  $w$ , observe that  $u'$  and  $v$  are true twins in  $H$  (if a vertex  $s$  distinguishes  $u'$  and  $v$  in  $H$ ,  $s$  cannot be before  $u$ , since otherwise  $s$  would distinguish  $u$  and  $v$ , not between  $u$  and  $w$  because it would be adjacent to  $u'$  and  $v$ , and not after  $w$ , by choice of  $w$ ). This leads to a contradiction since we assumed that  $N_M(x) \subseteq N_M(y)$  for any true twins  $x$  and  $y$  with  $x <_{\sigma_H} y <_{\sigma_H} h_f$ .

The cases where  $u \in L$  are similar, what concludes the proof of Claim 2.8  $\diamond$

Now, we will insert vertices of  $M$  into the graph  $H$  while preserving an umbrella ordering. For simplicity, once one vertex of  $M$  is inserted into  $H$ , we still denote the obtained graph by  $H$  and consider the new vertex as a vertex of  $H$ , for the next add. We then prove the following.

**Claim 2.9.** *Let  $m$  be a vertex of  $M$ . Then  $m$  can be added to the graph  $H$  while preserving an umbrella ordering.*

*Proof.* Let  $m$  be a vertex of  $M$  and  $h_i$  (resp.  $h_j$ ) be the vertex with minimal (resp. maximal) index in  $\sigma_H$  such that  $h_i m \in E(G)$  (resp.  $h_j m \in E(G)$ ). By definition, we have  $h_{i-1}m \notin E(G)$ ,  $h_{j+1}m \notin E(G)$  and through Claim 2.8, we know that  $N_H(m) = \{w \in V(H) : h_i \leq_{\sigma_H} w \leq_{\sigma_H} h_j\}$ . Moreover, since  $B_H \cup M$  is a clique by Claim 2.7, it follows that  $h_{i-1} <_{\sigma_H} h_f$  and  $h_l <_{\sigma_H} h_{j+1}$ . Hence, by Claim 2.8, we know that  $h_{i-1}h_{j+1} \notin E(G)$ , otherwise the ordering  $\sigma'_H$  defined in Claim 2.8 would not be an umbrella ordering. The situation is depicted in Figure 7 (a). For any vertex  $v \in N_H(m)$ , let  $N^-(v)$  (resp.  $N^+(v)$ ) denote the set of vertices  $\{w \in V(H) : w \leq_{\sigma_H} h_{i-1} \text{ and } wv \in E(H)\}$  (resp.  $\{w \in V(H) : w \geq_{\sigma_H} h_{j+1} \text{ and } wv \in E(H)\}$ ). Observe that for any vertex

$v \in N_H(m)$ , if there exist two vertices  $x \in N^-(v)$  and  $y \in N^+(v)$  such that  $xv \in E(G)$  and  $yv \in E(G)$ , then the set  $\{v, x, y, m\}$  defines a claw containing  $m$  in  $G$ , which cannot be. We now consider  $c_{h_{i-1}}$  the neighbor of  $h_{i-1}$  with maximal index in  $\sigma_H$ . Similarly we let  $c_{h_{j+1}}$  be the neighbor of  $h_{j+1}$  with minimal index in  $\sigma_H$ . Since  $h_{i-1}h_{j+1} \notin E(G)$ , we have  $c_{h_{i-1}}, c_{h_{j+1}} \in N_H(m)$ . We study the behavior of  $c_{h_{i-1}}$  and  $c_{h_{j+1}}$  in order to conclude.

Assume first that  $c_{h_{j+1}} \leq_{\sigma_H} c_{h_{i-1}}$ . Let  $X$  be the set of vertices  $\{w \in V(H) : c_{h_{j+1}} \leq_{\sigma_H} w \leq_{\sigma_H} c_{h_{i-1}}\}$ . Remark that we have  $c_{h_{i-1}} \leq_{\sigma_H} h_i$  and  $h_j \leq_{\sigma_H} c_{h_{j+1}}$ , otherwise for instance, if we have  $c_{h_{i-1}} >_{\sigma_H} h_i$ , then  $B_H \subseteq N_H(h_{i-1})$  implying, as usual, that  $h_{i-1}m \in E(G)$  which is not. So, we know that  $X \subseteq B_H$ . Then, let  $X_1 \subseteq X$  be the set of vertices  $x \in X$  such that there exists  $w \in N^+(x)$  with  $xw \in E(G)$  and  $X_2 = X \setminus X_1$ . Let  $x \in X_1$ : observe that by construction  $xw' \in F$  for any  $w' \in N^-(x)$ . Similarly, given  $x \in X_2$ ,  $xw'' \in F$  for any  $w'' \in N^+(x)$ . Now, we reorder the vertices of  $X$  as follows: we first put the vertices from  $X_2$  and then the vertices from  $X_1$ , preserving the order induced by  $\sigma_H$  for both sets. Moreover, we remove from  $E(H)$  all edges between  $X_1$  and  $N^-(X_1)$  and between  $X_2$  and  $N^+(X_2)$ . Recall that such edges have to belong to  $F$ . We claim that inserting  $m$  between  $X_2$  and  $X_1$  yields an umbrella ordering (see Figure 7 b). Indeed, by Claim 2.8, we know that the umbrella ordering is preserved between  $m$  and the vertices of  $H \setminus B_H$ .

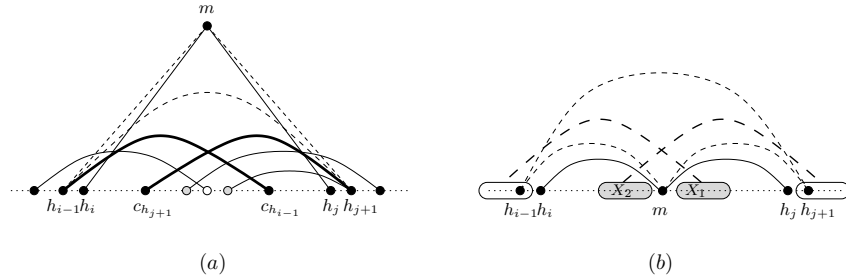


Figure 7: Illustration of the reordering applied to  $\sigma_H$ . The thin edges stand for edges of  $G$ . On the left, the gray vertices represent vertices of  $X_1$  while the white vertex is a vertex of  $X_2$ .

Now, remark that there is no edge between  $X_1$  and  $\{w \in V(H) : w \leq_{\sigma_H} h_{i-1}\}$ , that there is no edge between  $X_2$  and  $\{w \in V(H) : w \geq_{\sigma_H} h_{j+1}\}$ , that there are still all the edges between  $N_H(m)$  and  $X_1 \cup X_2$  and that the edges between  $X_1$  and  $\{w \in V(H) : w \geq_{\sigma_H} h_{j+1}\}$  and the edges between  $X_2$  and  $\{w \in V(H) : w \leq_{\sigma_H} h_{i-1}\}$  are unchanged. So, it follows that the new ordering respects the umbrella property, and we are done.

Next, assume that  $c_{h_{i-1}} <_{\sigma_H} c_{h_{j+1}}$ . We let  $c_{h_i}$  (resp.  $c_{h_j}$ ) be the neighbor of  $h_i$  (resp.  $h_j$ ) with maximal (resp. minimal) index in  $N_H(m)$ . Notice that  $c_{h_{i-1}} \leq_{\sigma_H} c_{h_i}$  and  $c_{h_j} \leq_{\sigma_H} c_{h_{j+1}}$  (see Figure 8). Two cases may occur:

- (i) First, assume that  $c_{h_i} <_{\sigma_H} c_{h_j}$ , case depicted in Figure 8 (a). In particular, this means that  $h_i h_j \notin E(G)$ . If  $c_{h_i}$  and  $c_{h_j}$  are consecutive in  $\sigma_H$ , then inserting  $m$  between  $c_{h_i}$  and  $c_{h_j}$  yields an umbrella ordering (since  $c_{h_j}$  (resp.  $c_{h_i}$ ) does not have any neighbor before (resp. after)  $h_i$  (resp.  $h_j$ ) in  $\sigma_H$ ). Now, if there exists  $w \in V(H)$  such that  $c_{h_i} <_{\sigma_H} w <_{\sigma_H} c_{h_j}$ , then one can see that the set  $\{m, h_i, w, h_j\}$  forms a claw containing  $m$  in  $G$ , which is impossible.
- (ii) The second case to consider is when  $c_{h_j} \leq_{\sigma_H} c_{h_i}$ . In such a case, one can see that  $m$  and

the vertices of  $\{w \in V(H) : c_{h_j} \leq_{\sigma_H} w \leq_{\sigma_H} c_{h_i}\}$  are true twins in  $H \cup \{m\}$ , because their common neighborhood is exactly  $\{w \in V(H) : h_i \leq_{\sigma_H} w \leq_{\sigma_H} h_j\}$ . Hence, inserting  $m$  just before  $c_{h_i}$  (or anywhere between  $c_{h_i}$  and  $c_{h_j}$  or just after  $c_{h_j}$ ) yields an umbrella ordering.

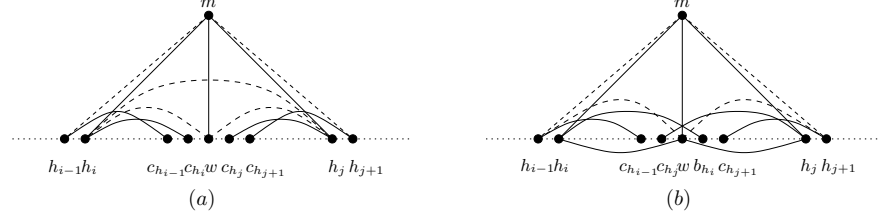


Figure 8: The possible cases for  $c_{h_{i-1}} <_{\sigma_H} c_{h_{j+1}}$ .

◇

As explained before, since the proof of Claim 2.9 does not use the fact that the vertices of  $H$  do not belong to  $M$ , it follows that we can iteratively insert the vertices of  $M$  into  $\sigma_H$ , preserving an umbrella ordering at each step. This concludes the proof of Lemma 2.4.  $\square$

The complexity needed to compute Rule 2.4 will be discussed in the next section. The following observation results from the application of Rule 2.4 and from Section 2.1.2.

**Observation 2.10.** *Let  $G = (V, E)$  be a positive instance of PROPER INTERVAL COMPLETION reduced under Rules 2.2 to 2.4. Any  $K$ -join of  $G$  contains at most  $2k + 2$  vertices which are not contained in any 4-cycle or claw of  $G$ .*

*Proof.* Let  $B$  be any  $K$ -join of  $G$ , and  $X$  be the set of vertices of  $B$  which are contained in a 4-cycle or a claw of  $G$ . As any subgraph of a  $K$ -join is a  $K$ -join,  $B \setminus X$  is a clean  $K$ -join of  $G$ . Then, after having applied Rule 2.4, we have  $|B \setminus X| \leq 2k + 2$ .  $\square$

#### 2.1.4 Cutting the 1-branches

We now turn our attention to branches of a graph  $G = (V, E)$ , proving how they can be reduced.

**Lemma 2.11.** *Let  $G = (V, E)$  be a connected graph which is a positive instance of PROPER INTERVAL COMPLETION, and let  $B$  be a 1-branch of  $G$  associated with the umbrella ordering  $\sigma_B$ . Assume that  $|B^R| \geq 2k + 1$  and let  $B_L$  be the  $2k + 1$  last vertices of  $B^R$  according to  $\sigma_B$ . Then, there exists a  $k$ -completion  $F$  of  $G$  into a proper interval graph and a vertex  $b \in B_L$  such that the umbrella ordering of  $G + F$  preserves the order induced by  $\sigma_B$  on the set  $B_b = \{w \in V(B) : b_1 \leq_{\sigma_B} w \leq_{\sigma_B} b_f\}$ , where  $f$  is the maximal index in  $\sigma_B$  such that  $bb_f \in E(G)$ . Moreover, the vertices of  $B_b$  are the first in an umbrella ordering of  $G + F$ .*

*Proof.* Let  $F$  be any  $k$ -completion of  $G$ ,  $H = G + F$  and  $\sigma_H$  be the umbrella ordering of  $H$ . Since  $|B_L| = 2k + 1$  and  $|F| \leq k$ , there exists a vertex  $b \in B_L$  not incident to any added edge of  $F$ . We let  $N_b$  be the set of neighbors of  $b$  that are after  $b$  in  $\sigma_B$ ,  $B_b = \{w \in V(B) : b_1 \leq_{\sigma_B} w \leq_{\sigma_B} b_f\}$ , where  $f$  is the maximal index in  $\sigma_B$  such that  $bb_f \in E(G)$  (i.e.  $b_f$  is the last vertex of  $N_b$ ), and  $C = V \setminus B_b$  (see Figure 9, which depicts the case where  $b_f \in B_1$ , but  $b_f \in B_L$  is possible too).

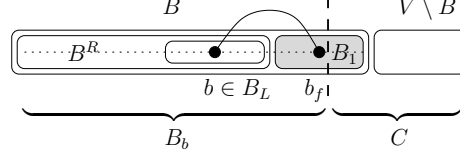


Figure 9: The different sets to cut a 1-branch (the set  $N_b$ , not shown in figure, is made by vertices lying under the edge  $bb_f$ , including  $b_f$ , not  $b$ ).

Remark that, by the definition of the attachment clique  $B_1$  (which is  $B_1 = N_B(b_{|B|})$ ), we have  $B_1 \not\subseteq B_b$  (because  $b$  is not a neighbor of  $b_{|B|}$ ) and then  $B \cap C \neq \emptyset$ .

**Claim 2.12.** *We have:*

- (i)  $G[C]$  is a connected graph and
- (ii) Either for every vertex  $u$  of  $C$  we have  $b <_{\sigma_H} u$  or for every vertex  $u$  of  $C$  we have  $u <_{\sigma_H} b$ .

*Proof.* The first point follows from the fact that by definition of a 1-branch, every vertex of  $V \setminus B$  which has a neighbor in  $B$  is a neighbor of  $b_{|B|}$  (which belongs to  $C$ ). So, as  $G$  is connected, every connected component of  $G[V \setminus B]$  contains a neighbor of  $b_{|B|}$ . As,  $C \cap B$  is a subset of the attachment clique  $B_1$  and then linked to  $b_B$ , we conclude that  $G[C]$  is a connected graph.

To see the second point, assume that there exist  $u, v \in C$  such that w.l.o.g.  $u <_{\sigma_H} b <_{\sigma_H} v$ . Since  $G[C]$  is a connected graph, there exists a path between  $u$  and  $v$  in  $G$  that avoids  $N_G[b]$ , which is equal to  $N_H[b]$  since  $b$  is not incident to any edge of  $F$ . Hence there exist  $u', v' \in C$ , consecutive along this path, such that  $u' <_{\sigma_H} b <_{\sigma_H} v'$  and  $u'v' \in E(G)$ . Then, as the neighborhood of  $b$  is the same in  $G$  than in  $H$ , we have  $u'b, v'b \notin E(H)$ , contradicting the fact that  $\sigma_H$  is an umbrella ordering for  $H$ .  $\diamond$

In the following, up to reversing the order  $\sigma_H$ , we assume that  $b <_{\sigma_H} u$  holds for any  $u \in C$ . We will then find  $B_b$  at the beginning of  $\sigma_H$ . We now consider the following ordering  $\sigma$  of  $H$ : we first put the set  $B_b$  according to the order of  $B$  and then put the remaining vertices  $C$  according to  $\sigma_H$  (see Figure 10). We construct a corresponding completion  $F'$  of  $G$  from  $F$  as follows: we remove from  $F$  the edges with both extremities in  $B_b$ , and remove all edges between  $B_b \setminus N_b$  and  $C$ . In other words, we set:

$$F' = F \setminus (F[B_b \times B_b] \cup F[(B_b \setminus N_b) \times C])$$

Finally, we inductively remove from  $F'$  any extremal edge of  $\sigma$  that belongs to  $F'$ , and abusively still call  $F'$  the obtained edge set.

**Claim 2.13.** *The set  $F'$  is a  $k$ -completion of  $G$ .*

*Proof.* We prove that  $\sigma$  is an umbrella ordering of  $H' = G + F'$ . Since  $|F'| \leq |F|$  by construction, the result will follow. Assume this is not the case. By definition of  $F'$ ,  $H'[B_b]$  and  $H'[C]$  induce proper interval graphs. This means that there exists a set of vertices  $S = \{u, v, w\}$ ,  $u <_{\sigma} v <_{\sigma} w$ , intersecting both  $B_b$  and  $C$  and violating the umbrella property. We either have (1)  $uw \in E, uv \notin E$  or (2)  $uw \in E, vw \notin E$ . Since neither  $F'$  nor  $G$  contain an edge between  $B_b \setminus N_b$  and  $C$ , it follows that  $S$  intersects  $N_b$  and  $C$ . We study the different cases:

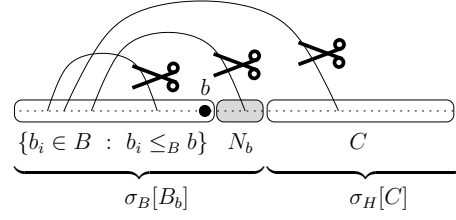


Figure 10: The construction of the ordering  $\sigma$  and the set  $F'$  (possible cut edges are from  $F$ ).

- (i) (1) holds and  $u \in N_b$ ,  $v, w \in C$ : since the edge set between  $N_b$  and  $C$  is the same in  $H$  and  $H'$ , it follows that  $uv \notin E(H)$ . Since  $\sigma_H$  is an umbrella ordering of  $H$ , we either have  $v <_{\sigma_H} u <_{\sigma_H} w$  or  $v <_{\sigma_H} w <_{\sigma_H} u$  (recall that  $C$  is in the same order in both  $\sigma$  and  $\sigma_H$ ). Now, recall that  $b <_{\sigma_H} \{v, w\}$  by assumption. In particular, since  $bu \in E(G)$ , this implies in both cases that  $\sigma_H$  is not an umbrella ordering, what leads to a contradiction.
- (ii) (1) holds and  $u, v \in N_b$ ,  $w \in C$ : this case cannot happen since  $N_b$  is a clique of  $H'$ .
- (iii) (2) holds and  $u \in N_b$ ,  $v, w \in C$ : this case is similar to (i). Observe that we may assume  $uw \in E(H)$  (otherwise (i) holds). By construction of  $F'$ , we have  $vw \notin E(H)$  and hence  $v <_{\sigma_H} w <_{\sigma_H} u$  or  $v <_{\sigma_H} u <_{\sigma_H} w$ . The former case contradicts the fact that  $\sigma_H$  is an umbrella ordering since  $wu \in E(H)$ . In the latter case, since  $\sigma_H$  is an umbrella ordering this means that  $bv \in E(H)$  (as  $bu \in E(H)$  and  $v <_{\sigma_H} u <_{\sigma_H} w$ ). Since  $b$  is non affected vertex and  $v \in C$ , we have  $bv \notin E(G)$ , which leads to a contradiction.
- (iv) (2) holds and  $u, v \in N_b$ ,  $w \in C$ : first, if  $uw \in E(G)$ , then we have a contradiction since  $N_C(u) \subseteq N_C(v)$ . So, we have  $uw \in F'$ . By construction of  $F'$ , we know that  $uw$  is not an extremal edge. Hence there exists an extremal edge (of  $G$ ) above  $uw$ , which is either  $uw'$  with  $w <_{\sigma} w'$ ,  $u'w$  with  $u' <_{\sigma} u$  or  $u'w'$  with  $u' <_{\sigma} u <_{\sigma} w <_{\sigma} w'$ . The three situation are depicted in Figure 11. In the first case,  $vw' \in E(G)$  (since  $N_C(u) \subseteq N_C(v)$  in  $G$ ) and hence we are in configuration (i) with vertex set  $\{v, w, w'\}$ . In the second case,  $u'w \in E(G)$  and  $vw \notin E(G)$  are in contradiction with  $N_C(u') \subseteq N_C(v)$  in  $G$  (since  $u' \in B_b$ ). Finally, in the third case,  $vw' \in E(G)$  (since  $N_C(u') \subseteq N_C(v)$  in  $G$ ), and we are in configuration (i) with vertex set  $\{v, w, w'\}$ .

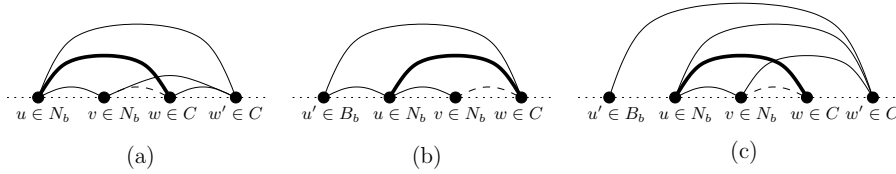


Figure 11: Illustration of the different cases of configuration (iv) (the bold edges belong to  $F'$ ).

◇

Altogether, we proved that there exists a  $k$ -completion of  $G$  associated with an umbrella ordering where the vertices of  $B_b$  are ordered in the same way than in the ordering of  $B$  and stand at the beginning of this ordering, what concludes the proof.  $\square$

**Rule 2.5** (1-branches). *Let  $B$  be a 1-branch such that  $|B^R| > 2k + 1$ . Remove  $B^R \setminus B_L$  from  $G$ , where  $B_L$  denotes the  $2k + 1$  last vertices of  $B^R$ .*

**Lemma 2.14.** *Rule 2.5 is safe.*

*Proof.* Let  $G' = G \setminus (B^R \setminus B_L)$  denote the reduced graph. Observe that any  $k$ -completion of  $G$  is a  $k$ -completion of  $G'$  since proper interval graphs are closed under induced subgraphs. So let  $F$  be a  $k$ -completion of  $G'$ . We denote by  $H = G' + F$  the resulting proper interval graph and let  $\sigma_H$  be the corresponding umbrella ordering. Without loss of generality, we assume that the connected component of  $H$  containing  $B_L$  is the first according to  $\sigma_H$ . Remark that  $B_1 \cup B_L$  forms a 1-branch of  $G'$ , which we denote by  $B'$ . The umbrella ordering associated with  $B'$  is induced by  $\sigma_B$ . So, as previously, for a vertex  $b$  of  $B_L$ , we denote  $\{wV(B') : b_1 \leq_{\sigma_{B'}} w \leq_{\sigma_{B'}} b_f\}$  by  $B_b$ . By Lemma 2.11 we know that there exists a vertex  $b \in B_L$  such that the order of  $B_b$  in  $\sigma_H$  is the same than in  $\sigma_{B'}$  and the vertices of  $B_b$  are the first of  $\sigma_H$ . Since  $N_G(B^R \setminus B_L) \subseteq B_b$ , it follows that the vertices of  $B^R \setminus B_L$  can be inserted into  $\sigma_H$  while respecting the umbrella property. Hence,  $F$  is a  $k$ -completion for  $G$ , implying the result.  $\square$

Here again, the time complexity needed to compute Rule 2.5 will be discussed in the next section. The following property of a reduced graph will be used to bound the size of our kernel.

**Observation 2.15.** *Let  $G = (V, E)$  be a positive instance of PROPER INTERVAL COMPLETION reduced under Rules 2.2 to 2.5. Every 1-branch of  $G$  contains at most  $4k + 3$  vertices which are not contained in any 4-cycle or claw of  $G$ .*

*Proof.* Let  $B$  be a 1-branch of a graph  $G = (V, E)$  reduced under Rules 2.2 to 2.5. As  $B$  has been reduced under Rule 2.5, we know that  $B \setminus B_1$  contains at most  $2k + 1$  vertices. Furthermore  $B_1$  forms a  $K$ -join of  $G$ , and then, by Observation 2.10, contains at most  $2k + 2$  vertices which are not contained in any 4-cycle or claw of  $G$ .  $\square$

### 2.1.5 Cutting the 2-branches

We now focus on 2-branches of the graph and explain how to reduce them. Let  $(G, k)$  be an instance of PROPER INTERVAL COMPLETION and  $B = \{b_1, \dots, b_{|B|}\}$  be a 2-branch of  $G$  associated with the umbrella ordering  $\sigma_B$ . Recall that the attachment cliques of  $B$  are  $B_1 = \{b \in V(B) : b_1 \leq_{\sigma_B} b \leq_{\sigma_B} b_{l'}\}$ , where  $b_{l'}$  is the neighbor of  $b_1$  with maximal index in  $\sigma_B$ , and  $B_2 = \{b \in V(B) : b_l \leq_{\sigma_B} b \leq_{\sigma_B} b_{|B|}\}$ , where  $b_l$  is the neighbor of  $b_{|B|}$  with minimal index in  $\sigma_B$ . Now, we define the next cliques in the 2-branch  $B$  (see Figure 12), namely  $B'_1 = \{b \in V(B) : b_{l'+1} \leq_{\sigma_B} b \leq_{\sigma_B} b_{\bar{l}}\}$ , where  $b_{\bar{l}}$  is the neighbor of  $b_{l'+1}$  with maximal index in  $\sigma_B$ , and  $B'_2 = \{b \in V(B) : b_{\bar{l}} \leq_{\sigma_B} b \leq_{\sigma_B} b_{l-1}\}$ , where  $b_{\bar{l}}$  is the neighbor of  $b_{l-1}$  with minimal index in  $\sigma_B$ . Finally, we denote by  $B_M$  the set  $B \setminus (B_1 \cup B'_1 \cup B'_2 \cup B_2)$ . Remark that by definition, we have  $B^R = B'_1 \cup B_M \cup B'_2$ . Remark also that  $B_M$  could be empty if  $B$  is made with four  $K$ -join or less. However, we are interested in 2-branches  $B$  with  $B_M$  large enough, to reduce it.

**Rule 2.6** (2-branches). *Let  $G$  be a connected instance of PROPER INTERVAL COMPLETION and  $B$  be a 2-branch such that  $G[V \setminus B^R]$  is not connected. Assume that  $|B_M| \geq 4k + 2$  and let  $B_M^f$  be the set of the  $2k + 1$  vertices after  $B'_1$  according to  $\sigma_B$  and  $B_M^l$  be the set of the  $2k + 1$  vertices*



*Proof.* Let  $B$  be a 2-branch of  $G$ , reduced under Rules 2.2 to 2.6, and  $C$  be the connected component containing  $B$ . The sets  $B_1$ ,  $B'_1$ ,  $B'_2$  and  $B_2$  form four  $K$ -joins of  $G$ , and then by Observation 2.10, they contain in all at most  $4 \cdot (2k + 2) = 8k + 8$  vertices which are not contained in any 4-cycle or claw of  $G$ . Furthermore, if  $G[C \setminus B^R]$  is not connected, then, as  $G$  is reduced under Rule 2.6,  $B \setminus (B_1 \cup B'_1 \cup B'_2 \cup B_2)$  contains at most  $4k + 2$  vertices, what provides the announced bound.  $\square$

## 2.2 Detecting the branches

We now turn our attention to the complexity needed to compute reduction rules 2.4 to 2.6. Mainly, we indicate how to obtain the maximum branches in order to reduce them. The detection of a branch is straightforward except for the attachment cliques, where several choices are possible. So, first, we detect the maximum 1-branches of  $G$ . Remark that for every vertex  $x$  of  $G$ , the set  $\{x\}$  is a 1-branch of  $G$ . The next lemma indicates how to compute a maximum 1-branch that contains a fixed vertex  $x$  as first vertex.

**Lemma 2.18.** *Let  $G = (V, E)$  be a graph and  $x$  a vertex of  $G$ . In time  $O(nm)$ , it is possible to detect a maximum 1-branch of  $G$  containing  $x$  as first vertex.*

*Proof.* To detect such a 1-branch, we design an algorithm which has two parts. Roughly speaking, we first try to detect the set  $B^R$  of a 1-branch  $B$  containing  $x$ . We set  $B_0^R = \{x\}$  and  $\sigma_0 = x$ . Once  $B_{i-1}^R$  has been defined, we construct the set  $C_i$  of vertices of  $G \setminus (\cup_{l=1}^{i-1} B_l^R)$  that are adjacent to at least one vertex of  $B_{i-1}^R$ . Two cases can appear. First, assume that  $C_i$  is a clique and that it is possible to order the vertices of  $C_i$  such that for every  $1 \leq j < |C_i|$ , we have  $N_{B_{i-1}^R}(c_{j+1}) \subseteq N_{B_{i-1}^R}(c_j)$  and  $(N_G(c_j) \setminus B_{i-1}^R) \subseteq (N_G(c_{j+1}) \setminus B_{i-1}^R)$ . In this case, the vertices of  $C_i$  correspond to a new  $K$ -join of the searched 1-branch (remark that, along this inductive construction, there is no edge between  $C_i$  and  $\cup_{l=1}^{i-2} B_l^R$ ). So, we let  $B_i^R = C_i$  and  $\sigma_i$  be the concatenation of  $\sigma_{i-1}$  and the ordering defined on  $C_i$ . In the other case, such an ordering of  $C_i$  can not be found, meaning that while detecting a 1-branch  $B$ , we have already detected the vertices of  $B^R$  and at least one (possibly more) vertex of the attachment clique  $B_1$  with neighbors in  $B^R$ . Assume that the process stops at step  $p$  and let  $C$  be the set of vertices of  $G \setminus \cup_{l=1}^p B_l^R$  which have neighbors in  $\cup_{l=1}^p B_l^R$  and  $B'_1 \subseteq B_p^R$  be the set of vertices that are adjacent to all the vertices of  $C$ . Remark that  $B'_1 \neq \emptyset$ , as  $B'_1$  contains at least the last vertex of  $\sigma_p$ . We denote by  $B^R$  the set  $(\cup_{l=1}^p B_l^R) \setminus B'_1$  and we will construct the largest  $K$ -join containing  $B'_1$  in  $G \setminus B^R$  which is compatible with  $\sigma_p$ , in order to define the attachment clique  $B_1$  of the desired 1-branch. The vertices of  $C$  are the candidates to complete the attachment clique. On  $C$ , we define the following oriented graph: there is an arc from  $u$  to  $v$  if:  $uv$  is an edge of  $G$ ,  $N_{B^R}(v) \subseteq N_{B^R}(u)$  and  $N_{G \setminus B^R}[u] \subseteq N_{G \setminus B^R}[v]$ . This graph can be computed in time  $O(nm)$ . Now, it is easy to check that the obtained oriented graph is a transitive graph, in which the equivalent classes are made of true twins in  $G$ . A path in this oriented graph corresponds, by definition, to a  $K$ -join containing  $B'_1$  and compatible with  $\sigma_p$ . As it is possible to compute a longest path in linear time in this oriented graph, we obtain a maximum 1-branch of  $G$  that contains  $x$  as first vertex.  $\square$

So, we detect all the maximum 1-branches of  $G$  in time  $O(n^2m)$ .

Now, to detect the 2-branches, we first detect for all pairs of vertices a maximum  $K$ -join with these vertices as ends. More precisely, if  $\{x, y\}$  are two vertices of  $G$  linked by an edge, then  $\{x, y\}$  is a  $K$ -join of  $G$ , with  $N = N_G(x) \cap N_G(y)$ ,  $L = N_G(x) \setminus N_G[y]$  and  $R = N_G(y) \setminus N_G[x]$ . So, there exist  $K$ -joins with  $x$  and  $y$  as ends, and we will compute such a  $K$ -join with maximum cardinality.

**Lemma 2.19.** *Let  $G = (V, E)$  be a graph and  $x$  and  $y$  two adjacent vertices of  $G$ . It is possible to compute in  $O(nm)$  time a maximum (in cardinality)  $K$ -join that admits  $x$  and  $y$  as ends.*



*Proof.* We denote  $N_G[x] \cap N_G[y]$  by  $N$ ,  $N_G(x) \setminus N_G[y]$  by  $L$  and  $N_G(y) \setminus N_G[x]$  by  $R$ . Let us denote by  $N'$  the set of vertices of  $N$  that contains  $N$  in their closed neighborhood. The vertices of  $N'$  are the candidates to belong to the desired  $K$ -join, and we can identify them in time  $O(n^2)$ . Now, we construct on  $N'$  an oriented graph  $D$ , putting, for every vertices  $u$  and  $v$  of  $N'$ , an arc from  $u$  to  $v$  if:  $N_G(v) \cap L \subseteq N_G(u) \cap L$  and  $N_G(u) \cap R \subseteq N_G(v) \cap R$ . Basically, it could take a  $O(n)$  time to decide if there is an arc from  $u$  to  $v$  or not, and so the whole oriented graph could be computed in time  $O(n \cdot |N'|^2)$ . As  $N'$  is a clique of  $G$ , we have  $|N'|^2 = O(m)$ . Now, it is easy to check that the obtained oriented graph is a transitive graph in which the equivalent classes are made of true twins in  $G$ . In this oriented graph, it is possible to compute a longest path from  $x$  to  $y$  in linear time. Such a path corresponds to a maximal  $K$ -join that admits  $x$  and  $y$  as ends. It follows that the desired  $K$ -join can be identified in  $O(nm)$  time.  $\square$

Now, for every edge  $xy$  of  $G$ , we compute a maximum  $K$ -join that contains  $x$  and  $y$  as ends and a reference to all the vertices that this  $K$ -join contains. This computation takes a  $O(nm^2)$  time and gives, for every vertex, some maximum  $K$ -joins that contain this vertex. These  $K$ -joins will be useful to compute the 2-branches of  $G$ , in particular through the next lemma.

**Lemma 2.20.** *Let  $B$  be a 2-branch of  $G$  with  $B^R \neq \emptyset$ , and  $x$  a vertex of  $B^R$ . Then, for every maximal (by inclusion)  $K$ -join  $B'$  that contains  $x$  there exists an extremal edge  $uv$  of  $\sigma_B$  such that  $B' = \{w \in B : u \leq_{\sigma_B} w \leq_{\sigma_B} v\}$ .*

*Proof.* As usually, we denote by  $L$ ,  $R$  and  $C$  the partition of  $G \setminus B$  associated with  $B$  and by  $\sigma_B$  the umbrella ordering associated with  $B$ . Let  $B'$  be a maximal  $K$ -join that contains  $x$  and define by  $b_f$  (resp.  $b_l$ ) the first (resp. last) vertex of  $B'$  according to  $\sigma_B$ . As there is no edge between  $\{u \in B : u <_{\sigma_B} b_f\} \cup L \cup C$  and  $b_l$  and no edge between  $\{u \in B : b_l <_{\sigma_B} u\} \cup R \cup C$  and  $b_f$ , we have  $B' \subseteq \{u \in B : b_f \leq_{\sigma_B} u \leq_{\sigma_B} b_l\}$ . Furthermore, as  $\{u \in B : b_f \leq_{\sigma_B} u \leq_{\sigma_B} b_l\}$  is a  $K$ -join and  $B'$  is maximal, we have  $B' = \{u \in B : b_f \leq_{\sigma_B} u \leq_{\sigma_B} b_l\}$ . Now, if  $b_f b_l$  was not an extremal edge of  $\sigma_B$ , it would be possible to extend  $B'$ , contradicting the maximality of  $B'$ .  $\square$

Now, we can detect the 2-branches  $B$  with a set  $B^R$  non empty.

**Lemma 2.21.** *Let  $G = (V, E)$  be a graph,  $x$  a vertex of  $G$  and  $B'$  a given maximal  $K$ -join that contains  $x$ . There is a  $O(nm)$  time algorithm to decide if there exists a 2-branch  $B$  of  $G$  which contains  $x$  as a vertex of  $B^R$ , and if it exists, to find a maximum 2-branch with this property.*

*Proof.* By Lemma 2.20, if there exists a 2-branch  $B$  of  $G$  which contains  $x$  as a vertex of  $B^R$ , then  $B'$  corresponds to a set  $\{u \in B : b_f \leq_{\sigma_B} u \leq_{\sigma_B} b_l\}$  where  $b_f b_l$  is an extremal edge of  $B$ . We denote by  $L'$ ,  $R'$  and  $C'$  the usual partition of  $G \setminus B'$  associated with  $B'$ , and by  $\sigma_{B'}$  the umbrella ordering of  $B'$ . In  $G$ , we remove the set of vertices  $\{u \in B' : u <_{\sigma_{B'}} x\}$  and the edges between  $L'$  and  $\{u \in B' : x \leq_{\sigma_{B'}} u\}$  and denote by  $H_1$  the resulting graph. From the definition of the 2-branch  $B$ ,  $\{u \in B : x \leq_{\sigma_B} u\}$  is a 1-branch of  $H_1$  that contains  $x$  as first vertex. So, using Lemma 2.18, we find a maximal 1-branch  $B_1$  that contains  $x$  as first vertex. Remark that  $B_1$  has to contain  $\{u \in B : x \leq_{\sigma_B} u\} \cap B^R$  at its beginning. Similarly, we define  $H_2$  from  $G$  by removing the vertex set  $\{u \in B' : x <_{\sigma_{B'}} u\}$  and the edges between  $R'$  and  $\{u \in B' : u \leq_{\sigma_{B'}} x\}$ . We detect in  $H_2$  a maximum 1-branch  $B_2$  that contains  $x$  as last vertex, and as previously,  $B_2$  has to contain  $\{u \in B : u \leq_{\sigma_B} x\} \cap B^R$  at its end. So,  $B_1 \cup B_2$  forms a maximum 2-branch of  $G$  containing  $x$ .  $\square$

We would like to mention that it could be possible to improve the execution time of our detecting branches algorithm, using possibly more involved techniques (as for instance, inspired from [7]).

However, this is not our main objective here.

Anyway, using the  $O(n^2m)$  time algorithm explained in Lemma 2.1 to localize all the 4-cycles and the claws, we obtain the following result.

**Lemma 2.22.** *Given a graph  $G = (V, E)$ , the reduction rules 2.4 to 2.6 can be carried out in polynomial time, namely in time  $O(nm(n + m))$ .*

### 2.3 Kernelization algorithm

We are now ready to state the main result of this Section. The kernelization algorithm consists of an exhaustive application of Rules 2.1 to 2.6.

**Theorem 2.23.** *The PROPER INTERVAL COMPLETION problem admits a kernel with  $O(k^3)$  vertices, computable in time  $O(nm(n + m))$ .*

*Proof.* Let  $G = (V, E)$  be a positive instance of PROPER INTERVAL COMPLETION reduced under Rules 2.1 to 2.6. Let  $F$  be a  $k$ -completion of  $G$ ,  $H = G + F$  and  $\sigma_H$  be the umbrella ordering of  $H$ . Since  $|F| \leq k$ ,  $G$  contains at most  $2k$  affected vertices (i.e. incident to an added edge). Let  $A = \{a_1 <_{\sigma_H} \dots <_{\sigma_H} a_i <_{\sigma_H} \dots <_{\sigma_H} a_p\}$  be the set of such vertices, with  $p \leq 2k$ . The size of the kernel is due to the following observations, which we admit without proof (see Figure 13).

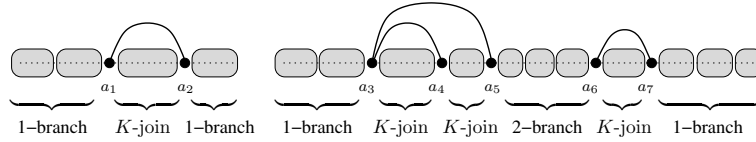


Figure 13: Illustration of the size of the kernel. The figure represents the graph  $H = G + F$ , the  $a_i$ 's are the affected vertices, and the bold edges are edges of  $F$ .

Between two consecutive affected vertices  $a_i$  and  $a_{i+1}$ , the interval of vertices of  $G$ , denoted by  $I$ , forms:

- Either a  $K$ -join, if  $I$  lies under an edge of  $F$ . For instance, on Figure 13, it corresponds to intervals of vertices between  $a_1$  and  $a_2$ , or between  $a_3$  and  $a_4$ , or between  $a_4$  and  $a_5$  or between  $a_6$  and  $a_7$ . So, by Observation 2.10, we know that such a  $I$  contains at most  $2k + 2$  vertices which are not contained in any claw or 4-cycle of  $G$ .
- Either a 1-branch or two disjoint 1-branch. If  $I$  lies at the beginning or at the end of  $\sigma_H$ , then  $I$  forms a 1-branch (for instance, on Figure 13, it corresponds to intervals of vertices before  $a_1$  or after  $a_7$ ). If  $I$  lies between two vertices  $a_i$  and  $a_{i+1}$  which are respectively the last (according to  $\sigma_H$ ) of a connected component of  $G$  and the first (according to  $\sigma_H$ ) of another connected component of  $G$ , then  $I$  forms two disjoint 1-branches (for instance, on Figure 13, it corresponds to the interval of vertices between  $a_2$  and  $a_3$ ). So, by Observation 2.15, we know that such a  $I$  contains at most  $2 \cdot (4k + 3) = 8k + 6$  vertices which are not contained in any claw or 4-cycle of  $G$ .
- Or a 2-branch, if  $I$  lies between two vertices  $a_i$  and  $a_{i+1}$  which belongs to the same connected component  $C$  of  $G$  and such that there is no edge of  $F$  standing above  $I$ . In this case the

2-branch  $B$  forms by the vertices of  $I$  is such that  $G[C \setminus B^R]$  is not connected, and then by Observation 2.17, we know that  $I$  contains at most  $12k + 10$  vertices which are not contained in any claw or 4-cycle of  $G$ .

Finally, as there is at most  $2k + 1$  such intervals  $I$ , the graph  $H$  (and hence  $G$ ) contains at most  $(2k + 1) \cdot (12k + 10)$  vertices different from the  $a_i$ 's and which are not contained in any claw or 4-cycle of  $G$ . Moreover, by Lemma 2.2, there is at most  $4k^3 + 15k^2 + 16k$  vertices of  $G$  contained in any claw or 4-cycle. Altogether,  $G$  contains at most  $4k^3 + 15k^2 + 16k + (2k + 1) \cdot (12k + 10) + 2k + 1 = 4k^3 + 39k^2 + 50k + 11$  vertices, which implies the claimed  $O(k^3)$  bound. The complexity directly follows from Lemma 2.22.  $\square$

### 3 A special case: BI-CLIQUE CHAIN COMPLETION

*Bipartite chain graphs* are defined as bipartite graphs whose parts are connected by a join. Equivalently, they are known to be the graphs that do not admit any  $\{2K_2, C_5, K_3\}$  as an induced subgraph [31] (see Figure 14). In [13], Guo proved that the so-called BIPARTITE CHAIN DELETION WITH FIXED BIPARTITION problem, where one is given a *bipartite* graph  $G = (V, E)$  and seeks a subset of  $E$  of size at most  $k$  whose deletion from  $E$  leads to a bipartite chain graph, admits a kernel with  $O(k^2)$  vertices. We define *bi-clique chain graph* to be the graphs formed by two disjoint cliques linked by a join. They correspond to interval graphs that can be covered by two cliques. Since the complement of a bipartite chain graph is a bi-clique chain graph, this result also holds for the BI-CLIQUE CHAIN COMPLETION WITH FIXED BI-CLIQUE PARTITION problem. Using similar techniques than in Section 2, we prove that when the bipartition is not fixed, both problems admit a quadratic-vertex kernel. For the sake of simplicity, we consider the completion version of the problem, defined as follows.

BI-CLIQUE CHAIN COMPLETION:

**Input:** A graph  $G = (V, E)$  and a positive integer  $k$ .

**Parameter:**  $k$ .

**Output:** A set  $F \subseteq (V \times V) \setminus E$  of size at most  $k$  such that the graph  $H = (V, E \cup F)$  is a bi-clique chain graph.

It follows from definition that bi-clique chain graphs do not admit any  $\{C_4, C_5, 3K_1\}$  as an induced subgraph, where a  $3K_1$  is an independent set of size 3 (see Figure 14). Observe in particular that bi-clique chain graphs are proper interval graphs, and hence admit an umbrella ordering.

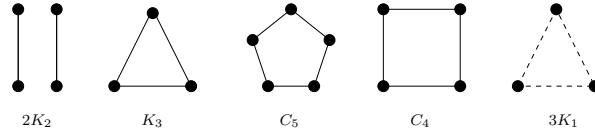


Figure 14: The forbidden induced subgraphs for bipartite and bi-clique chain graphs.

We provide a kernelization algorithm for the BI-CLIQUE CHAIN COMPLETION problem which follows the same lines that the one in Section 2.

**Rule 3.1** (Sunflower). Let  $\mathcal{S} = \{C_1, \dots, C_m\}$ ,  $m > k$  be a set of  $3K_1$  having two vertices  $u, v$  in common but distinct third vertex. Add  $uv$  to  $F$  and decrease  $k$  by 1.

Let  $\mathcal{S} = \{C_1, \dots, C_m\}$ ,  $m > k$  be a set of distinct 4-cycles having a non-edge  $uv$  in common. Add  $uv$  to  $F$  and decrease  $k$  by 1.

The following result is similar to Lemma 2.2.

**Lemma 3.1.** Let  $G = (V, E)$  be a positive instance of BI-CLIQUE CHAIN COMPLETION on which Rule 3.1 has been applied. There are at most  $k^2 + 2k$  vertices of  $G$  contained in  $3K_1$ 's. Furthermore, there are at most  $2k^2 + 2k$  vertices of  $G$  that are vertices of a 4-cycle.

We say that a  $K$ -join is *simple* whenever  $L = \emptyset$  or  $R = \emptyset$ . In other words, a simple  $K$ -join consists in a clique connected to the rest of the graph by a join. We will see it as a 1-branch which is a clique and use for it the classical notation devoted to the 1-branch. Moreover, we (re)define a *clean  $K$ -join* as a  $K$ -join whose vertices do not belong to any  $3K_1$  or 4-cycle. The following reduction rule is similar to Rule 2.4, the main ideas are identical, only some technical arguments change. Anyway, to be clear, we give the proof in all details.

**Rule 3.2** ( $K$ -join). Let  $B$  be a simple clean  $K$ -join of size at least  $2(k+1)$  associated with an umbrella ordering  $\sigma_B$ . Let  $B_L$  (resp.  $B_R$ ) be the  $k+1$  first (resp. last) vertices of  $B$  according to  $\sigma_B$ , and  $M = B \setminus (B_L \cup B_R)$ . Remove the set of vertices  $M$  from  $G$ .

**Lemma 3.2.** Rule 3.2 is safe and can be computed in polynomial time.

*Proof.* Let  $G' = G \setminus M$ . Observe that any  $k$ -completion of  $G$  is a  $k$ -completion of  $G'$  since bi-clique chain graphs are closed under induced subgraphs. So, let  $F$  be a  $k$ -completion for  $G'$ . We denote by  $H = G' + F$  the resulting bi-clique chain graph and by  $\sigma_H$  an umbrella ordering of  $H$ . We prove that we can always insert the vertices of  $M$  into  $\sigma_H$  and modify it if necessary, to obtain an umbrella ordering of a bi-clique chain graph for  $G$  without adding any edge. This will imply that  $F$  is a  $k$ -completion for  $G$ . To see this, we need the following structural property of  $G$ . As usual, we denote by  $R$  the neighbors in  $G \setminus B$  of the vertices of  $B$ , and by  $C$  the vertices of  $G \setminus (R \cup B)$ . For the sake of simplicity, we let  $N = \bigcap_{b \in B} N_G(b) \setminus B$ , and remove the vertices of  $N$  from  $R$ . We abusively still denote by  $R$  the set  $R \setminus N$ , see Figure 15.

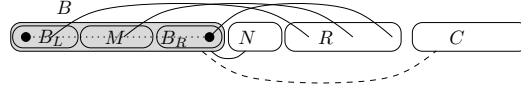


Figure 15: The  $K$ -join decomposition for the BI-CLIQUE CHAIN COMPLETION problem.

**Claim 3.3.** The set  $R \cup C$  is a clique of  $G$ .

*Proof.* Observe that no vertex of  $R$  is a neighbor of  $b_1$ , since otherwise such a vertex must be adjacent to all the vertices of  $B$  and then must stand in  $N$ . So, if  $R \cup C$  contains two vertices  $u, v$  such that  $uv \notin E$ , we form the  $3K_1$   $\{b_1, u, v\}$ , contradicting the fact that  $B$  is clean.  $\diamond$

The following observation comes from the definition of a simple  $K$ -join.

**Observation 3.4.** Given any vertex  $r \in R$ , if  $N_B(r) \cap B_L \neq \emptyset$  holds then  $M \subseteq N_B(r)$ .

We use these facts to prove that an umbrella ordering of a bi-clique chain graph can be obtained for  $G$  by inserting the vertices of  $M$  into  $\sigma_H$ . Let  $b_f, b_l$  be the first and last vertex of  $B \setminus M$  appearing in  $\sigma_H$ , respectively. We let  $B_H$  denote the set  $\{u \in V(H) : b_f <_{\sigma_H} u <_{\sigma_H} b_l\}$ . Now, we modify  $\sigma_H$  by ordering the twins in  $H$  according to their neighborhood in  $M$ : if  $x$  and  $y$  are twins in  $H$ , are consecutive in  $\sigma_H$ , verify  $x <_{\sigma_H} y <_{\sigma_H} b_f$  and  $N_M(y) \subset N_M(x)$ , then we exchange  $x$  and  $y$  in  $\sigma_H$ . This process stops when the considered twins are ordered following the join between  $\{u \in V(H) : u <_{\sigma_H} b_f\}$  and  $M$ . We proceed similarly on the right of  $B_H$ , i.e. for  $x$  and  $y$  consecutive twins with  $b_l <_{\sigma_H} x <_{\sigma_H} y$  and  $N_M(x) \subset N_M(y)$ . The obtained order is clearly an umbrella ordering of a bi-clique chain graph too (in fact, we just re-labeled some vertices in  $\sigma_H$ , and we abusively still denote it by  $\sigma_H$ ).

**Claim 3.5.** *The set  $B_H \cup \{m\}$  is a clique of  $G$  for any  $m \in M$ , and consequently  $B_H \cup M$  is a clique of  $G$ .*

*Proof.* Let  $u$  be any vertex of  $B_H$ . We claim that  $um \in E(G)$ . Observe that if  $u \in B$  then the claim trivially holds. So, assume that  $u \notin B$ . By definition of  $\sigma_H$ ,  $B_H$  is a clique in  $H$  since  $b_f b_l \in E(G)$ . It follows that  $u$  is incident to every vertex of  $B \setminus H$  in  $H$ . Since  $B_L$  contains  $k+1$  vertices, it follows that  $N_G(u) \cap B_L \neq \emptyset$ . Hence,  $u$  belongs to  $N \cup R$  and  $um \in E$  by Observation 2.6.  $\diamond$

**Claim 3.6.** *Let  $m$  be any vertex of  $M$  and  $\sigma'_H$  be the ordering obtained from  $\sigma_H$  by removing  $B_H$  and inserting  $m$  to the position of  $B_H$ . The ordering  $\sigma'_H$  respects the umbrella property.*

*Proof.* Assume that  $\sigma'_H$  does not respect the umbrella property, i.e. that there exist (w.l.o.g.) two vertices  $u, v \in H \setminus B_H$  such that either (1)  $u <_{\sigma'_H} v <_{\sigma'_H} m$ ,  $um \in E(H)$  and  $uv \notin E(H)$  or (2)  $u <_{\sigma'_H} m <_{\sigma'_H} v$ ,  $um \notin E(H)$  and  $uv \in E(H)$  or (3)  $u <_{\sigma'_H} v <_{\sigma'_H} m$ ,  $um \in E(H)$  and  $vm \notin E(H)$ . First, assume that (1) holds. Since  $uv \notin E$  and  $\sigma_H$  is an umbrella ordering,  $uw \notin E(H)$  for any  $w \in B_H$ , and hence  $uw \notin E(G)$ . This means that  $B_R \cap N_G(u) = \emptyset$ , which is impossible since  $um \in E(G)$ . If (2) holds, since  $uv \in E(H)$  and  $\sigma_H$  is an umbrella ordering of  $H$ , we have  $B_H \subseteq N_H(u)$ . In particular,  $B_L \subseteq N_H(u)$  holds, and as  $|B_L| = k+1$ , we have  $B_L \cap N_G(u) \neq \emptyset$  and  $um$  should be an edge of  $G$ , what contradicts the assumption  $um \notin E(H)$ . So, (3) holds, and we choose the first  $u$  satisfying this property according to the order given by  $\sigma'_H$ . So we have  $vm \notin E(G)$  for any  $w <_{\sigma'_H} u$ . Similarly, we choose  $v$  to be the first vertex satisfying  $vm \notin E(G)$ . Since  $um \in E(G)$ , we know that  $u$  belongs to  $N \cup R$ . Moreover, since  $vm \notin E(G)$ ,  $v \in R \cup C$ . There are several cases to consider:

- (i)  $u \in N$ : in this case we know that  $B \subseteq N_G(u)$ , and in particular that  $ub_l \in E(G)$ . Since  $\sigma_H$  is an umbrella ordering for  $H$ , it follows that  $vb_l \in E(H)$  and that  $B_L \subseteq N_H(v)$ . Since  $|B_L| = k+1$  we know that  $N_G(v) \cap B_L \neq \emptyset$  and hence  $v \in R$ . It follows from Observation 2.6 that  $vm \in E(G)$ .
- (ii)  $u \in R, v \in R \cup C$ : in this case  $uv \in E(G)$ , by Claim 3.3, but  $u$  and  $v$  are not true twins in  $H$  (otherwise  $v$  would be placed before  $u$  in  $\sigma_H$  due to the modification we have applied to  $\sigma_H$ ). This means that there exists a vertex  $w \in V(H)$  that *distinguishes*  $u$  from  $v$  in  $H$ .

Assume first that  $w <_{\sigma_H} u$  and that  $uw \in E(H)$  and  $vw \notin E(H)$ . We choose the first  $w$  satisfying this according to the order given by  $\sigma'_H$ . Since  $vm, wm, vw \notin E(H)$ , it follows that  $\{v, w, m\}$  defines a  $3K_1$  of  $G$ , which cannot be since  $B$  is clean. Hence we can assume that for any  $w'' <_{\sigma_H} u$ ,  $uw'' \in E(H)$  implies that  $vw'' \in E(H)$ . Now, suppose that  $b_l <_{\sigma_H} w$  and  $uw \notin E(H)$ ,  $vw \in E(H)$ . In particular, this means that  $B_L \subseteq N_H(v)$ . Since  $|B_L| = k+1$

we have  $N_G(v) \cap B_L \neq \emptyset$ , implying  $vm \in E(G)$  (Observation 2.6). Assume now that  $v <_{\sigma_H} w <_{\sigma_H} b_f$ . In this case, since  $uw \notin E(H)$ ,  $B \cap N_H(u) = \emptyset$  holds and hence  $B \cap N_G(u) = \emptyset$ , which cannot be since  $u \in R$ . Finally, assume that  $w \in B_H$  and choose the last vertex  $w$  satisfying this according to the order given by  $\sigma'_H$  (i.e.  $vw' \notin E(H)$  for any  $w <_{\sigma_H} w'$  and  $w' \in B_H$ ). If  $vw \in E(G)$  then  $\{u, m, w, v\}$  is a 4-cycle in  $G$  containing a vertex of  $B$ , which cannot be (recall that  $B_H \cup \{m\}$  is a clique of  $G$  by Claim 2.7). Hence  $vw \in F$  and there exists an extremal edge above  $vw$ . The only possibility is that this edge is some edge  $u'w$  for some  $u'$  with  $u' \in V(H)$ ,  $u <_{\sigma_H} u' <_{\sigma_H} v$  and  $u'w \in E(G)$ . By the choice of  $v$  we know that  $u'm \in E(G)$ . Moreover, by the choice of  $w$ , observe that  $u'$  and  $v$  are true twins in  $H$  (if a vertex  $s$  distinguishes  $u'$  and  $v$  in  $H$ ,  $s$  cannot be before  $u$ , since otherwise  $s$  would distinguish  $u$  and  $v$ , and not before  $w$ , by choice of  $w$ ). This leads to a contradiction because  $v$  should have been placed before  $u$  through the modification we have applied to  $\sigma_H$ .  $\diamond$

**Claim 3.7.** *Every vertex  $m \in M$  can be added to the graph  $H$  while preserving an umbrella ordering.*

*Proof.* Let  $m$  be any vertex of  $M$ . The graph  $H$  is a bi-clique chain graph. So, we know that in its associated umbrella ordering  $\sigma_H = b_1, \dots, b_{|H|}$ , there exists a vertex  $b_i$  such that  $H_1 = \{b_1, \dots, b_i\}$  and  $H_2 = \{b_{i+1}, \dots, b_{|H|}\}$  are two cliques of  $H$  linked by a join. We study the behavior of  $B_H$  according to the partition  $(H_1, H_2)$ .

- (i) Assume first that  $B_H \subseteq H_1$  (the case  $B_H \subseteq H_2$  is similar). We claim that the set  $H_1 \cup \{m\}$  is a clique. Indeed, let  $v \in H_1 \setminus B_H$ : since  $H_1$  is a clique,  $B_H \subseteq N_H(v)$  and hence  $N_G(v) \cap B_L \neq \emptyset$ . In particular, this means that  $vm \in E(G)$  by Observation 3.4. Since  $B_H \cup \{m\}$  is a clique by Claim 3.5, the result follows. Now, let  $u$  be the neighbor of  $m$  with maximal index in  $\sigma_H$ , and  $b_u$  the neighbor of  $u$  with minimal index in  $\sigma_H$ . Observe that we may assume  $u \in H_2$  since otherwise  $N_H(m) \cap H_2 = \emptyset$  and hence we insert  $m$  at the beginning of  $\sigma_H$ . First, if  $b_u \in H_1$ , we prove that the order  $\sigma_m$  obtained by inserting  $m$  directly before  $b_u$  in  $\sigma_H$  yields an umbrella ordering of a bi-clique chain graph. Since  $H_1 \cup \{m\}$  is a clique, we only need to show that  $N_{H_2}(v) \subseteq N_{H_2}(m)$  for any  $v \leq_{\sigma_m} b_u$  and  $N_{H_2}(m) \subseteq N_{H_2}(w)$  for any  $w \in H_2$  with  $w \geq_{\sigma_m} b_u$ . Observe that by Claim 3.6 the set  $\{w \in V : m \leq_{\sigma_m} w \leq_{\sigma_m} u\}$  is a clique. Hence the former case holds since  $vu' \notin E(G)$  for any  $v \leq_{\sigma_m} b_u$  and  $u' \geq_{\sigma_m} u$ . The latter case also holds since  $N_H(m) \subseteq N_H(b_u)$  by construction. Finally, if  $b_u \in H_2$ , then  $b_u = b_{|H_1|+1}$  since  $H_2$  is a clique. Hence, using similar arguments one can see that inserting  $m$  directly after  $b_{|H_1|}$  in  $\sigma_H$  yields an umbrella ordering of a bi-clique chain graph.
- (ii) Assume now that  $B_H \cap H_1 \neq \emptyset$  and  $B_H \cap H_2 \neq \emptyset$ . In this case, we claim that  $H_1 \cup \{m\}$  or  $H_2 \cup \{m\}$  is a clique in  $H$ . Let  $u$  and  $u'$  be the neighbors of  $m$  with minimal and maximal index in  $\sigma_H$ , respectively. If  $u = b_1$  or  $u' = b_{|H|}$  then Claims 3.5 and 3.6 imply that  $H_1 \cup \{m\}$  or  $H_2 \cup \{m\}$  is a clique and we are done. So, none of these two conditions hold and  $mb_1 \notin E(H)$  and  $mb_{|H|} \notin E(H)$ . Then, by Claim 3.6, we know that  $b_1 b_{|H|}$  and the set  $\{b_1, b_{|H|}, m\}$  defines a  $3K_1$  containing  $m$  in  $G$ , which cannot be. This means that we can assume w.l.o.g. that  $H_1 \cup \{m\}$  is a clique, and we can conclude using similar arguments than in (i).  $\diamond$

Since the proof of Claim 3.7 does not use the fact that the vertices of  $H$  do not belong to  $M$ , it follows that we can iteratively insert the vertices of  $M$  into  $\sigma_H$ , preserving an umbrella ordering at each step. To conclude, observe that the reduction rule can be computed in polynomial time using Lemma 2.19.  $\square$

**Observation 3.8.** *Let  $G = (V, E)$  be a positive instance of BI-CLIQUE CHAIN COMPLETION reduced under Rule 3.2. Any simple  $K$ -join  $B$  of  $G$  has size at most  $3k^2 + 6k + 2$ .*

*Proof.* Let  $B$  be any simple  $K$ -join of  $G$ , and assume  $|B| > 3k^2 + 6k + 2$ . By Lemma 3.1 we know that at most  $3k^2 + 2k$  vertices of  $B$  are contained in a  $3K_1$  or a 4-cycle. Hence  $B$  contains a set  $B'$  of at least  $2k + 3$  vertices not contained in any  $3K_1$  or a 4-cycle. Now, since any subset of a  $K$ -join is a  $K$ -join, it follows that  $B'$  is a *clean* simple  $K$ -join. Since  $G$  is reduced under rule 3.2, we know that  $|B'| \leq 2(k + 1)$  what gives a contradiction.  $\square$

Finally, we can prove that Rules 3.1 and 3.2 form a kernelization algorithm.

**Theorem 3.9.** *The BI-CLIQUE CHAIN COMPLETION problem admits a kernel with  $O(k^2)$  vertices.*

*Proof.* Let  $G = (V, E)$  be a positive instance of BI-CLIQUE CHAIN COMPLETION reduced under Rules 3.1 and 3.2, and  $F$  be a  $k$ -completion for  $G$ . We let  $H = G + F$  and  $H_1, H_2$  be the two cliques of  $H$ . Observe in particular that  $H_1$  and  $H_2$  both define simple  $K$ -joins. Let  $A$  be the set of affected vertices of  $G$ . Since  $|F| \leq k$ , observe that  $|A| \leq 2k$ . Let  $A_1 = A \cap H_1$ ,  $A_2 = A \cap H_2$ ,  $A'_1 = H_1 \setminus A_1$  and  $A'_2 = H_2 \setminus A_2$  (see Figure 16). Observe that since  $H_1$  is a simple  $K$ -join in  $H$ ,  $A'_1 \subseteq H_1$  is a simple  $K$ -join of  $G$  (recall that the vertices of  $A'_1$  are not affected). By Observation 3.8, it follows that  $|A'_1| \leq 3k^2 + 6k + 2$ . The same holds for  $A'_2$  and  $H$  contains at most  $2(3k^2 + 6k + 2) + 2k$  vertices.

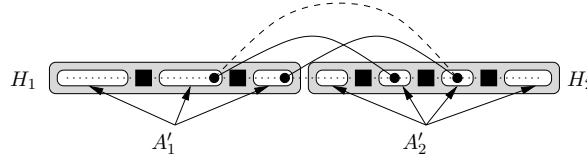


Figure 16: Illustration of the bi-clique chain graph  $H$ . The square vertices stand for affected vertices, and the sets  $A'_1 = H_1 \setminus A_1$  and  $A'_2 = H_2 \setminus A_2$  are simple  $K$ -joins of  $G$ , respectively.

$\square$

**Corollary 3.10.** *The BIPARTITE CHAIN DELETION problem admits a kernel with  $O(k^2)$  vertices.*

## 4 Conclusion

In this paper we prove that the PROPER INTERVAL COMPLETION problem admits a kernel with  $O(k^3)$  vertices. Two natural questions arise from our results: firstly, does the INTERVAL COMPLETION problem admit a polynomial kernel? Observe that this problem is known to be FPT not for long [29]. The techniques we developed here intensively use the fact that there are few claws in the graph, what help us to reconstruct parts of the umbrella ordering. Of course, these considerations no more hold in general interval graphs. The second question is: does the PROPER INTERVAL EDGE-DELETION problem admit a polynomial kernel? Again, this problem admits a fixed-parameter algorithm [27], and we believe that our techniques could be applied to this problem as well. Finally, we proved that the BI-CLIQUE CHAIN COMPLETION problem admits a kernel with  $O(k^2)$  vertices, which completes a result of Guo [13]. In all cases, a natural question is thus whether these bounds can be improved?

## References

- [1] S. Bessy, C. Paul, and A. Perez. Polynomial kernels for 3-leaf power graph modification problems. *Discrete Applied Mathematics*, 158(16):1732–1744, 2010.
- [2] H. L. Bodlaender. Kernelization: New upper and lower bound techniques. In *IWPEC*, pages 17–37, 2009.
- [3] H. L. Bodlaender, R. G. Downey, M. R. Fellows, and D. Hermelin. On problems without polynomial kernels. *J. Comput. Syst. Sci.*, 75(8):423–434, 2009.
- [4] L. Cai. Fixed-parameter tractability of graph modification problems for hereditary properties. *Inf. Process. Lett.*, 58(4):171–176, 1996.
- [5] J. Chen and J. Meng. A  $2k$  kernel for the cluster editing problem. In *COCOON*, volume 6196 of *LNCS*, pages 459–468, 2010.
- [6] D. G. Corneil. A simple 3-sweep LBFS algorithm for the recognition of unit interval graphs. *Discrete Appl. Math.*, 138:371–379, April 2004.
- [7] D. G. Corneil, H. Kim, S. Natarajan, S. Olariu, and A. P. Sprague. Simple linear time recognition of unit interval graphs. *Information Processing Letters*, 55(2):99 – 104, 1995.
- [8] F. K. Dehne, M. R. Fellows, F. A. Rosamond, and P. Shaw. Greedy localization, iterative compression, modeled crown reductions: New FPT techniques, an improved algorithm for set splitting, and a novel  $2k$  kernelization for vertex cover. In *IWPEC*, volume 3162 of *LNCS*, pages 271–280, 2004.
- [9] X. Deng, P. Hell, and J. Huang. Linear-time representation algorithms for proper circular-arc graphs and proper interval graphs. *SIAM J. Comput.*, 25(2):390–403, 1996.
- [10] R. G. Downey and M. R. Fellows. *Parameterized complexity*. Springer, 1999.
- [11] M. C. Golumbic, H. Kaplan, and R. Shamir. On the complexity of DNA physical mapping. *ADVAM: Advances in Applied Mathematics*, 15, 1994.
- [12] S. Guillemot, C. Paul, and A. Perez. On the (non-)existence of polynomial kernels for  $P_t$ -free edge modification problems. In *IPEC*, volume 6478 of *LNCS*, pages 147–157, 2010.
- [13] J. Guo. Problem kernels for NP-complete edge deletion problems: Split and related graphs. In *ISAAC*, volume 4835 of *LNCS*, pages 915–926, 2007.
- [14] M. Habib and C. Paul. A survey on algorithmic aspects of modular decomposition. *Computer Science Review*, 4(1):41–59, 2010.
- [15] P. Hell, R. Shamir, and R. Sharan. A fully dynamic algorithm for recognizing and representing proper interval graphs. *SIAM Journal on Computing*, 31(1):289–305, 2001.
- [16] H. Kaplan, R. Shamir, and R. E. Tarjan. Tractability of parameterized completion problems on chordal and interval graphs: Minimum fill-in and physical mapping. In *FOCS*, pages 780–791, 1994.



- [17] H. Kaplan, R. Shamir, and R. E. Tarjan. Tractability of parameterized completion problems on chordal, strongly chordal, and proper interval graphs. *SIAM J. Comput.*, 28(5):1906–1922, 1999.
- [18] S. Kratsch and M. Wahlström. Two edge modification problems without polynomial kernels. In *IWPEC*, volume 5917 of *LNCS*, pages 264–275. Springer, 2009.
- [19] P. J. Looges and S. Olariu. Optimal greedy algorithms for indifference graphs. *Computers & Mathematics with Applications*, 25(7):15 – 25, 1993.
- [20] F. Mancini. *Graph modification problems related to graph classes*. PhD thesis, University of Bergen, Norway, 2008.
- [21] R. Niedermeier. *Invitation to fixed parameter algorithms*, volume 31 of *Oxford Lectures Series in Mathematics and its Applications*. Oxford University Press, 2006.
- [22] R. Niedermeier and P. Rossmanith. A general method to speed up fixed-parameter-tractable algorithms. *Inf. Process. Lett.*, 73(3-4):125–129, 2000.
- [23] R. Shamir, R. Sharan, and D. Tsur. Cluster graph modification problems. *Discrete Applied Mathematics*, 144(1-2):173–182, 2004.
- [24] R. Sharan. *Graph modification problems and their applications to genomic research*. PhD thesis, Tel-Aviv University, 2002.
- [25] R. E. Tarjan and M. Yannakakis. Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM J. Comput.*, 13(3):566–579, 1984.
- [26] S. Thomassé. A  $4k^2$  kernel for feedback vertex set. *ACM Transactions on Algorithms*, 6(2), 2010.
- [27] Y. Villanger. [www.lirmm.fr/~paul/ANR/CIRM-TALKS-2010/Villanger-cirm-2010.pdf](http://www.lirmm.fr/~paul/ANR/CIRM-TALKS-2010/Villanger-cirm-2010.pdf), 2010.
- [28] Y. Villanger. Proper interval vertex deletion. In *IPEC*, volume 6478 of *LNCS*, pages 228–238, 2010.
- [29] Y. Villanger, P. Heggernes, C. Paul, and J. A. Telle. Interval completion is fixed parameter tractable. *SIAM J. Comput.*, 38(5):2007–2020, 2009.
- [30] G. Wegner. *Eigenschaften der nerven homologische-einfactor familien in  $\mathbb{R}^n$* . PhD thesis, Universität Göttingen, Göttingen, Germany, 1967.
- [31] M. Yannakakis. Computing the minimum fill-in is NP-Complete. *SIAM J. Alg. and Discr. Meth.*, 2(1):77–79, 1981.

## 2.9 A poly-kernel for FAS in tournaments

Foundations of Software Technology and Theoretical Computer Science (2009) Submission

# Kernels for Feedback Arc Set In Tournaments

**Stéphane Bessy<sup>1</sup>, Fedor V. Fomin<sup>2</sup>, Serge Gaspers<sup>1</sup>, Christophe Paul<sup>1</sup>, Anthony Perez<sup>1</sup>, Saket Saurabh<sup>2</sup>, Stéphan Thomassé<sup>1</sup>**

<sup>1</sup> LIRMM – Université de Montpellier 2, CNRS,  
Montpellier, France.

{bessy|gaspers|paul|perez|thomasse}@lirmm.fr

<sup>2</sup> Department of Informatics, University of Bergen,  
Bergen, Norway.

{fedor.fomin|saket.saurabh}@ii.uib.no

**ABSTRACT.** A tournament  $T = (V, A)$  is a directed graph in which there is exactly one arc between every pair of distinct vertices. Given a digraph on  $n$  vertices and an integer parameter  $k$ , the **FEEDBACK ARC SET** problem asks whether the given digraph has a set of  $k$  arcs whose removal results in an acyclic digraph. The **FEEDBACK ARC SET** problem restricted to tournaments is known as the  **$k$ -FEEDBACK ARC SET IN TOURNAMENTS ( $k$ -FAST)** problem. In this paper we obtain a linear vertex kernel for  $k$ -FAST. That is, we give a polynomial time algorithm which given an input instance  $T$  to  $k$ -FAST obtains an equivalent instance  $T'$  on  $O(k)$  vertices. In fact, given any fixed  $\epsilon > 0$ , the kernelized instance has at most  $(2 + \epsilon)k$  vertices. Our result improves the previous known bound of  $O(k^2)$  on the kernel size for  $k$ -FAST. For our kernelization algorithm we find a subclass of tournaments where one can find a minimum sized feedback arc set in polynomial time and use the known polynomial time approximation scheme for  $k$ -FAST.

## 1 Introduction

Given a directed graph  $G = (V, A)$  on  $n$  vertices and an integer parameter  $k$ , the **FEEDBACK ARC SET** problem asks whether the given digraph has a set of  $k$  arcs whose removal results in an acyclic directed graph. In this paper, we consider this problem in a special class of directed graphs, *tournaments*. A tournament  $T = (V, A)$  is a directed graph in which there is exactly one directed arc between every pair of vertices. More formally the problem we consider is defined as follows.

**$k$ -FEEDBACK ARC SET IN TOURNAMENTS ( $k$ -FAST):** Given a tournament  $T = (V, A)$  and a positive integer  $k$ , does there exist a subset  $F \subseteq A$  of at most  $k$  arcs whose removal makes  $T$  acyclic.

In the weighted version of  $k$ -FAST, we are also given integer weights (each weight is at least one) on the arcs and the objective is to find a feedback arc set of weight at most  $k$ . This problem is called  **$k$ -WEIGHTED FEEDBACK ARC SET IN TOURNAMENTS ( $k$ -WFAST)**.

Feedback arc sets in tournaments are well studied from the combinatorial [16, 17, 23, 24, 26, 30], statistical [25] and algorithmic [1, 2, 11, 20, 28, 29] points of view. The problems  $k$ -FAST and  $k$ -WFAST have several applications. In *rank aggregation* we are given several

NOT FOR DISTRIBUTION

## 2 KERNELS FOR FEEDBACK ARC SET IN TOURNAMENTS

rankings of a set of objects, and we wish to produce a single ranking that on average is as consistent as possible with the given ones, according to some chosen measure of consistency. This problem has been studied in the context of voting [6, 10], machine learning [9], and search engine ranking [14, 15]. A natural consistency measure for rank aggregation is the number of pairs that occur in a different order in the two rankings. This leads to *Kemeney-Young rank aggregation* [18, 19], a special case of  $k$ -WFAST.

The  $k$ -FAST problem is known to be NP-complete by recent results of Alon [2] and Charbit *et al.* [8] while  $k$ -WFAST is known to be NP-complete by Dwork *et al.* [14, 15]. From an approximation perspective,  $k$ -WFAST admits a polynomial time approximation scheme [20]. The problem is also well studied in parameterized complexity. In this area, a problem with input size  $n$  and a parameter  $k$  is said to be fixed parameter tractable (FPT) if there exists an algorithm to solve this problem in time  $f(k) \cdot n^{O(1)}$ , where  $f$  is an arbitrary function of  $k$ . Raman and Saurabh [22] showed that  $k$ -FAST and  $k$ -WFAST are FPT by obtaining an algorithm running in time  $O(2.415^k \cdot k^{4.752} + n^{O(1)})$ . Recently, Alon *et al.* [3] have improved this result by giving an algorithm for  $k$ -WFAST running in time  $O(2^{O(\sqrt{k} \log k)} + n^{O(1)})$ . This algorithm runs in sub-exponential time, a trait uncommon to parameterized algorithms. In this paper we investigate  $k$ -FAST from the view point of kernelization, currently one of the most active subfields of parameterized algorithms.

A parameterized problem is said to admit a *polynomial kernel* if there is a polynomial time algorithm, called a *kernelization* algorithm, that reduces the input instance to an instance whose size is bounded by a polynomial  $p(k)$  in  $k$ , while preserving the answer. This reduced instance is called a  $p(k)$  *kernel* for the problem. When  $p(k)$  is a linear function of  $k$  then the corresponding kernel is a linear kernel. Kernelization has been at the forefront of research in parameterized complexity in the last couple of years, leading to various new polynomial kernels as well as tools to show that several problems do not have a polynomial kernel under some complexity-theoretic assumptions [4, 5, 7, 13, 27]. In this paper we continue the current theme of research on kernelization and obtain a *linear vertex* kernel for  $k$ -FAST. That is, we give a polynomial time algorithm which given an input instance  $T$  to  $k$ -FAST obtains an equivalent instance  $T'$  on  $O(k)$  vertices. More precisely, given any fixed  $\epsilon > 0$ , we find a kernel with at most  $(2 + \epsilon)k$  vertices in polynomial time. The reason we call it a linear *vertex* kernel is that, even though the number of vertices in the reduced instance is at most  $O(k)$ , the number of arcs is still  $O(k^2)$ . Our result improves the previous known bound of  $O(k^2)$  on the vertex kernel size for  $k$ -FAST [3, 12]. For our kernelization algorithm we find a subclass of tournaments where one can find a minimum sized feedback arc set in polynomial time (see Lemma 12) and use the known polynomial time approximation scheme for  $k$ -FAST by Kenyon-Mathieu and Schudy [20]. The polynomial time algorithm for a subclass of tournaments could be of independent interest.

The paper is organized as follows. In Section 2, we give some definition and preliminary results regarding feedback arc sets. In Section 3 we give a linear vertex kernel for  $k$ -FAST. Finally we conclude with some remarks in Section 4.

## 2 Preliminaries

Let  $T = (V, A)$  be a tournament on  $n$  vertices. We use  $T_\sigma = (V_\sigma, A)$  to denote a tournament whose vertices are ordered under a fixed ordering  $\sigma = v_1, \dots, v_n$  (we also use  $D_\sigma$  for an ordered directed graph). We say that an arc  $v_i v_j$  of  $T_\sigma$  is a *backward arc* if  $i > j$ , otherwise we call it a *forward arc*. Moreover, given any partition  $\mathcal{P} := \{V_1, \dots, V_l\}$  of  $V_\sigma$ , where every  $V_i$  is an interval according to the ordering of  $T_\sigma$ , we use  $A_B$  to denote all arcs between the intervals (having their endpoints in different intervals), and  $A_I$  for all arcs within the intervals. If  $T_\sigma$  contains no backward arc, then we say that it is *transitive*.

For a vertex  $v \in V$  we denote its *in-neighborhood* by  $N^-(v) := \{u \in V \mid uv \in A\}$  and its *out-neighborhood* by  $N^+(v) := \{u \in V \mid vu \in A\}$ . A set of vertices  $M \subseteq V$  is a *module* if and only if  $N^+(u) \setminus M = N^+(v) \setminus M$  for every  $u, v \in M$ . For a subset of arcs  $A' \subseteq A$ , we define  $T[A']$  to be the digraph  $(V', A')$  where  $V'$  is the union of endpoints of the arcs in  $A'$ . Given an ordered digraph  $D_\sigma$  and an arc  $e = v_i v_j$ ,  $S(e) = \{v_i, \dots, v_j\}$  denotes the *span* of  $e$ . The number of vertices in  $S(e)$  is called the *length* of  $e$  and is denoted by  $l(e)$ . Thus, for every arc  $e = v_i v_j$ ,  $l(e) = |i - j| + 1$ . Finally, for every vertex  $v$  in the span of  $e$ , we say that  $e$  is *above*  $v$ .

In this paper, we will use the well-known fact that every acyclic tournament admits a transitive ordering. In particular, we will consider *maximal transitive modules*. We also need the following result for our kernelization algorithm.

**LEMMA 1.** ([22]) *Let  $D = (V, A)$  be a directed graph and  $F$  be a minimal feedback arc set of  $D$ . Let  $D'$  be the graph obtained from  $D$  by reversing the arcs of  $F$  in  $D$ , then  $D'$  is acyclic.*

In this paper whenever we say *circuit*, we mean a directed cycle. Next we introduce a definition which is useful for a lemma we prove later.

**DEFINITION 2.** *Let  $D_\sigma = (V_\sigma, A)$  be an ordered directed graph and let  $f = vu$  be a backward arc of  $D_\sigma$ . We call certificate of  $f$ , and denote it by  $c(f)$ , any directed path from  $u$  to  $v$  using only forward arcs in the span of  $f$  in  $D_\sigma$ .*

Observe that such a directed path together with the backward arc  $f$  forms a directed cycle in  $D_\sigma$  whose only backward arc is  $f$ .

**DEFINITION 3.** *Let  $D_\sigma = (V_\sigma, A)$  be an ordered directed graph, and let  $F \subseteq A$  be a set of backward arcs of  $D_\sigma$ . We say that we can certify  $F$  whenever it is possible to find a set  $\mathcal{F} = \{c(f) : f \in F\}$  of arc-disjoint certificates for the arcs in  $F$ .*

Let  $D_\sigma = (V_\sigma, A)$  be an ordered directed graph, and let  $F \subseteq A$  be a subset of backward arcs of  $D_\sigma$ . We say that we can certify the set  $F$  using *only arcs from*  $A' \subseteq A$  if  $F$  can be certified by a collection  $\mathcal{F}$  such that the union of the arcs of the certificates in  $\mathcal{F}$  is contained in  $A'$ . In the following,  $fas(D)$  denotes the *size* of a minimum feedback arc set, that is, the cardinality of a minimum sized set  $F$  of arcs whose removal makes  $D$  acyclic.

**LEMMA 4.** *Let  $D_\sigma$  be an ordered directed graph, and let  $\mathcal{P} = \{V_1, \dots, V_l\}$  be a partition of  $D_\sigma$  into intervals. Assume that the set  $F$  of all backward arcs of  $D_\sigma[A_B]$  can be certified using only arcs from  $A_B$ . Then  $fas(D_\sigma) = fas(D_\sigma[A_I]) + fas(D_\sigma[A_B])$ . Moreover, there exists a minimum sized feedback arc set of  $D_\sigma$  containing  $F$ .*

**PROOF.** For any bipartition of the arc set  $A$  into  $A_1$  and  $A_2$ ,  $fas(D_\sigma) \geq fas(D_\sigma[A_1]) + fas(D_\sigma[A_2])$ . Hence, in particular for a partition of the arc set  $A$  into  $A_I$  and  $A_B$  we have

#### 4 KERNELS FOR FEEDBACK ARC SET IN TOURNAMENTS

that  $fas(D_\sigma) \geq fas(D_\sigma[A_I]) + fas(D_\sigma[A_B])$ . Next, we show that  $fas(D_\sigma) \leq fas(D_\sigma[A_I]) + fas(D_\sigma[A_B])$ . This follows from the fact that once we reverse all the arcs in  $F$ , each remaining circuit lies in  $D_\sigma[V_i]$  for some  $i \in \{1, \dots, l\}$ . In other words once we reverse all the arcs in  $F$ , every circuit is completely contained in  $D_\sigma[A_I]$ . This concludes the proof of the first part of the lemma. In fact, what we have shown is that there exists a minimum sized feedback arc set of  $D_\sigma$  containing  $F$ . This concludes the proof of the lemma.  $\blacksquare$

### 3 Kernels for $k$ -FAST

In this section we first give a subquadratic vertex kernel of size  $O(k\sqrt{k})$  for  $k$ -FAST and then improve on it to get our final vertex kernel of size  $O(k)$ . We start by giving a few reduction rules that will be needed to bound the size of the kernels.

**Rule 3.1** *If a vertex  $v$  is not contained in any triangle, delete  $v$  from  $T$ .*

**Rule 3.2** *If there exists an arc  $uv$  that belongs to more than  $k$  distinct triangles, then reverse  $uv$  and decrease  $k$  by 1.*

We say that a reduction rule is *sound*, if whenever the rule is applied to an instance  $(T, k)$  to obtain an instance  $(T', k')$ ,  $T$  has a feedback arc set of size at most  $k$  if and only if  $T'$  has a feedback arc set of size at most  $k'$ .

**LEMMA 5.** *([3, 12]) Rules 3.1 and 3.2 are sound and can be applied in polynomial time.*

The Rules 3.1 and 3.2 together led to a quadratic kernel for  $k$ -WFAST [3]. Earlier, these rules were used by Dom *et al.* [12] to obtain a quadratic kernel for  $k$ -FAST. We now add a new reduction rule that will allow us to obtain the claimed bound on the kernel sizes for  $k$ -FAST. Given an ordered tournament  $T_\sigma = (V_\sigma, A)$ , we say that  $\mathcal{P} = \{V_1, \dots, V_l\}$  is a *safe partition* of  $V_\sigma$  into intervals whenever it is possible to certify the backward arcs of  $T_\sigma[A_B]$  using only arcs from  $A_B$ .

**Rule 3.3** *Let  $T_\sigma$  be an ordered tournament,  $\mathcal{P} = \{V_1, \dots, V_l\}$  be a safe partition of  $V_\sigma$  into intervals and  $F$  be the set of backward arcs of  $T_\sigma[A_B]$ . Then reverse all the arcs of  $F$  and decrease  $k$  by  $|F|$ .*

**LEMMA 6.** *Rule 3.3 is sound.*

**PROOF.** Let  $\mathcal{P}$  be a safe partition of  $T_\sigma$ . Observe that it is possible to certify all the backward arcs, that is  $F$ , using only arcs in  $A_B$ . Hence using Lemma 4 we have that  $fas(T_\sigma) = fas(T_\sigma[A_I]) + fas(T_\sigma[A_B])$ . Furthermore, by Lemma 4 we also know that there exists a minimum sized feedback arc set of  $D_\sigma$  containing  $F$ . Thus,  $T_\sigma$  has a feedback arc set of size at most  $k$  if and only if the tournament  $T'_\sigma$  obtained from  $T_\sigma$  by reversing all the arcs of  $F$  has a feedback arc set of size at most  $k - |F|$ .  $\blacksquare$

#### 3.1 A subquadratic kernel for $k$ -FAST

In this section, we show how to obtain an  $O(k\sqrt{k})$  sized vertex kernel for  $k$ -FAST. To do so, we introduce the following reduction rule.

**Rule 3.4** Let  $V_m$  be a maximal transitive module of size  $p$ , and  $I$  and  $O$  be the set of in-neighbors and out-neighbors of the vertices of  $V_m$  in  $T$ , respectively. Let  $Z$  be the set of arcs  $uv$  such that  $u \in O$  and  $v \in I$ . If  $q = |Z| < p$  then reverse all the arcs in  $Z$  and decrease  $k$  by  $q$ .

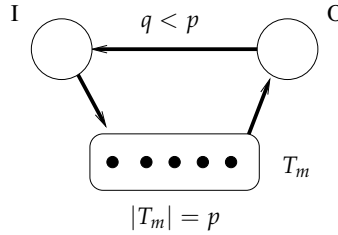


Figure 1: A transitive module on which Rule 3.4 applies.

**LEMMA 7.** Rule 3.4 is sound and can be applied in linear time.

**PROOF.** We first prove that the partition  $\mathcal{P} = \{I, V_m, O\}$  forms a safe partition of the input tournament. Let  $V'_m = \{w_1, \dots, w_q\} \subseteq V_m$  be an arbitrary subset of size  $q$  of  $V_m$  and let  $Z = \{u_i v_i \mid 1 \leq i \leq q\}$ . Consider the collection  $\mathcal{F} = \{v_i w_i u_i \mid u_i v_i \in Z, w_i \in V'_m\}$  and notice that it certifies all the arcs in  $Z$ . In fact we have managed to certify all the backwards arcs of the partition using only arcs from  $A_B$  and hence  $\mathcal{P}$  forms a safe partition. Thus, by Rule 3.3, it is safe to reverse all the arcs from  $O$  to  $I$ . The time complexity follows from the fact that computing a modular decomposition tree can be done in  $O(n + m)$  time on directed graphs [21]. ■

We show that any YES-instance to which none of the Rules 3.1, 3.2 and 3.4 could be applied has at most  $O(k\sqrt{k})$  vertices.

**THEOREM 8.** Let  $(T = (V, A), k)$  be a YES-instance to  $k$ -FAST which has been reduced according to Rules 3.1, 3.2 and 3.4. Then  $T$  has at most  $O(k\sqrt{k})$  vertices.

**PROOF.** Let  $S$  be a feedback arc set of size at most  $k$  of  $T$  and let  $T'$  be the tournament obtained from  $T$  by reversing all the arcs in  $S$ . Let  $\sigma$  be the transitive ordering of  $T'$  and  $T_\sigma = (V_\sigma, A)$  be the ordered tournament corresponding to the ordering  $\sigma$ . We say that a vertex is *affected* if it is incident to some arc in  $S$ . Thus, the number of affected vertices is at most  $2|S| \leq 2k$ . The reduction Rule 3.1 ensures that the first and last vertex of  $T_\sigma$  are affected. To see this note that if the first vertex in  $V_\sigma$  is not affected then it is a source vertex (vertex with in-degree 0) and hence it is not part of any triangle and thus Rule 3.1 would have applied. We can similarly argue for the last vertex. Next we argue that there is no backward arc  $e$  of length greater than  $2k + 2$  in  $T_\sigma$ . Assume to the contrary that  $e = uv$  is a backward arc with  $S(e) = \{v, x_1, x_2, \dots, x_{2k+1}, \dots, u\}$  and hence  $l(e) > 2k + 2$ . Consider the collection  $\mathcal{T} = \{vx_i u \mid 1 \leq i \leq 2k\}$  and observe that at most  $k$  of these triples can contain an arc from  $S \setminus \{e\}$  and hence there exist at least  $k + 1$  triplets in  $\mathcal{T}$  which corresponds to distinct triangles all containing  $e$ . But then  $e$  would have been reversed by an application

## 6 KERNELS FOR FEEDBACK ARC SET IN TOURNAMENTS

of Rule 3.2. Hence, we have shown that there is no backward arc  $e$  of length greater than  $2k + 2$  in  $T_\sigma$ . Thus  $\sum_{e \in S} l(e) \leq 2k^2 + 2k$ .

We also know that between two consecutive affected vertices there is exactly one maximal transitive module. Let us denote by  $t_i$  the number of vertices in these modules, where  $i \in \{1, \dots, 2k-1\}$ . The objective here is to bound the number of vertices in  $V_\sigma$  or  $V$  using  $\sum_{i=1}^{2k-1} t_i$ . To do so, observe that since  $T$  is reduced under the Rule 3.4, there are at least  $t_i$  backward arcs above every module with  $t_i$  vertices, each of length at least  $t_i$ . This implies that  $\sum_{i=1}^{2k-1} t_i^2 \leq \sum_{e \in S} l(e) \leq 2k^2 + 2k$ . Now, using the Cauchy-Schwarz inequality we can show the following.

$$\sum_{i=1}^{2k-1} t_i = \sum_{i=1}^{2k-1} t_i \cdot 1 \leq \sqrt{\sum_{i=1}^{2k-1} t_i^2 \cdot \sum_{i=1}^{2k-1} 1} \leq \sqrt{(2k^2 + 2k) \cdot (2k-1)} = \sqrt{4k^3 + 2k^2 - k}.$$

Thus every reduced YES-instance has at most  $\sqrt{4k^3 + 2k^2 - k} + 2k = O(k\sqrt{k})$  vertices.  $\blacksquare$

### 3.2 A linear kernel for $k$ -FAST

We begin this subsection by showing some general properties about tournaments which will be useful in obtaining a linear kernel for  $k$ -FAST.

#### Backward Weighted Tournaments

Let  $T_\sigma$  be an ordered tournament with weights on its backward arcs. We call such a tournament a *backward weighted tournament* and denote it by  $T_\omega$ , and use  $\omega(e)$  to denote the weight of a backward arc  $e$ . For every interval  $I := [v_i, \dots, v_j]$  we use  $\omega(I)$  to denote the total weight of all backward arcs having both their endpoints in  $I$ , that is,  $\omega(I) = \sum_{e=uv} \omega(e)$  where  $u, v \in I$  and  $e$  is a backward arc.

**DEFINITION 9.(Contraction)** Let  $T_\omega = (V_\sigma, A)$  be an ordered tournament with weights on its backward arcs and  $I = [v_i, \dots, v_j]$  be an interval. The contracted tournament is defined as  $T_{\omega'} = (V_{\sigma'}, A')$  where  $V_{\sigma'} = V_\sigma \setminus \{I\} \cup \{c_I\}$ . The arc set  $A'$  is defined as follows.

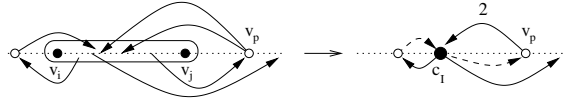
- It contains all the arcs  $A_1 = \{uv \mid uv \in A, u \notin I, v \notin I\}$
- Add  $A_2 = \{uc_I \mid uv \in A, u \notin I, v \in I\}$  and  $A_3 = \{c_Iv \mid uv \in A, u \in I, v \notin I\}$ .
- Finally, we remove every forward arc involved in a 2-cycle after the addition of arcs in the previous step.

The order  $\sigma'$  for  $T_{\omega'}$  is provided by  $\sigma' = v_1, \dots, v_{i-1}, c_I, v_{j+1}, \dots, v_n$ . We define the weight of a backward arc  $e = xy$  of  $A'$  as follows.

$$w'(xy) = \begin{cases} w(xy) & \text{if } xy \in A_1 \\ \sum_{\{xz \in A \mid z \in I\}} w(xz) & \text{if } xy \in A_2 \\ \sum_{\{zy \in A \mid z \in I\}} w(zy) & \text{if } xy \in A_3 \end{cases}$$

We refer to Figure 2 for an illustration.

Next we generalize the notions of certificate and certification (Definitions 2 and 3) to backward weighted tournaments.

Figure 2: Illustration of the contraction step for the interval  $I := [v_i, \dots, v_j]$ .

**DEFINITION 10.** Let  $T_\omega = (V_\sigma, A)$  be a backward weighted tournament, and let  $f = vu \in A$  be a backward arc of  $T_\omega$ . We call  $\omega$ -certificate of  $f$ , and denote it by  $\mathcal{C}(f)$ , a collection of  $\omega(f)$  arc-disjoint directed paths going from  $u$  to  $v$  and using only forward arcs in the span of  $f$  in  $T_\omega$ .

**DEFINITION 11.** Let  $T_\omega = (V_\sigma, A)$  be a backward weighted tournament, and let  $F \subseteq A$  be a subset of backward arcs of  $T_\omega$ . We say that we can  $\omega$ -certify  $F$  whenever it is possible to find a set  $\mathcal{F} = \{\mathcal{C}(f) : f \in F\}$  of arc-disjoint  $\omega$ -certificates for the arcs in  $F$ .

**LEMMA 12.** Let  $T_\omega = (V_\sigma, A)$  be a backward weighted tournament such that for every interval  $I := [v_i, \dots, v_j]$  the following holds:

$$2 \cdot \omega(I) \leq |I| - 1 \quad (1)$$

Then it is possible to  $\omega$ -certify the backward arcs of  $T_\omega$ .

**PROOF.** Let  $V_\sigma = v_1, \dots, v_n$ . The proof is by induction on  $n$ , the number of vertices. Note that by applying (1) to the interval  $I = [v_1, \dots, v_n]$ , we have that there exists a vertex  $v_i$  in  $T_\omega$  that is not incident to any backward arc. Let  $T'_\omega = (V'_\sigma, A')$  denote the tournament  $T_\omega \setminus \{v_i\}$ . We say that an interval  $I$  is *critical* whenever  $2 \cdot \omega(I) = |I| - 1$ . We now consider several cases, based on different types of critical intervals.

- (i) Suppose that there are no critical intervals. Thus, in  $T'_\omega$ , every interval satisfies (1), and hence by induction on  $n$  the result holds.
- (ii) Suppose now that the only critical interval is  $I = [v_1, \dots, v_n]$ , and let  $e = vu$  be a backward arc above  $v_i$  with the maximum length. Note that since  $v_i$  does not belong to any backward arc, we can use it to form a directed path  $c(e) = uv_i v$ , which is a certificate for  $e$ . We now consider  $T'_\omega$  where the weight of  $e$  has been decreased by 1. In this process if  $\omega(e)$  becomes 0 then we reverse the arc  $e$ . We now show that every interval of  $T'_\omega$  respects (1). If an interval  $I' \in T'_\omega$  does not contain  $v_i$  in the corresponding interval in  $T_\omega$ , then by our assumption we have that  $2 \cdot \omega(I') \leq |I'| - 1$ . Now we assume that the interval corresponding to  $I'$  in  $T_\omega$  contains  $v_i$  but either  $u \notin I' \cup \{v_i\}$  or  $v \notin I' \cup \{v_i\}$ . Then we have  $2 \cdot \omega(I') = 2 \cdot \omega(I) < |I| - 1 = |I'|$  and hence we get that  $2 \cdot \omega(I') \leq |I'| - 1$ . Finally, we assume that the interval corresponding to  $I'$  in  $T_\omega$  contains  $v_i$  and  $u, v \in I' \cup \{v_i\}$ . In this case,  $2 \cdot \omega(I') = 2 \cdot (\omega(I) - 1) \leq |I| - 1 - 2 < |I'| - 1$ . Thus, by the induction hypothesis, we obtain a family of arc-disjoint  $\omega$ -certificates  $\mathcal{F}'$  which  $\omega$ -certify the backward arcs of  $T'_\omega$ . Observe that the maximality of  $I(e)$  ensures that if  $e$  is reversed then it will not be used in any  $\omega$ -certificate of  $\mathcal{F}'$ , thus implying that  $\mathcal{F}' \cup c(e)$  is a family  $\omega$ -certifying the backward arcs of  $T_\omega$ .



## 8 KERNELS FOR FEEDBACK ARC SET IN TOURNAMENTS

- (iii) Finally, suppose that there exists a critical interval  $I \subset V_\sigma$ . Roughly speaking, we will show that  $I$  and  $V_\sigma \setminus I$  can be certified separately. To do so, we first show the following claim.

*Claim.* Let  $I \subset V_\sigma$  be a critical interval. Then the tournament  $T_{\omega'} = (V_{\sigma'}, A')$  obtained from  $T_\omega$  by contracting  $I$  satisfies the conditions of the lemma.

**PROOF.** Let  $H'$  be any interval of  $T_{\omega'}$ . As before if  $H'$  does not contain  $c_I$  then the result holds by hypothesis. Otherwise, let  $H$  be the interval corresponding to  $H'$  in  $T_\omega$ . We will show that  $2\omega(H') \leq |H'| - 1$ . By hypothesis, we know that  $2\omega(H) \leq |H| - 1$  and that  $2\omega(I) = |I| - 1$ . Thus we have the following.

$$2\omega(H') = 2 \cdot (\omega(H) - \omega(I)) \leq |H| - 1 - |I| + 1 = (|H| + 1 - |I|) - 1 = |H'| - 1$$

Thus, we have shown that the tournament  $T_{\omega'}$  satisfies the conditions of the lemma. ■

We now consider a minimal critical interval  $I$ . By induction, and using the claim, we know that we can obtain a family of arc-disjoint  $\omega$ -certificates  $\mathcal{F}'$  which  $\omega$ -certifies the backward arcs of  $T_{\omega'}$  without using any arc within  $I$ . Now, by minimality of  $I$ , we can use (ii) to obtain a family of arc-disjoint  $\omega$ -certificates  $\mathcal{F}''$  which  $\omega$ -certifies the backward arcs of  $I$  using only arcs within  $I$ . Thus,  $\mathcal{F}' \cup \mathcal{F}''$  is a family  $\omega$ -certifying all backward arcs of  $T_\omega$ .

This concludes the proof of the lemma. ■

In the following, any interval that does not respect condition (1) is said to be a *dense interval*.

**LEMMA 13.** Let  $T_\omega = (V_\sigma, A)$  be a backward weighted tournament with  $|V_\sigma| \geq 2p + 1$ , having at most  $p$  backward arcs and every backward arc has weight one. Then there exists a safe partition of  $V_\sigma$  with at least one backward arc between the intervals and it can be computed in polynomial time.

**PROOF.** The proof is by induction on  $n = |V_\sigma|$ . Observe that the statement is true for  $n = 3$ , which is our base case.

For the inductive step, we assume first that there is no dense interval in  $T_\omega$ . In this case Lemma 12 ensures that the partition of  $V_\sigma$  into singletons of vertices is a safe partition. So from now on we assume that there exists at least one dense interval.

Let  $I$  be a dense interval. By definition of  $I$ , we have that  $\omega(I) \geq \frac{1}{2} \cdot |I|$ . We now contract  $I$  and obtain the backward weighted tournament  $T_{\omega'} = (V_{\sigma'}, A')$ . In the contracted tournament  $T_{\omega'}$ , we have:

$$\begin{cases} |V_{\sigma'}| & \geq 2p + 1 - (|I| - 1) = 2p - |I| + 2; \\ \omega'(V_{\sigma'}) & \leq p - \frac{1}{2} \cdot |I|. \end{cases}$$

Thus, if we set  $r := p - \frac{1}{2} \cdot |I|$ , we get that  $|V_{\sigma'}| \geq 2r + 1$  and  $\omega'(V_{\sigma'}) \leq r$ . Since  $|V_{\sigma'}| < |V_\sigma|$ , by the induction hypothesis we can find a safe partition  $\mathcal{P}$  of  $T_{\omega'}$ , and thus obtain a family  $\mathcal{F}_{\omega'}$  that  $\omega$ -certifies the backward arcs of  $T_{\omega'}[A_B]$  using only arcs in  $A_B$ .

We claim that  $\mathcal{P}'$  obtained from  $\mathcal{P}$  by substituting  $c_I$  by its corresponding interval  $I$  is a safe partition in  $T_\omega$ . To see this, first observe that if  $c_I$  has not been used to  $\omega$ -certify the

backward arcs in  $T_{\omega'}[A_B]$ , that is,  $c_I$  is not an end point of any arc in the  $\omega$ -certificates, then we are done. So from now on we assume that  $c_I$  has been part of a  $\omega$ -certificate for some backward arc. Let  $e$  be a backward arc in  $T_{\omega'}[A_B]$ , and let  $c_{\omega'}(e) \in \mathcal{F}_{\omega'}$  be a  $\omega$ -certificate of  $e$ . First we assume that  $c_I$  is not the first vertex of the certificate  $c_{\omega'}(e)$  (with respect to ordering  $\sigma'$ ), and let  $c_1$  and  $c_2$  be the left (in-) and right (out-) neighbors of  $c_I$  in  $c_{\omega'}(e)$ . By definition of the contraction step together with the fact that there is a forward arc between  $c_1$  and  $c_I$  and between  $c_I$  and  $c_2$  in  $T_{\omega'}$ , we have that there were no backward arcs between any vertex in the interval corresponding to  $c_I$  and  $c_1$  and  $c_2$  in the original tournament  $T_\omega$ . So we can always find a vertex in  $I$  to replace  $c_I$  in  $c_{\omega'}(e)$ , thus obtaining a certificate  $c(e)$  for  $e$  in  $T_\omega[A_B]$  (observe that  $e$  remains a backward arc even in  $T_\omega$ ). Now we assume that  $c_I$  is either a first or last vertex in the certificate  $c_{\omega'}(e)$ . Let  $e'$  be an arc corresponding to  $e$  in  $T_{\omega'}$  with one of its endpoints being  $e_I \in I$ . To certify  $e'$  in  $T_\omega[A_B]$ , we need to show that we can construct a certificate  $c(e')$  using only arcs of  $T_\omega[A_B]$ . We have two cases to deal with.

- (i) If  $c_I$  is the first vertex of  $c_{\omega'}(e)$  then let  $c_1$  be its right neighbor in  $c_{\omega'}(e)$ . Using the same argument as before, there are only forward arcs between any vertex in  $I$  and  $c_1$ . In particular, there is a forward arc  $e_I c_1$  in  $T_\omega$ , meaning that we can construct a  $\omega$ -certificate for  $e'$  in  $T_\omega$  by setting  $c(e') := (c_{\omega'}(e) \setminus \{c_I\}) \cup \{e_I\}$ .

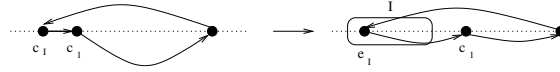


Figure 3: On the left, the  $\omega$ -certificate  $c_{\omega'}(e) \in \mathcal{F}_{\omega'}$ . On the right, the corresponding  $\omega$ -certificate obtained in  $T_\omega$  by replacing  $c_I$  by the interval  $I$ .

- (ii) If  $c_I$  is the last vertex of  $c_{\omega'}(e)$  then let  $c_q$  be its left neighbor in  $c_{\omega'}(e)$ . Once again, we have that there are only forward arcs between  $c_q$  and vertices in  $I$ , and thus between  $c_q$  and  $e_I$ . So using this we can construct a  $\omega$ -certificate for  $e'$  in  $T_\omega$ .

Notice that the fact that all  $\omega$ -certificates are pairwise arc-disjoint in  $T_{\omega'}[A_B]$  implies that the corresponding  $\omega$ -certificates are arc-disjoint in  $T_\omega[A_B]$ , and so  $\mathcal{P}'$  is indeed a safe partition of  $V_\sigma$ . ■

We are now ready to give the linear size kernel for  $k$ -FAST. To do so, we make use of the fact that there exists a polynomial time approximation scheme for this problem [20].

**THEOREM 14.** *For every fixed  $\epsilon > 0$ , there exists a vertex kernel for  $k$ -FAST with at most  $(2 + \epsilon)k$  vertices that can be computed in polynomial time.*

**PROOF.** Let  $(T = (V, A), k)$  be an instance of  $k$ -FAST. For a fixed  $\epsilon > 0$ , we start by computing a feedback arc set  $S$  of size at most  $(1 + \frac{\epsilon}{2})k$ . To find such a set  $S$ , we use the known polynomial time approximation scheme for  $k$ -FAST [20]. Then, we order  $T$  with the transitive ordering of the tournament obtained by reversing every arc of  $S$  in  $T$ . Let  $T_\sigma$  denote the resulting ordered tournament. By the upper bound on the size of  $S$ , we know that  $T_\sigma$  has at most  $(1 + \frac{\epsilon}{2})k$  backward arcs. Thus, if  $T_\sigma$  has more than  $(2 + \epsilon)k$  vertices then Lemma 13 ensures that we can find a safe partition with at least one backward arc between the intervals in polynomial time. Hence we can reduce the tournament by applying Rule 3.3. We then apply Rule 3.1, and repeat the previous steps until we do not find a safe partition or

## 10 KERNELS FOR FEEDBACK ARC SET IN TOURNAMENTS

$k = 0$ . In the former case, we know by Lemma 13 that  $T$  can have at most  $(2 + \epsilon)k$  vertices, thus implying the result. In all other cases we return NO. This concludes the proof of our main theorem. ■

## 4 Conclusion

In this paper we obtained linear vertex kernel for  $k$ -FAST, in fact, a vertex kernel of size  $(2 + \epsilon)k$  for any fixed  $\epsilon > 0$ . The new bound on the kernel size improves the previous known bound of  $O(k^2)$  on the vertex kernel size for  $k$ -FAST given in [3, 12]. It would be interesting to see if one can obtain kernels for other problems using either polynomial time approximation schemes or a constant factor approximation algorithm for the corresponding problem. An interesting problem which remains unanswered is, whether there exists a linear or even a  $o(k^2)$  vertex kernel for the  $k$ -FEEDBACK VERTEX SET IN TOURNAMENTS ( $k$ -FVST) problem. In the  $k$ -FVST problem we are given a tournament  $T$  and a positive integer  $k$  and the aim is to find a set of at most  $k$  vertices whose deletion makes the input tournament acyclic. The smallest known kernel for  $k$ -FVST has size  $O(k^2)$ .

## References

- [1] N. Ailon, M. Charikar, and A. Newman. Aggregating inconsistent information: ranking and clustering. In *ACM Symposium on Theory of Computing (STOC)*, pages 684–693, 2005.
- [2] N. Alon. Ranking tournaments. *SIAM J. Discrete Math.*, 20(1):137–142, 2006.
- [3] N. Alon, D. Lokshtanov, and S. Saurabh. Fast FAST. In *International Colloquium on Automata, Languages and Programming (ICALP)*, 2009. (to appear).
- [4] H. L. Bodlaender, R. G. Downey, M. R. Fellows, and D. Hermelin. On problems without polynomial kernels (extended abstract). In *International Colloquium on Automata, Languages and Programming (ICALP)*, volume 5125 of *LNCS*, pages 563–574, 2008.
- [5] H. L. Bodlaender, F. V. Fomin, D. Lokshtanov, E. Penninkx, S. Saurabh, and D. M. Thi-likos. (meta) kernelization. *CoRR*, abs/0904.0727, 2009.
- [6] J. Borda. Mémoire sur les élections au scrutin. *Histoire de l’Académie Royale des Sciences*, 1781.
- [7] N. Bousquet, J. Daligault, S. Thomassé, and A. Yeo. A polynomial kernel for multicut in trees. In *Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 183–194, 2009.
- [8] P. Charbit, S. Thomassé, and A. Yeo. The minimum feedback arc set problem is NP-hard for tournaments. *Combin. Probab. Comput.*, 16(1):1–4, 2007.
- [9] W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. In *Advances in neural information processing systems (NIPS)*, pages 451–457, 1997.
- [10] M. Condorcet. Essai sur l’application de l’analyse à la probabilité des décisions rendues à la pluralité des voix, 1785.
- [11] D. Coppersmith, L. Fleischer, and A. Rudra. Ordering by weighted number of wins gives a good ranking for weighted tournaments. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 776–782, 2006.

- [12] M. Dom, J. Guo, F. Hüffner, R. Niedermeier, and A. Truß. Fixed-parameter tractability results for feedback set problems in tournaments. In *Conference on Algorithms and Complexity (CIAC)*, volume 3998 of *LNCS*, pages 320–331, 2006.
- [13] M. Dom, D. Lokshtanov, and S. Saurabh. Incompressibility through colors and IDs. In *International Colloquium on Automata, Languages and Programming (ICALP)*, 2009. (to appear).
- [14] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation revisited. Technical report.
- [15] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *World Wide Web Conference (WWW)*, 2001.
- [16] P. Erdős and J. W. Moon. On sets on consistent arcs in tournaments. *Canad. Math. Bull.*, 8:269–271, 1965.
- [17] H. A. Jung. On subgraphs without cycles in tournaments. *Combinatorial Theory and its Applications II*, pages 675–677, 1970.
- [18] J. Kemeny. Mathematics without numbers. *Daedalus*, 88:571–591, 1959.
- [19] J. Kemeny and J. Snell. *Mathematical models in the social sciences*. Blaisdell, 1962.
- [20] C. Kenyon-Mathieu and W. Schudy. How to rank with few errors. In *ACM Symposium on Theory of Computing (STOC)*, pages 95–103, 2007.
- [21] R. M. McConnell and F. de Montgolfier. Linear-time modular decomposition of directed graphs. *Discrete Appl. Math.*, 145(2):198–209, 2005.
- [22] V. Raman and S. Saurabh. Parameterized algorithms for feedback set problems and their duals in tournaments. *Theor. Comput. Sci.*, 351(3):446–458, 2006.
- [23] K. D. Reid and E. T. Parker. Disproof of a conjecture of Erdős and Moser on tournaments. *J. Combin. Theory*, 9:225–238, 1970.
- [24] S. Seshu and M. B. Reed. *Linear Graphs and Electrical Networks*. Addison-Wesley, 1961.
- [25] P. Slater. Inconsistencies in a schedule of paired comparisons. *Biometrika*, 48:303–312, 1961.
- [26] J. Spencer. Optimal ranking of tournaments. *Networks*, 1:135–138, 1971.
- [27] S. Thomassé. A quadratic kernel for feedback vertex set. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 115–119, 2009.
- [28] A. van Zuylen. Deterministic approximation algorithms for ranking and clusterings. Technical Report 1431, Cornell ORIE, 2005.
- [29] A. van Zuylen, R. Hegde, K. Jain, and D. P. Williamson. Deterministic pivoting algorithms for constrained ranking and clustering problems. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 405–414, 2007.
- [30] D. H. Younger. Minimum feedback arc sets for a directed graph. *IEEE Trans. Circuit Theory*, 10:238–245, 1963.

## 2.10 Counting the number of edge-colourings

### Enumerating the edge-colourings and total colourings of a regular graph\*

S. Bessy<sup>†</sup> and F. Havet<sup>‡</sup>

November 5, 2011

#### Abstract

In this paper, we are interested in computing the number of edge colourings and total colourings of a connected graph. We prove that the maximum number of  $k$ -edge-colourings of a connected  $k$ -regular graph on  $n$  vertices is  $k \cdot ((k-1)!)^{n/2}$ . Our proof is constructive and leads to a branching algorithm enumerating all the  $k$ -edge-colourings of a connected  $k$ -regular graph in time  $O^*((k-1)!)^{n/2}$  and polynomial space. In particular, we obtain a algorithm to enumerate all the 3-edge-colourings of a connected cubic graph in time  $O^*(2^{n/2}) = O^*(1.4143^n)$  and polynomial space. This improves the running time of  $O^*(1.5423^n)$  of the algorithm due to Golovach et al. [10]. We also show that the number of 4-total-colourings of a connected cubic graph is at most  $3 \cdot 2^{3n/2}$ . Again, our proof yields a branching algorithm to enumerate all the 4-total-colourings of a connected cubic graph.

## 1 Introduction

We refer to [5] for standard notation and concepts for graphs. In this paper, all the considered graphs are loopless, but may have parallel edges. A graph with no parallel edges is said to be *simple*. Let  $G$  be a graph. We denote by  $n(G)$  the number of vertices of  $G$ , and for each integer  $k$ , we denote by  $n_k(G)$  the number of degree  $k$  vertices of  $G$ . Often, when the graph  $G$  is clearly understood, we abbreviate  $n(G)$  to  $n$  and  $n_k(G)$  to  $n_k$ .

Graph colouring is one of the classical subjects in graph theory. See for example the book of Jensen and Toft [12]. From an algorithmic point of view, for many colouring type problems, like vertex colouring, edge colouring and total colouring, the existence problem asking whether an input graph has a colouring with an input number of colours is NP-complete. Even more, these colouring problems remain NP-complete when the question is whether there is a colouring of the input graph with a fixed (and greater than 2) number of colours [9, 11, 17].

Exact algorithms to solve NP-hard problems are a challenging research subject in graph algorithms. Many papers on exact exponential time algorithms have been published in the last decade. One of the major results is the  $O^*(2^n)$ -time inclusion-exclusion algorithm to compute the chromatic number of a graph found independently by Björklund, Husfeldt [2] and Koivisto [14], see [4]. This

\*This work is supported by the French *Agence Nationale de la Recherche* under reference AGAPE ANR-09-BLAN-0159.

<sup>†</sup>Université Montpellier 2 - CNRS, LIRMM. e-mail: Stephane.Bessy@lirmm.fr.

<sup>‡</sup>Projet Mascotte, I3S (CNRS, UNSA) and INRIA, Sophia Antipolis.  
email: Frederic.Havet@inria.fr.

approach may also be used to establish a  $O^*(2^n)$ -time algorithm to count the  $k$ -colourings and to compute the chromatic polynomial of a graph. It also implies a  $O^*(2^m)$ -time algorithm to count the  $k$ -edge-colourings and a  $O^*(2^{n+m})$ -time algorithm to count the  $k$ -total-colourings of a given graph.

Since edge colouring and total colouring are particular cases of vertex colouring, a natural question is to ask if faster algorithms than the general one may be designed in these cases. For instance, very recently Björklund et al. [3] showed how to detect whether a  $k$ -regular graph admits a  $k$ -edge-colouring in time  $O^*(2^{(k-1)n/2})$ .

The existence problem asking whether a graph has a colouring with a fixed and small number  $k$  of colours also attracted a lot of attention. For vertex colourability the fastest algorithm for  $k = 3$  has running time  $O^*(1.3289^n)$  and was proposed by Beigel and Eppstein [1], and the fastest algorithm for  $k = 4$  has running time  $O^*(1.7272^n)$  and was given by Fomin et al. [8]. They also established algorithms for counting  $k$ -vertex-colourings for  $k = 3$  and 4. The existence problem for a 3-edge-colouring is considered in [1, 15, 10]. Kowalik [15] gave an algorithm deciding if a graph is 3-edge-colourable in time  $O^*(1.344^n)$  and polynomial space and Golovach et al. [10] presented an algorithm counting the number of 3-edge-colourings of a graph in time  $O^*(3^{n/6}) = O^*(1.201^n)$  and exponential space. Golovach et al. [10] also showed a branching algorithm to enumerate all the 3-edge-colourings of a connected cubic graph in time  $O^*(25^{n/8}) = O^*(1.5423^n)$  and polynomial space. In particular, this implies that every connected cubic graph of order  $n$  has at most  $O(1.5423^n)$  3-edge-colourings. They give an example of a connected cubic graph of order  $n$  having  $\Omega(1.2820^n)$  3-edge-colourings. In Section 2, we prove that a connected cubic graph of order  $n$  has at most  $3 \cdot 2^{n/2}$  3-edge-colourings and give an example reaching this bound. Our proof can be translated into a branching algorithm to enumerate all the 3-edge-colourings of a connected cubic graph in time  $O^*(2^{n/2}) = O^*(1.4143^n)$  and polynomial space. Furthermore, we extend our result proving that every  $k$ -regular connected graph of order  $n$  admits at most  $k \cdot ((k-1)!)^{n/2}$   $k$ -edge-colourings. And, similarly, we derive a branching algorithm to enumerate all the  $k$ -edge-colourings of a connected  $k$ -regular graph in time  $O^*(k \cdot ((k-1)!)^{n/2})$  and polynomial space.

Regarding total colouring, very little has been done. Golovach et al. [10] showed a branching algorithm to enumerate the 4-total-colourings of a connected cubic graph in time  $O^*(2^{13n/8}) = O^*(3.0845^n)$ , implying that the maximum number of 4-total-colourings in a connected cubic graph of order  $n$  is at most  $O^*(2^{13n/8}) = O^*(3.0845^n)$ . In Section 3, we lower this bound to  $3 \cdot 2^{3n/2} = O(2.8285^n)$ . Again, our proof yields a branching algorithm to enumerate all the 4-total-colourings of a connected cubic graph in time  $O^*(2.8285^n)$  and polynomial space.

## 2 Edge colouring

A (*proper*) *edge colouring* of a graph is a colouring of its edges such that two adjacent edges receive different colours. An edge colouring with  $k$  colours is a *k-edge-colouring*. We denote by  $c_k(G)$  the number of  $k$ -edge-colourings of a graph  $G$ .

### 2.1 General bounds for $k$ -regular graphs

In this section, we are interested in computing the number of  $k$ -edge-colourings of  $k$ -regular connected graphs. We start by computing exactly the number of 3-edge-colourings of the cycles.

**Proposition 1.** *Let  $C_n$  be the cycle of length  $n$ .*

$$c_3(C_n) = \begin{cases} 2^n + 2, & \text{if } n \text{ is even,} \\ 2^n - 2, & \text{if } n \text{ is odd.} \end{cases}$$

*Proof.* By induction on  $n$ . It is easy to check that  $c_3(C_2) = c_3(C_3) = 6$ .

Let  $C_n = (v_1, v_2, \dots, v_n, v_1)$ . Let  $A$  be the set of 3-edge-colourings of  $C_n$  such that  $c(v_{n-1}v_n) \neq c(v_1v_2)$  and  $B$  the set of 3-edge-colourings of  $C_n$  such that  $c(v_{n-1}v_n) = c(v_1v_2)$ . The 3-edge-colourings of  $A$  are in one-to-one correspondence with those of  $C_{n-1}$  and the pair of colourings of  $B$  agreeing everywhere except on  $v_nv_1$  are in one-to-one correspondence with the 3-edge-colourings of  $C_{n-2}$ . Thus  $c_3(C_n) = c_3(C_{n-1}) + 2c_3(C_{n-2})$ . Hence, if  $n$  is even, then  $c_3(C_n) = 2^{n-1} - 2 + 2(2^{n-2} + 2) = 2^n + 2$ , and if  $n$  is odd, then  $c_3(C_n) = 2^{n-1} + 2 + 2(2^{n-2} - 2) = 2^n - 2$ .  $\square$

Let us now present our method which is based on a classical tool:  $(s, t)$ -ordering.

**Definition 2.** Let  $G$  be a graph and  $s$  and  $t$  be two distinct vertices of  $G$ . An  $(s, t)$ -ordering of  $G$  is an ordering of its vertices  $v_1, \dots, v_n$  such that  $s = v_1$  and  $t = v_n$ , and for all  $1 < i < n$ ,  $v_i$  has a neighbour in  $\{v_1, \dots, v_{i-1}\}$  and a neighbour in  $\{v_{i+1}, \dots, v_n\}$ .

**Lemma 3** (Lempel et al. [16]). *A graph  $G$  is a 2-connected graph if, and only if, for every pair  $(s, t)$  of vertices, it admits an  $(s, t)$ -ordering.*

In fact, Lempel et al. established Lemma 3 only for simple graphs but it can be trivially extended to graphs since replacing all the parallel edges between two vertices by a unique edge does not change the connectivity.

**Theorem 4.** *Let  $G$  be a 2-connected subcubic graph. Then  $c_3(G) \leq 3 \cdot 2^{n - \frac{n_3}{2}}$ .*

*Proof.* If  $G$  is a cycle, then the result follows from Proposition 1. Hence we may assume that  $G$  is not a cycle and thus has at least two vertices of degree 3, say  $s$  and  $t$ . By Lemma 3, there exists an  $(s, t)$ -ordering  $v_1, v_2, \dots, v_n$  of  $G$ . Orient each edge of  $G$  according to this order, that is from the lower-indexed end-vertex towards its higher-indexed one. Let us denote by  $D$  the obtained digraph. Observe that  $d^+(v_1) = 3 = d^-(v_n)$  and  $d^-(v_1) = 0 = d^+(v_n)$ . Let  $A^+$  (resp.  $A^-$ ) be the set of vertices with outdegree 2 (resp. indegree 2) in  $D$  and  $A_2$  be the set of vertices with degree 2 in  $G$  (and thus with indegree 1 and outdegree 1 in  $D$ ). Clearly,  $(A_2, A^-, A^+)$  is a partition of  $V(D) \setminus \{v_1, v_n\}$ . Observe that  $|A_2| = n - n_3$ . Since  $\sum_{v \in V(D)} d^+(v) = \sum_{v \in V(D)} d^-(v)$ , we have  $|A^+| = |A^-|$ , and so  $|A^+| = (n_3 - 2)/2$ .

Now for  $i = 1$  to  $n - 1$ , we enumerate the  $p_i$  partial 3-edge-colourings of the arcs whose tail is in  $\{v_1, \dots, v_i\}$ . For  $i = 1$ , there are 6 such colourings, since  $d^+(v_1) = 3$ .

Now, for each  $i$ , when we want to extend the partial colourings, two cases may arise.

- If  $d_D^-(v_i) = 1$ , then we need to colour one or two arcs, and one colour (the one of the arc entering  $v_i$ ) is forbidden, so there are at most 2 possibilities. Hence  $p_i \leq 2p_{i-1}$ .
- If  $d_D^-(v_i) = 2$ , then we need to colour one arc, and at least two colours (the ones of the arcs entering  $v_i$ ) are forbidden, so there is at most one possibility. Hence  $p_i \leq p_{i-1}$ .

At the end, all the edges of  $G$  are coloured, and a simple induction shows that  $c_3(G) = p_{n-1} \leq 6 \cdot 2^{|A_2| + |A^+|} = 3 \cdot 2^{n - \frac{n_3}{2}}$ .  $\square$

In particular, for a connected cubic graph  $G$ , we obtain  $c_3(G) \leq 3 \cdot 2^{n/2}$ . We now extend this result to  $k$ -regular graphs.

**Theorem 5.** *Let  $G$  be a connected  $k$ -regular graph, with  $k \geq 3$ . Then  $c_k(G) \leq k \cdot ((k-1)!)^{n/2}$ .*

*Proof.* First, remark that if a connected  $k$ -regular graph  $G$  admits a  $k$ -edge-colouring, then every colour induces a perfect matching of  $G$ , and then  $n$  is even. Furthermore, observe that  $G$  is 2-connected. Indeed, assume that  $G$  has a cutvertex  $x$  and admits a  $k$ -edge-colouring  $c$ . As  $G$  has an even number of vertices, one of the connected components, say  $C$ , of  $G - x$  has odd cardinality. A colour appearing on an edge between  $x$  and a connected component of  $G - x$  different from  $C$  must form a perfect matching on  $C$  which is impossible. So,  $G$  is 2-connected.

Hence we can use the method of the proof of Theorem 4 and consider an  $(s, t)$ -ordering  $v_1, \dots, v_n$  of  $G$  and  $D$  the orientation of  $G$  obtained from this ordering (i.e.  $v_i v_j \in A(D)$  if and only if  $v_i v_j \in E(G)$  and  $i < j$ ). The analysis made in the proof of Theorem 4 yields  $c_k(G) \leq \prod_{x \in V(G)} (d^+(x)!)^i$ . For  $i = 1, \dots, k-1$ , we define  $A_i = \{x \in V(G) \setminus \{v_1, v_n\} : d^+(x) = i\}$ . It is clear that  $(A_i)_{1 \leq i \leq k-1}$  form a partition of  $V(G) \setminus \{v_1, v_n\}$ . If we denote  $|A_i|$  by  $a_i$ , then  $c_k(G) \leq P := k! \prod_{i=1}^{k-1} (i!)^{a_i}$ . Moreover  $S_1 := \sum_{i=1}^{k-1} a_i = n-2$  (by counting the number of vertices of  $G$ ) and  $S_2 := \sum_{i=1}^{k-1} i \cdot a_i = k(n-2)/2$  (by counting the number of arcs of  $D - v_1$ ).

Let us now find the maximum value of  $P$  under the conditions  $S_1 = n-2$  and  $S_2 = k(n-2)/2$ . If we can find  $1 < p \leq q < k-1$  with  $a_p \neq 0$  and  $a_q \neq 0$  (or  $a_p \geq 2$  if  $p = q$ ), then we decrease  $a_p$  and  $a_q$  by one and increase  $a_{p-1}$  and  $a_{q+1}$  by one. Doing this,  $S_1$  and  $S_2$  are unchanged and  $P$  is multiplied by  $\frac{q+1}{p} > 1$ . We repeat this operation as many times as possible and stop when (a) for every  $i = 2, \dots, k-2$ ,  $a_i = 0$  or (b) there exists  $j \in \{2, \dots, k-2\}$  such that for every  $i = 2, \dots, k-2$  and  $i \neq j$ ,  $a_i = 0$  and  $a_j = 1$ . In case (b),  $S_1$  gives  $a_1 + 1 + a_{k-1} = n-2$  and  $S_2$  is  $a_1 + j + (k-1)a_{k-1} = k(n-2)/2$ . Combining  $S_1$  and  $S_2$ , we obtain  $2(k-2)a_{k-1} + 2(j-1) = (k-2)(n-2)$  and we conclude that  $k-2$  divides  $2(j-1)$  and so that  $j = k/2$ . Solving  $S_1$  and  $S_2$  we have in particular that  $a_1 = a_{k-1}$  and  $2a_1 = n-3$  which is impossible, as  $n$  is even. Hence, we are in case (a), and solving  $S_1$  and  $S_2$  yields to  $a_1 = a_k = (n-2)/2$ . We conclude that  $P \leq k!((k-1)!)^{n/2-1}$ .  $\square$

We turn now the proof of Theorem 5 into an algorithm to enumerate all the  $k$ -edge-colourings of a connected  $k$ -regular graph.

**Corollary 6.** *There is an algorithm to enumerate all the  $k$ -edge-colourings of a connected  $k$ -regular graph on  $n$  vertices in time  $O^*((k-1)!)^{n/2}$  and polynomial space.*

*Proof.* Let  $G$  be a connected  $k$ -regular graph. We first check the 2-connectivity of  $G$ . If it is not 2-connected, then we return ‘The graph is not  $k$ -edge-colourable’.

If it is 2-connected, then we proceed as follows. We compute an  $(s, t)$ -ordering  $v_1, \dots, v_n$  of  $G$ , which can be done in polynomial time (see [6] and [7] for instance), and orient  $G$  accordingly to this ordering. Now, it is classical to enumerate all the permutations of a set of size  $p$  in time  $O(p!)$  and linear space, in such way that, being given a permutation we compute in average constant time the next permutation in the enumeration (with the Steinhaus-Johnson-Trotter algorithm for instance, see [13]).

Using this and the odometer principle, it is now easy to enumerate all the edge colourings we want. In the enumeration of all the permutations of  $\{1, \dots, k\}$ , we take the first one and assign the corresponding colours to the arc with tail  $x_1$ . For any index  $i$  with  $2 \leq i \leq k$ , we assign to the arcs



with tail  $x_i$  the first permutation in the enumeration of the permutations of the possible colours for these arcs (i.e. all the colours of  $\{1, \dots, k\}$  minus the one of the arcs entering in  $x_i$ ). Then, we have the first colouring, and we check if it is a proper edge colouring of  $G$  (in polynomial time). To obtain the next colouring, we take the next permutation on the colours possible on the arcs with tail  $x_{n-1}$ , and so on. Once all the possible permutations have been enumerated for these arcs, we take the next permutation on the colours possible on the arcs with tail  $x_{n-2}$  and re-enumerate the permutation of possible colours for the arcs with tail  $x_{n-1}$ , and so on, following the odometer principle.  $\square$

The bound given by Theorem 5 is optimal on the class of connected  $k$ -regular graphs. For all  $k \geq 3$ , and  $n \geq 2$ ,  $n = 2p$  even, the  $k$ -noodle necklace  $N_n^k$  is the  $k$ -regular graph obtained from a cycle on  $2p$  vertices  $(v_1, v_2, \dots, v_{2p}, v_1)$  by replacing all the edges  $v_{2i-1}v_{2i}$ ,  $1 \leq i \leq p$  by  $k-1$  parallel edges.

**Proposition 7.** *Let  $k \geq 3$  and  $n \geq 2$ ,*

$$c_k(N_n^k) = k \cdot ((k-1)!)^{n/2}.$$

*Proof.* Observe that in every  $k$ -edge-colouring of  $N_n^k$  the edges which are not multiplied (i.e.  $v_{2i}v_{2i+1}$ ) are coloured the same. There are  $k$  choices for such a colour. Once this colour is fixed, there are  $(k-1)!$  choices for each set of  $k-1$  parallel edges. Hence  $c_k(N_n^k) = k \cdot ((k-1)!)^{n/2}$ .  $\square$

## 2.2 A more precise bound for cubic graphs

For simple cubic graphs, we lower the bound on the number of 3-edge-colourings from  $3 \cdot 2^{n/2}$  to  $\frac{9}{4} \cdot 2^{n/2}$ .

**Lemma 8.** *If  $G$  is a connected cubic simple graph, then  $c_3(G) \leq \frac{9}{4} \cdot 2^{n/2}$ .*

*Proof.* As in Theorem 4, let us consider an  $(s, t)$ -ordering  $v_1, v_2, \dots, v_n$  of the vertices and the acyclic digraph  $D$  obtained by orienting all the edges of  $G$  according to this ordering.

Let  $i$  be the smallest integer such that  $d^-(v_i) = 2$ . Since every vertex (except  $v_1$ ) has an inneighbour, there exists  $j$  such that there are two internally-disjoint directed paths from  $v_j$  to  $v_i$  in  $D$ . In  $G$ , the union of these two paths forms a cycle  $C$ . By definition of  $i$ , all vertices of  $C$  but  $v_i$  have outdegree 2. So, if there is  $k$  such that  $j < k < i$  and  $v_k \notin V(C)$ , then  $v_k$  has no outneighbour in  $C$  and the ordering  $v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_i, v_k, v_{i+1}, \dots, v_n$  is also an  $(s, t)$ -ordering. Repeating this operation as many times as necessary, we may obtain that all the vertices of  $C$  are consecutive in the ordering, that is  $C = (v_j, v_{j+1}, \dots, v_i, v_j)$ .

We enumerate the 3-edge-colourings of  $G$  in a similar way to the proof of Theorem 4, except that instead of examining the colour of arcs with tail in  $\{v_j, \dots, v_i\}$  one after another, we look at  $C$  globally. If  $j = 1$ , then there are exactly  $c_3(C)$  3-edge-colourings of  $C$ , because no arcs has head in  $C$ . If  $j > 1$ , then there are  $c_3(C)/3$  3-edge-colourings of  $C$ , because one arc has head  $v_j$  and we need the colours of the two arcs with tail  $v_j$  to have a colour distinct from it.

Recall that in  $D$ ,  $v_1$  has indegree 0,  $v_p$  has indegree 3,  $\frac{n-2}{2}$  vertices have indegree 2 and  $\frac{n-2}{2}$  vertices have indegree 1. If  $j > 1$  (resp.  $j = 1$ ), then there are  $i-j$  (resp.  $i-2$ ) vertices of indegree 1 in  $C$ , so there are  $\frac{n-2i+2j-2}{2}$  (resp.  $\frac{n-2i+2}{2}$ ) vertices of indegree 1 in  $V(G) \setminus V(C)$ .

If  $j = 1$ , then we start by colouring  $C$  and then extend the colouring to  $G$ . Once  $C$  is coloured, there is at most one possibility to colour each arc with tail in  $C$ , so  $c_3(G) \leq c_3(C) \cdot 2^{\frac{n-2i+2}{2}} =$

$(c_3(C)/2^{i-1}) \cdot 2^{n/2}$ . If  $j > 1$ , we colour the arcs with tail in  $\{v_1, \dots, v_{j-1}\}$  as usual. Remark that, by the choice of  $C$ , there is exactly one of these arcs, denoted by  $e$ , which has head in  $C$  and more precisely  $e$  has tail  $v_j$ . Then, we consider all the 3-edge-colourings of  $C$  that agree with the colour of  $e$  (i.e.  $c(v_j v_{j+1}) \neq c(e)$  and  $c(v_j v_i) \neq c(e)$ ). There are exactly  $c_3(C)/3$  such colourings. Finally, we extend the edge colourings in all possible ways to  $G$  using the usual method. So, in this case, we obtain  $c_3(G) \leq 6 \cdot c_3(C)/3 \cdot 2^{\frac{n-2i+2j-2}{2}} = (c_3(C)/2^{i-j}) \cdot 2^{n/2}$ .

In all cases, we have to bound the value  $\frac{c_3(C)}{2^{i-j}}$  for  $1 \leq i < j$ . Since  $G$  has no 2-cycles,  $C$  has length at least 3, and so  $i - j \geq 2$ . By Proposition 1,  $c_3(G) = 2^{i-j+1} + 2$  if  $i - j$  is odd, and  $c_3(G) = 2^{i-j+1} - 2$  if  $i - j$  is even. Easily one sees that the value  $\frac{c_3(C)}{2^{i-j}}$  is maximized when  $i - j = 3$  (i.e.  $C$  has length four), and so  $\frac{c_3(C)}{2^{i-j}} \leq \frac{18}{8} = \frac{9}{4}$ . Thus  $c_3(G) \leq \frac{9}{4} \cdot 2^{n/2}$ .  $\square$

Theorem 5 for  $k = 3$  states that a connected cubic graph  $G$  has at most  $3 \cdot 2^{n/2}$  3-edge-colourings. We shall now describe all connected cubic graphs attaining this bound.

Let  $G$  be a cubic graph and  $C = uvu$  be a 2-cycle in  $G$ . Then  $G/C$  is the graph obtained from  $G - \{u, v\}$  by adding an edge between the neighbour of  $u$  distinct from  $v$  and the neighbour of  $v$  distinct from  $u$ .

Let  $\Theta$  be the graph with two vertices joined by three edges. And let  $\mathcal{L}$  be the family of graphs defined recursively as follows:

- $\Theta \in \mathcal{L}$ .
- if  $G$  has a 2-cycle  $C$  such that  $G/C$  is in  $\mathcal{L}$ , then  $G$  is in  $\mathcal{L}$ .

Remark that the 3-noodle necklaces  $N_n^3$  (with  $n$  even) belongs to the family  $\mathcal{L}$ .

**Theorem 9.** *Let  $G$  be a connected cubic graph. If  $G \in \mathcal{L}$ , then  $c_3(G) = 3 \cdot 2^{n/2}$ . Otherwise  $c_3(G) \leq \frac{9}{4} \cdot 2^{n/2}$ .*

*Proof.* By induction on  $n$ , the result holding for simple graphs by Lemma 8 and for  $\Theta$  because  $c_3(\Theta) = 6$ .

Assume that  $n \geq 4$  and that  $G$  has a 2-cycle  $C = uvu$ . In any 3-edge-colouring of  $G$ , the edges not in  $C$  incident to  $u$  and  $v$  are coloured the same. Hence to each 3-edge-colouring  $c$  of  $G/C$  corresponds the two 3-edge-colourings of  $G$  that agrees with  $c$  on  $G - \{u, v\}$ . Hence  $c_3(G) = 2c_3(G/C)$ .

If  $G/C$  is in  $\mathcal{L}$ , then  $G$  is also in  $\mathcal{L}$ . Moreover, by the induction hypothesis,  $c_3(G/C) = 3 \cdot 2^{(n-2)/2}$ . So  $c_3(G) = 3 \cdot 2^{n/2}$ .

If  $G/C$  is not in  $\mathcal{L}$ , then  $G$  is not in  $\mathcal{L}$ . Moreover, by the induction hypothesis,  $c_3(G/C) \leq \frac{9}{4} \cdot 2^{(n-2)/2}$ . So  $c_3(G) \leq \frac{9}{4} \cdot 2^{n/2}$ .  $\square$

We have no example of cubic simple graphs admitting exactly  $\frac{9}{4} \cdot 2^{n/2}$  3-edge-colourings, and we believe that  $\frac{9}{4}$  could be replaced by a lower constant in the statement of Theorem 9. In fact, we conjecture that the maximum number of 3-edge-colourings of cubic simple graphs of order  $n$  is attained by some special graphs that we now describe.

For all  $n \geq 2$ ,  $n = 2p$  even, the *hamster wheel*  $H_n$  is the cubic graph obtained from two cycles on  $p$  vertices  $C_v = (v_1, v_2, \dots, v_p, v_1)$  and  $C_w = (w_1, w_2, \dots, w_p, w_1)$  by adding the matching  $M = \{v_i w_i : 1 \leq i \leq p\}$ . This construction for a lower bound was proposed by Pyatkin as it is mentioned in [10].

**Proposition 10.**

$$c_3(H_n) = \begin{cases} 2^{n/2} + 8, & \text{if } n/2 \text{ is even,} \\ 2^{n/2} - 2, & \text{if } n/2 \text{ is odd.} \end{cases}$$

*Proof.* Let  $\phi$  is a 3-edge-colouring of  $C_v$ .

If the three colours appear on  $C_v$ , then there is a unique 3-edge-colouring of  $H_n$  extending  $\phi$ . Indeed, to extend  $\phi$ , the colours of the edges of  $M$  are forced. Since the three colours appear on  $C_v$ , there are two edges of  $M$  which are coloured differently. Without loss of generality, we may assume that these two edges are consecutive, that is there exists  $i$  such that they are  $v_j w_j$  and  $v_{j+1} w_{j+1}$ . But then the colour of  $w_j w_{j+1}$  must be equal to the one of  $v_j v_{j+1}$ . Then, from edge to edge along the cycle, one shows that for all  $i$ , the colour of  $w_i w_{i+1}$  is the one of  $v_i v_{i+1}$ .

If only two colours appear on  $C_v$ , then there are two 3-edge-colourings of  $H_n$  extending  $\phi$ . Indeed in this case,  $n$  is even and all the edges of  $M$  must be coloured by the colour not appearing on  $C_v$ . So, there are two possible 3-edge-colourings of  $C_w$  with the colours appearing on  $C_v$ .

Hence the number of 3-edge-colourings of  $G$  is equal to the number of 3-edge-colourings of  $C_v$  plus the number of 3-edge-colourings of  $C_v$  in which two colours appear. If  $n/2$  is odd, this last number is 0, and if  $n/2$  is even, this number is 6. So, by Proposition 1,  $c_3(H_n) = 2^{n/2} - 2$  if  $n/2$  is odd and  $c_3(H_n) = 2^{n/2} + 8$  if  $n/2$  is even.  $\square$

For all  $n \geq 2$ ,  $n = 2p$  even, the *Möbius ladder*  $M_n$  is the cubic graph obtained from a cycle on  $n$  vertices  $C = (v_1, v_2, \dots, v_n, v_1)$  by adding the matching  $M = \{v_i v_{i+p} : 1 \leq i \leq p\}$  (indices are modulo  $n$ ).

Two edges  $e$  and  $f$  of the cycle  $C$  are said to be *antipodal*, if there exists  $1 \leq i \leq p$  such that  $\{e, f\} = \{v_i v_{i+1}, v_{i+p} v_{i+p+1}\}$ . A 3-edge-colouring  $c$  of  $M_n$  is said to be *antipodal* if  $c(e) = c(f)$  for every pair  $(e, f)$  of antipodal edges.

**Proposition 11.** *Let  $c$  be a 3-edge-colouring of  $M_n$ . If  $c$  is not antipodal, then  $n/2$  is odd and all the arcs  $v_i v_{i+n/2+1}$  are coloured the same.*

*Proof.* Suppose that two antipodal edges are not coloured the same. Without loss of generality,  $c(v_n v_1) = 2$  and  $c(v_p v_{p+1}) = 3$ , where  $n = 2p$ . Hence, we have  $c(v_1 v_{p+1}) = 1$ ,  $c(v_1 v_2) = 3$  and  $c(v_{p+1} v_{p+2}) = 2$ . And so on by induction, for all  $1 \leq i \leq p$ ,  $c(v_i v_{i+p}) = 1$  and  $\{c(v_i v_{i+1}), c(v_{i+p} v_{i+p+1})\} = \{2, 3\}$ . Hence, the edges of  $C$  are coloured alternately with 2 and 3. Since  $c(v_{2p} v_1) = 2$  and  $c(v_p v_{p+1}) = 3$ ,  $p$  must be odd.  $\square$

**Proposition 12.**

$$c_3(M_n) = \begin{cases} 2^{n/2} + 2, & \text{if } n/2 \text{ is even,} \\ 2^{n/2} + 4, & \text{if } n/2 \text{ is odd.} \end{cases}$$

*Proof.* Clearly, there is a one-to-one mapping between the antipodal 3-edge-colourings of  $M_n$  and the 3-edge-colourings of  $C_{n/2}$ . Hence, by Proposition 11, if  $n/2$  is even, then  $c_3(M_n) = c_3(C_{n/2}) = 2^{n/2} + 2$  by Proposition 1.

If  $n/2$  is odd, non-antipodal 3-edge-colourings are those such that all arcs  $v_i v_{i+n/2+1}$  are coloured the same, by Proposition 11. There are 6 such edge colourings (three choices for the colour of the edges  $v_i v_{i+n/2+1}$  and for each of these choices, two possible edge colourings of  $C$ ). Hence  $c_3(M_n) = c_3(C_{n/2}) + 6 = 2^{n/2} + 4$  by Proposition 1.  $\square$

We think that  $H_n$  and  $M_n$  are the connected cubic graphs which admit the maximum number of 3-edge-colourings. Precisely, we raise the following conjecture.

**Conjecture 13.** Let  $G$  be a connected cubic simple graph on  $n$  vertices. If  $n/2$  is even, then  $c_3(G) \leq c_3(H_n)$  and if  $n/2$  is odd, then  $c_3(G) \leq c_3(M_n)$ .

### 3 Total colouring

A *total colouring* of a graph  $G$  into  $k$  colours is a colouring of its vertices and edges such that two adjacent vertices receive different colours, two adjacent edges receive different colours and a vertex and an edge incident to it receive different colours. A total colouring with  $k$  colours is a *k-total-colouring*. For every graph  $G$ , let  $c_k^T(G)$  be the number of  $k$ -total-colourings of  $G$ .

For each 4-edge-colouring  $c$  of a cubic graph  $G$ , there is at most one 4-total-colouring of  $G$  whose restriction to  $E(G)$  equals  $c$ . Indeed, the colours of the three edges incident to a vertex force the colour of this vertex. Hence if  $G$  is cubic, we have that  $c_4^T(G) \leq c_4(G)$ .

By the method described in the previous section, one can show that if  $G$  is 2-connected, then  $c_4(G) = O(2^{n/2} \cdot 6^{n/2})$ , and so  $c_4^T(G) = O(2^{n/2} \cdot 6^{n/2})$ . We now obtain better upper bounds for  $c_4^T$ .

**Theorem 14.** Let  $G$  be a 2-connected subcubic graph. Then  $c_4^T(G) \leq 3 \cdot 2^{2n-n_3/2}$ .

*Proof.* Assume first that  $G$  is a cycle  $(v_1, \dots, v_n, v_1)$ . Let us totally colour it greedily starting from  $v_1$ . There are 4 possible colours for  $v_1$ , and then 3 possible colours for  $v_1v_2$ . Afterwards for every  $i \geq 2$ , there are at most two possible colours for  $v_i$  (the ones distinct from the colours of  $v_{i-1}$  and  $v_{i-1}v_i$ ) and then at most two possible colours for  $v_iv_{i+1}$  (the ones distinct from the colours of  $v_{i-1}v_i$  and  $v_i$ ). Hence  $c_4^T(G) \leq 4 \cdot 3 \cdot 2^{2n-2} = 3 \cdot 2^{2n}$ .

Assume now that  $G$  is not a cycle. Let  $s$  and  $t$  be two distinct vertices of degree 3. Consider an  $(s, t)$ -ordering  $v_1, v_2, \dots, v_n$  of  $V(G)$ , which exists by Lemma 3, and the orientation  $D$  of  $G$  according to this ordering. Then  $d^+(v_1) = 3 = d^-(v_n)$  and  $d^-(v_1) = 0 = d^+(v_n)$ . Let  $A^+$  (resp.  $A^-$ ) be the set of vertices of outdegree 2 (resp. indegree 2) in  $D$  and  $A_2$  be the set of vertices of degree 2 in  $G$ . As in the proof of Theorem 4, we have  $|A_2| = n - n_3$ , and  $|A^+| = |A^-| = (n_3 - 2)/2$ .

Now for  $i = 1$  to  $n - 1$ , we enumerate the  $p_i$  partial 4-total-colourings of vertices in  $\{v_1, \dots, v_i\}$  and arcs with tail in  $\{v_1, \dots, v_i\}$ . For  $i = 1$ , there are  $4! = 24$  such colourings, since  $v_1$  and its three incident arcs must receive different colours.

For each  $1 < i < n$ , when we extend the partial total colourings. Two cases may arise.

- If  $d_D^-(v_i) = 1$ , then there are two choices to colour  $v_i$  and then two other choices to colour the (at most two) arcs leaving  $v_i$ . Hence  $p_i \leq 4p_{i-1}$ .
- If  $d_D^-(v_i) = 2$ , then there are at most two choices to colour  $v_i$  and then the colour of the arc leaving  $v_i$  is forced since three colours are forbidden by  $v_i$  and its two entering arcs. Hence  $p_i \leq 2p_{i-1}$ .

Finally, we need to colour  $v_n$ . Since its three entering arcs are coloured its colour is forced (or it is impossible to extend the colouring).

Hence an easy induction shows that  $c_4^T(G) = p_{n-1} \leq 24 \cdot 4^{|A_2|+|A^+|} \cdot 2^{|A^-|} = 3 \cdot 2^{2n-n_3/2}$ .  $\square$

A *leaf* of a tree is a degree one vertex. A vertex of a tree which is not a leaf is called a *node*. A tree is *binary* if all its nodes have degree 3.

**Proposition 15.** If  $T$  is a binary tree of order  $n$ , then  $c_4^T(T) = 3 \cdot 2^{3n/2}$ .

*Proof.* By induction on  $n$ , the results holding easily when  $n = 2$ , that is when  $T = K_2$ .

Suppose now that  $T$  has more than two vertices. There is a node  $x$  which is adjacent to two leaves  $y_1$  and  $y_2$ . Consider the tree  $T' = T - \{y_1, y_2\}$ . By the induction hypothesis,  $c_4^T(T') = 3 \cdot 2^{3(n-2)/2}$ . Now each 4-total-colouring of  $T'$  may be extended into exactly eight 4-total-colourings of  $T$ . Indeed the two colours of  $x$  and its incident edge in  $T'$  are forbidden for  $xy_1$  and  $xy_2$ , so there are two possibilities to extend the colouring to these edges, and then for each  $y_i$ , there are two possible colours available. Hence  $c_4^T(T) = 8 \cdot c_4^T(T') = 3 \cdot 2^{3n/2}$ .  $\square$

**Theorem 16.** *Let  $G$  be a connected cubic graph. Then  $c_4^T(G) \leq 3 \cdot 2^{3n/2}$ .*

*Proof.* Let  $F$  be the subgraph induced by the cutedges of  $G$ . Then  $F$  is a forest. Consider a tree of  $F$ . It is binary, its leaves are in different non-trivial 2-connected components of  $G$ , and every node is a trivial 2-connected component of  $G$ .

A subgraph  $H$  of  $G$  is *full* if it is induced on  $G$ , connected and such that for every non-trivial 2-connected component  $C$ ,  $H \cap C$  is empty or is  $C$  itself and for every tree  $T$  of  $F$ ,  $H \cap T$  is empty, or is just one leaf of  $T$  or is  $T$  itself. Observe that a full subgraph has minimum degree at least 2.

We shall prove that for every full subgraph  $H$ ,  $c_4^T(H) \leq 3 \cdot 2^{2n(H) - n_3(H)/2}$ . We proceed by induction on the number of 2-connected components of  $H$ . If  $H$  is 2-connected, then the result holds by Theorem 14.

Suppose now that  $H$  is not 2-connected. Then  $H$  contains a tree  $T$  of  $F$ . Let  $v_1, \dots, v_p$  be the leaves of  $T$ ,  $e_i$ ,  $1 \leq i \leq p$  the edge incident to  $v_i$  in  $T$  and  $N$  the set of nodes of  $T$ . Then  $H - N$  has  $p$  connected components  $H_1, \dots, H_p$  such that  $v_i \in H_i$  for all  $1 \leq i \leq p$ . Furthermore, each  $H_i$  is a full subgraph of  $G$ .

Let  $c$  be a 4-total-colouring of  $T$ . It can be extended to  $H_i$  by any 4-total-colouring of  $H_i$  such that  $v_i$  is coloured  $c(v_i)$  and the two edges incident to  $v_i$  in  $H_i$  are coloured in  $\{1, 2, 3, 4\} \setminus \{c(v_i), c(e_i)\}$ . There are  $\frac{1}{12}c_4^T(H_i)$  such colourings because each of them correspond to exactly twelve 4-total colourings of  $H_i$  obtained by permuting the colour of  $v_i$  (there are 4 possibilities) and then the colour of  $e_i$  (there are 3 possibilities). Hence each 4-total-colouring of  $T$  can be extended into  $\prod_{i=1}^p \frac{1}{12}c_4^T(H_i)$  4-total-colourings of  $H$  and so

$$c_4^T(H) = c_4^T(T) \cdot \prod_{i=1}^p \frac{1}{12}c_4^T(H_i).$$

Now by Proposition 15,  $T$  has  $3 \cdot 2^{3n(T)/2}$  4-total-colourings, and since  $H_i$  is full  $c_4^T(H_i) \leq 3 \cdot 2^{2n(H_i) - n_3(H_i)/2}$  by the induction hypothesis. Moreover,  $n_3(H) = n(T) + \sum_{i=1}^p n_3(H_i)$  and  $n(H) = n(T) + \sum_{i=1}^p n(H_i) - p$ . Hence  $c_4^T(H) \leq 3 \cdot 2^{2n(H) - n_3(H)/2}$ .  $\square$

As previously for the edge colourings of graphs, we derive from Theorem 16 an algorithm to enumerate all the 4-total-colourings of a cubic graph. The proof is similar to the one of Corollary 6.

**Corollary 17.** *There is an algorithm to enumerate all the 4-total-colourings of a connected cubic graph on  $n$  vertices in time  $O^*(2^{3n/2})$  and polynomial space.*

The bound of Theorem 16 is seemingly not tight. Indeed, in Theorem 14, the equation  $p_i \leq 2p_{i-1}$  when  $d_D(v_i)$  often overestimates  $p_i$ , because there are two choices to colour  $v_i$  only if the two colours appearing on its two entering arcs are the same two as the ones assigned to the tails of these arcs. If not the colour of  $v_i$  is forced or  $v_i$  cannot be coloured.

**Problem 18.** What is  $c_4^T(n)$ , the maximum of  $c_4^T(G)$  over all connected graphs of order  $n$ ?

We shall now give a lower bound on  $c_4^T(n)$ . A binary tree is *nice* if its set of leaves may be partitioned into pairs of *twins*, i.e. leaves at distance 2. Clearly, every nice binary tree  $T$  has an even number of leaves and thus  $n(T) \equiv 2 \pmod{4}$ . Moreover if  $n(T) = 4p + 2$ , then  $T$  has  $2p$  nodes and  $p + 1$  pairs of twins. A *noodle tree* is a cubic graph obtained from a nice binary tree by adding two parallel edges between each pair of twins.

**Proposition 19.** *Let  $p$  be a positive integer and  $n = 4p + 2$ . If  $G$  is a noodle tree  $G$  of order  $n$ , then  $c_4^T(G) = \frac{3}{\sqrt{2}} \cdot 2^{5n/4}$ .*

*Proof.* Let  $X_1, \dots, X_{p+1}$  be the pairs of twins of  $G$ , and let  $T$  be the binary tree  $G - \bigcup_{i=1}^{p+1} X_i$ . Let us label the leaves of  $T$ ,  $y_1, \dots, y_{p+1}$  such that for all  $1 \leq i \leq p+1$ ,  $y_i$  is adjacent to the two vertices of  $X_i$  in  $G$ .

Every 4-total-colouring of  $T$ , may extended in exactly 4 ways to each pair of twins  $X_i = \{x_i, x'_i\}$  and their incident edges. Indeed, without loss of generality we may assume that  $y_i$  is coloured 1 and its incident edge in  $T$  is coloured 2. Then the edges  $y_i x_i$  and  $y_i x'_i$  must be coloured in  $\{3, 4\}$ , which can be done in two possible ways. For each of these possibilities, the parallel edges between  $x_i$  and  $x'_i$  must be coloured in  $\{1, 2\}$ , which again can be done in two possible ways. Finally, we must colour  $x_i$  (resp.  $x'_i$ ) with the colour of  $y_i x'_i$  (resp.  $y_i x_i$ ).

Hence  $c_4(G) = 4^{p+1} \cdot c_4(T)$ , and so by Proposition 15,  $c_4(G) = 3 \cdot 2^{5p+2}$ .  $\square$

## Acknowledgement

The authors would like to thank their children for suggesting them some graph names.

## References

- [1] R. Beigel and D. Eppstein. 3-coloring in time  $O(1.3289^n)$ . *Journal of Algorithms* 54:168–204, 2005.
- [2] A. Björklund and T. Husfeldt. Inclusion-exclusion algorithms for counting set partitions. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, pp. 575–582. IEEE, Los Alamitos, 2006.
- [3] A. Björklund, T. Husfeldt, P. Kaski, and M. Koivisto. Narrow sieves for parameterized paths and packings. arXiv:1007.1161v1, 2010.
- [4] A. Björklund, T. Husfeldt, and M. Koivisto. Set partitioning via Inclusion-Exclusion. *SIAM J. Comput.*, 39(2):546–563, 2009.
- [5] J. A. Bondy and U. S. R. Murty. *Graph theory*. Graduate Texts in Mathematics 244, Springer, 2008 (2nd ed.).
- [6] S. Even and R. E. Tarjan. Computing an st-numbering. *Theoretical Computer Science*, 2(3):339–344, 1976.

- [7] S. Even and R. E. Tarjan. Corrigendum: Computing an st-numbering. *Theoretical Computer Science*, 4(1):123, 1977.
- [8] F. V. Fomin, S. Gaspers, and S. Saurabh. Improved exact algorithms for counting 3- and 4-colorings. In *Lin, G. (ed.) COCOON 2007, Lecture Notes in Computer Science* 4598:65–74, 2007.
- [9] M. R. Garey and D. S. Johnson, *Computers and intractability. A guide to the theory of NP-completeness*. A Series of Books in the Mathematical Sciences. W. H. Freeman and Co., San Francisco, Calif., 1979.
- [10] P. A. Golovach, D. Kratsch, and J. F. Couturier. Colorings with few colors: counting, enumeration and combinatorial bounds. In *Proceedings of WG 2010, Lecture Notes in Computer Science* 6410:39–50, 2010.
- [11] I. Holyer. The NP-completeness of edge-colouring. *SIAM J. Computing* 2:225–231, 1981.
- [12] T. R. Jensen and B. Toft. *Graph coloring problems*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., New-York, 1995.
- [13] D. E. Knuth. *The art of Computer Programming, volume 4A: Combinatorial Algorithms, Part 1*. Addison-Wesley Professional, 2011.
- [14] M. Koivisto. An  $O(2^n)$  algorithm for graph coloring and other partitioning problems via inclusion-exclusion. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, pp. 583–590. IEEE, Los Alamitos, 2006.
- [15] L. Kowalik. Improved edge-coloring with three colors. *Theoret. Comp. Sci.* 410:3733–3742, 2009.
- [16] A. Lempel, S. Even, and I. Cederbaum. An algorithm for planarity testing of graphs. In *P. Rosenstiehl, editor, Proceedings of the International Symposium on the Theory of Graphs (Rome, July 1966)*: 215–232. Gordon and Breach, 1967.
- [17] A. Sánchez-Arroyo. Determining the total colouring number is NP-hard. *Discrete Math.* 78(3):315–319, 1989.