



HAL
open science

Sur deux problèmes d'apprentissage automatique : la détection de communautés et l'appariement adaptatif

Lennart Gulikers

► **To cite this version:**

Lennart Gulikers. Sur deux problèmes d'apprentissage automatique : la détection de communautés et l'appariement adaptatif. Machine Learning [stat.ML]. Université Paris sciences et lettres, 2017. Français. NNT : 2017PSLEE062 . tel-01834967

HAL Id: tel-01834967

<https://theses.hal.science/tel-01834967v1>

Submitted on 11 Jul 2018

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.

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'École Normale Supérieure

Sur deux problèmes d'apprentissage automatique:
la détection de communautés et l'appariement adaptatif

École doctorale n°386

L'ÉCOLE DOCTORALE DE SCIENCES MATHÉMATIQUES DE PARIS-CENTRE

Spécialité MATHÉMATIQUES APPLIQUÉS

Soutenue par **Lennart GULIKERS**
le 13 novembre 2017

Dirigée par **Laurent MASSOULIÉ**
et **Marc LELARGE**

COMPOSITION DU JURY :

M Nicolas Vayatis
ENS Paris-Saclay,
Président

M Joe Neeman
UT Austin,
Rapporteur

M Alexandre Proutière
KTH Stockholm,
Rapporteur

Mme Lenka Zdeborová
CNRS et CEA,
Membre du jury

M Laurent Massoulié
INRIA et Microsoft,
Directeur de thèse

M Marc Lelarge
Safran, INRIA et ENS,
Co-Directeur de thèse



Abstract

In this thesis, we study two problems of machine learning: (I) community detection and (II) adaptive matching.

I) It is well-known that many networks exhibit a community structure. Finding those communities helps us understand and exploit general networks. In this thesis we focus on community detection using so-called spectral methods based on the eigenvectors of carefully chosen matrices. We analyse their performance on artificially generated benchmark graphs. Instead of the classical Stochastic Block Model (which does not allow for much degree-heterogeneity), we consider a Degree-Corrected Stochastic Block Model (DC-SBM) with weighted vertices, that is able to generate a wide class of degree sequences. We consider this model in both a dense and sparse regime. In the dense regime, we show that an algorithm based on a suitably normalized adjacency matrix correctly classifies all but a vanishing fraction of the nodes. In the sparse regime, we show that the availability of only a small amount of information entails the existence of an information-theoretic threshold below which no algorithm performs better than random guess. On the positive side, we show that an algorithm based on the non-backtracking matrix works all the way down to the detectability threshold in the sparse regime, showing the robustness of the algorithm. This follows after a precise characterization of the non-backtracking spectrum of sparse DC-SBM's. We further perform tests on well-known real networks.

II) Online two-sided matching markets such as Q&A forums and online labour platforms critically rely on the ability to propose adequate matches based on imperfect knowledge of the two parties to be matched. We develop a model of a task / server matching system describing platform operation in the presence of such uncertainty. For this model, we give a necessary and sufficient condition for an incoming stream of tasks to be manageable by the system. We further identify a so-called back-pressure policy under which the throughput that the system can handle is maximal. We show that this policy achieves strictly larger throughput than a natural greedy policy. Finally, we validate our model and confirm our theoretical findings with experiments based on user-contributed content on an online platform.

Keywords: Machine learning, community detection, social networks, degree-corrected stochastic block models, spectral methods, random matrices, non-backtracking matrix, random graphs, recommendation systems, reinforcement learning, queueing theory.

Résumé

Dans cette thèse, nous étudions deux problèmes d'apprentissage automatique: (I) la détection des communautés et (II) l'appariement adaptatif.

I) Il est bien connu que beaucoup de réseaux ont une structure en communautés. La détection de ces communautés nous aide à comprendre et exploiter des réseaux de tout genre. Cette thèse considère principalement la détection des communautés par des méthodes spectrales utilisant des vecteurs propres associés à des matrices choisies avec soin. Nous faisons une analyse de leur performance sur des graphes artificiels. Au lieu du modèle classique connu sous le nom de « Stochastic Block Model » (dans lequel les degrés sont homogènes) nous considérons un modèle où les degrés sont plus variables: le « Degree-Corrected Stochastic Block Model » (DC-SBM). Dans ce modèle les degrés de tous les noeuds sont pondérés - ce qui permet de générer des suites des degrés hétérogènes. Nous étudions ce modèle dans deux régimes: le régime dense et le régime « épars », ou « dilué ». Dans le régime dense, nous prouvons qu'un algorithme basé sur une matrice d'adjacence normalisée réussit à classifier correctement tous les noeuds sauf une fraction négligeable. Dans le régime épars il existe un seuil en termes de paramètres du modèle en-dessous lequel n'importe quel algorithme échoue par manque d'information. En revanche, nous prouvons qu'un algorithme utilisant la matrice « non-backtracking » réussit jusqu'au seuil - cette méthode est donc très robuste. Pour montrer cela nous caractérisons le spectre des graphes qui sont générés selon un DC-SBM dans son régime épars. Nous concluons cette partie par des tests sur des réseaux sociaux.

II) Les marchés d'intermédiation en ligne tels que des plateformes de Question-Réponse et des plateformes de recrutement nécessitent un appariement basé sur une information incomplète des deux parties. Nous développons un modèle de système d'appariement entre tâches et serveurs représentant le comportement de telles plateformes. Pour ce modèle nous donnons une condition nécessaire et suffisante pour que le système puisse gérer un certain flux de tâches. Nous introduisons également une politique de « back-pression » sous lequel le débit gérable par le système est maximal. Nous prouvons que cette politique atteint un débit strictement plus grand qu'une politique naturelle « gloutonne ». Nous concluons en validant nos résultats théoriques avec des simulations entrainées par des données de la plateforme Stack-Overflow.

Mots clés: Apprentissage automatique, détection des communautés, réseaux sociaux, degree-corrected stochastic block models, méthodes spectrales, matrices aléatoires, matrice non-backtracking, graphes aléatoires, appariement adaptatif, apprentissage par renforcement, théorie des files d'attente.

Remerciements

Tout d'abord, j'exprime toute ma gratitude à Laurent Massoulié et Marc Lelarge. Avec eux comme co-directeurs de ma thèse j'ai pu profiter d'une formation exceptionnelle. Ils m'ont accueilli chaleureusement dès mon arrivée à Paris. Je remercie Laurent pour les nombreuses discussions enrichissantes. Grâce à ces discussions j'ai appris énormément de choses. Je le remercie également pour tous les conseils scientifiques et non-scientifiques qu'il m'a donnés. Je remercie Marc pour toutes les discussions sur des sujets passionnants, pour sa disponibilité et pour ce qu'il m'a apporté sur le plan humain. Je me réjouis d'avance de poursuivre notre coopération.

Mes remerciements s'adressent également à Lenka Zdeborová et Nicolas Vayatis pour leur accord de participer au jury de cette thèse. J'exprime ma vive gratitude à Joe Neeman et Alexandre Proutière pour avoir accepté d'être rapporteurs. Je remercie Alexandre également pour toutes les discussions qu'on a eues. Je remercie également Milan Vojnovic pour notre collaboration scientifique.

De très sincères remerciements vont à Virag Shah: pour notre collaboration scientifique (voir chapitre 8), pour sa sagesse et pour m'avoir fait découvrir la cuisine indienne.

Par ailleurs, je remercie Nidhi Hegde, Céline Comte et Thomas Bonald pour les discussions scientifiques, les déjeuners et les pauses café au Lincs. Je remercie Nidhi notamment pour m'avoir aidé à améliorer mes présentations et pour sa contribution au travail présenté dans le chapitre 8 de cette thèse. Merci à Thomas pour m'avoir appris un peu d'argot.

Finalement, je voudrais remercier Anya pour sa présence dans ma vie et pour son soutien fidèle.

Contents

I	Community Detection	5
1	Introduction	6
1.1	Motivation and preview of results	6
1.2	What is a community?	7
1.3	Methods of community detection	8
1.3.1	Spectral algorithms	8
1.3.2	Modularity optimization	9
1.3.3	k -means	10
1.3.4	Divisive algorithms	10
1.3.5	Belief propagation	11
1.4	Real datasets and benchmark graphs	12
2	Random matrices	14
3	The Degree-Corrected Stochastic Block Model	17
3.1	Dense regime	18
3.1.1	Literature on the <i>ordinary</i> PPM	18
3.1.2	Literature on the DC-SBM	19
3.1.3	Results in this thesis	19
3.2	Sparse regime	20
3.2.1	Literature on the <i>ordinary</i> SBM	20
3.2.2	Results in this thesis	21
4	A spectral algorithm for the <i>dense</i> DC-SBM	22
4.1	Introduction	22
4.2	Conditions on the model	23
4.3	Main results	24
4.4	Relation to other spectral algorithms in the literature	25
4.4.1	Adjacency matrix	25
4.4.2	Spectral clustering on ratios of eigenvectors	26
4.4.3	Laplacian	26
4.4.4	Regularized spectral clustering	29
4.4.5	Degree-normalized adjacency matrix	29
4.5	Discussion	30
4.5.1	When does the degree-normalized adjacency matrix fail?	30
4.5.2	Interpretation of the conditions	30
4.5.3	Random Walk point of view	31
4.6	Outline of proof of main theorem	32
4.7	Proofs	35
4.7.1	Main theorem	35
4.7.2	Comparison to spectral analysis on the adjacency matrix	46
4.7.3	Interpretation of the conditions	50
4.8	Directions for future research	52
4.8.1	Exact recovery	52
4.8.2	Non-constant B	52

4.8.3	Sparser graphs	52
5	Information-theoretical limits for the <i>sparse</i> DC-SBM	53
5.1	Introduction	53
5.2	Main results	53
5.3	Proof heuristics	54
5.4	Description of q -type Poisson-mixture branching process	54
5.5	General proof idea and outline	55
5.6	Broadcasting on the branching process	56
5.7	Coupling of local neighbourhood	58
5.7.1	Alternative description of branching process	58
5.7.2	Coupling	59
5.8	No long-range correlation in DC-SBM	61
5.9	Proofs	63
5.9.1	Broadcasting on the branching process	63
5.9.2	Coupling of local neighbourhood	63
5.9.3	No long-range correlation in DC-SBM	66
6	Non-backtracking spectrum of the <i>sparse</i> DC-SBM	71
6.1	Introduction	71
6.1.1	Linearisation of belief propagation	72
6.1.2	Proof heuristics	73
6.2	Main Results	74
6.2.1	Quasi Ramanujan property	75
6.2.2	Notation	76
6.3	Outline and proof strategy	76
6.4	Preliminaries	78
6.4.1	Background on non-backtracking matrix	78
6.4.2	Extension of Bauer-Fike Theorem	78
6.5	Proof of Theorem 6.2.1	79
6.5.1	The case $\mu_2^2 > \rho$	79
6.5.2	The case $\mu_2^2 \leq \rho$	81
6.6	Poisson-mixture two-type branching processes	81
6.6.1	A theorem of Kesten and Stigum	81
6.6.2	Quantitative version of the Kesten-Stigum theorem	82
6.6.3	$B^\ell B^{*\ell} \tilde{\chi}_k$ on trees: a cross generation functional	83
6.6.4	Orthogonality: Decorrelation in branching process	85
6.7	Coupling of local neighbourhood	85
6.7.1	Coupling	85
6.7.2	Geometric growth	87
6.8	A weak law of large numbers for local functionals on the DC-SBM	87
6.8.1	Application with some specific local functionals	88
6.9	Proof of Propositions 6.5.1 and 6.5.3	90
6.10	Norm of non-backtracking matrices	91
6.10.1	Proof of Proposition 6.5.2	94
6.10.2	Comparison with the Stochastic Block Model in [17]	94
6.11	Detection: Proof of Theorem 6.2.2	95
6.12	Proofs of Section 6.6	97
6.13	Proofs of Section 6.7	107
6.14	Proofs of Section 6.8	111
6.15	Proofs of Section 6.10	115
6.15.1	Bound on $\ \Delta^{(k)}\ $	115
6.15.2	Bound on $\ \Delta^{(k)}\chi_i\ $	118
6.15.3	Bound on $\ R_k^{(\ell)}\ $	119
6.15.4	Bound $\ KB^{(k)}\ $	122
6.15.5	Bound on $\ S_k^{(l)}\ $	123

6.16	Proofs of Section 6.11	124
7	Tests on real networks	126
7.1	Normalized Adjacency Matrix	126
7.2	Non-backtracking Matrix	126
II	Adaptive Matching	129
8	Adaptive Matching for Expert Systems with Uncertain Task Types	130
8.1	Introduction	130
8.2	Problem Setting	132
	8.2.1 Single Task Scenario	132
	8.2.2 Online Task Arrivals	133
8.3	Baseline Policies	133
8.4	Optimal Stability	135
8.5	Experimental Results	136
8.6	Related Work	138
8.7	Conclusion	138
8.8	Preprocessing phase	139
	8.8.1 Estimation of success-probabilities	140
8.9	Proofs	143
	8.9.1 Proof of Theorem 8.3.4	143
	8.9.2 Stability Threshold under Random Policy for an Arbitrary System	145
	8.9.3 Proof of Theorem 8.4.1	146
	8.9.4 Proof of Lemma 8.9.3	151
	Appendices	159
A	Algebraic Preliminaries	160

Part I

Community Detection

Chapter 1

Introduction

1.1 Motivation and preview of results

We live in a time of unprecedented volumes of data and computing power. Many such datasets can be described as graphs of interacting items, and finding communities of "likewise" items is a first step in analysing them, either as a means by itself or as a preprocessing phase for other learning problems (e.g., as dimensionality reduction).

Community detection has many important applications. For example, it leads to a better understanding of social behaviour on networks as Facebook, Google+, Twitter and Flickr [121]. It is also useful to cluster web clients or discovering topical similarities on the web [74]. Other applications lay in shopping networks: targeted advertising, logistics and future recommendation [77]. Clustering is even used in natural language processing, see for instance the appendix in [9].

Finding communities is a central problem in unsupervised machine learning. It is challenging because the ground truth is often unknown, making it difficult to evaluate the result of community detection algorithms. Further, often one has access only to the topological configuration of graphs without any side information, so that even the number and size of communities is a priori unknown and algorithms thus necessarily need to learn those.

Many proposed community detection algorithms rely on heuristics and come without formal guarantees, except for their reported behaviour on specific networks. It is however important to understand if the outputted community structure of an algorithm is meaningful or just an artefact. In particular, what are the limitations of an algorithm: can we always (partly) recover the community-structure from an observed graph?

A principled way to evaluate the performance of algorithms is to analyse their behaviour on benchmark graphs. The most commonly used model for this purpose is the Stochastic Block Model (SBM): arguably the simplest extension of Erdős-Rényi graphs to a network with a community structure, see below for definitions. The SBM has as advantages that it comes with a known community structure (i.e., the ground truth) and that it is analytically tractable. However, it is seldom a good fit to observed real data, as the SBM does not allow for heterogeneity in its degree sequence (i.e., all vertices in a community have on average the same number of edges). Indeed, many real networks display a strong heterogeneity in their degree sequences, often following a power-law [4].

In this thesis, we consider the Degree-Corrected Stochastic Block Model (DC-SBM), an extension of the SBM which allows for strong degree-heterogeneity. We study this model in two regimes: a dense regime where the number of edges per node grows with the size of the network and a sparse regime where the number of edges per node stays constant when the network grows. We evaluate the performance of certain, so-called spectral algorithms on this model. In particular, we show that an algorithm based on a suitably normalized adjacency matrix recovers all but a vanishing fraction of vertices in the dense regime of the DC-SBM. Further, in the

sparse (and challenging, because there is not much information) regime, we show that on the one hand there exists an information-theoretic threshold (in terms of the model’s parameters) below which no algorithm does better than random guess. On the other hand, we prove that an algorithm based on the non-backtracking matrix does significantly better than random guess all the way down to this threshold. This is an important result as traditional methods are known to break down long before the information-theoretic limits are reached, see [125] and the figures therein. We conclude by implementing various community detection algorithms to demonstrate their performance on real networks.

1.2 What is a community?

It is an important task to define formally the notion of a community. According to [44], no definition is universally accepted. We summarize the observations in the latter article here: What we define to be a community depends on both the underlying network and the envisaged final application. Usually, however we expect a community to be a set of vertices having more internal edges than edges connecting it to its complement in the graph. Formally we could consider the *intra-cluster density* $\delta_i(C)$ and the *inter-cluster density* $\delta_e(C)$ of a subgraph C and compare those with the average edge density δ of the graph. Intuitively, the subgraph C would then be a community, if $\delta_i(C) \gg \delta \gg \delta_e(C)$. Maximizing $\sum_{C \in \mathcal{C}} (\delta_i(C) - \delta_e(C))$ over all partitions \mathcal{C} is however too expensive in practice if the graph is large, so that many community-detection algorithms find an approximative solution to this problem.

We give here the main definitions that can be found in the exposition of [44]: local definitions, global definitions and definitions based on vertex similarities.

We start with the main local definitions. A very stringent such definition is requiring that the subgraph forms a clique (i.e., there is an edge from any vertex in the subgraph to any other vertex in the subgraph). In social networks this would mean that every group member is friends with any other group member. Note that this definition does not allow for heterogeneous relationships between the vertices (or people). Further, finding cliques in a graph is known to be an NP-complete problem [16].

An *LS-set* or *strong community* [80, 107] is a subgraph such that the internal degree of a vertex is larger than its external degree. Weakening this definition leads to *weak communities* [107]: subgraphs with total internal degree exceeding its total external degree. The assortative SBM with two equal-sized communities for instance can be partitioned into two weak communities.

In general we expect communities to have a *small cut size*, i.e., few edges connecting a subset to its complement in the graph.

We now turn to the most popular global definition, based on *modularity* [101]. The latter measures the deviation of a graph with communities from a null model. In particular, the null model is the original graph with all its edges rewired at random so as to conserve the original degree sequence. A subgraph is then a community if its number of internal edges is larger than its expectation in the null model. Note that according to this definition a community is a more densely-connected subgraph than would have been expected from chance.

Communities can also be defined with respect to some kind of similarity measure on the vertices: a community is then a group of vertices that are similar to each other with respect to this measure. For instance, documents could be clustered according to their similarity (e.g., discussing the same topics). Or, if the vertices could be embedded into an Euclidean space, the inverse of the distance between them defines a similarity measure.

We further note that communities can also be overlapping rather than partitioning a graph: the number of possibilities is then much higher and therefore mathematically even more challenging. We shall not consider this direction in this thesis.

1.3 Methods of community detection

In this thesis we focus primarily on spectral algorithms, as their reliability can be formally verified. For completeness we also state some well-known alternative (heuristic) methods.

1.3.1 Spectral algorithms

The general idea of spectral methods is to associate a matrix to a set of objects and use only a few of its top eigenvectors to cluster the objects. The original objects might be embedded in a high-dimensional space and projecting them on the low-dimensional space spanned by the top eigenvectors is thus a **dimensional reduction**. It is of-course important that the top eigenvectors contain enough information regarding the community-structure. Commonly used matrices are the adjacency matrix and Laplacian. Both have their shortcomings and we focus in this thesis instead on a normalized version of the adjacency matrix (different from the Laplacian) and on the non-backtracking matrix. See Sections 3.1.1, 3.1.2 and 4.4 for a discussion of (traditional) spectral methods.

The spectrum of suitably chosen matrices often gives information about the underlying community structure. Indeed, the spectrum often consists of a bulk of eigenvalues confined to a bounded set plus a few outliers that correspond to the communities.

A good introduction to spectral clustering is given in [117], where the following is explained in more detail: Consider a graph on n vertices, with adjacency matrix A . Denote by D the diagonal matrix with $D_{uu} = \sum_v A_{uv}$. Then, the *unnormalized* Laplacian matrix is defined as $L = D - A$, the *symmetric normalized* Laplacian as $L_{\text{sym}} = D^{-1/2}LD^{-1/2}$ and the *random walk* Laplacian as $L_{\text{rw}} = D^{-1}L$.

Propositions 1 and 2 in [117] state that L has n non-negative real-valued eigenvalues $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. The multiplicity K of eigenvalue 0 equals the number of connected components C_1, \dots, C_K in the graph and eigenvectors corresponding to eigenvalue 0 are indicator functions on C_1, \dots, C_K . Similar statements hold for the normalized Laplacians, according to Proposition 4 in [117].

Thus, for a graph that has K communities with few in-between edges we expect the first K eigenvalues to be close to 0 and the corresponding eigenvectors to be close to the respective indicator functions. This leads to the following spectral algorithm: Compute the first K eigenvectors u_1, \dots, u_K of the Laplacian and form the matrix U containing u_1, \dots, u_K as its columns. Let $y_u \in \mathbb{R}^{n \times K}$ be the vector containing the u -th row of U . Apply k -means (see below) to partition the set $\{y_1, \dots, y_n\}$ into clusters A_1, \dots, A_K . Approximate the community-membership of vertex u by i if $y_u \in A_i$. Note that $\{y_u\}_u$ are thus low-dimensional representatives of the vertices.

Formal guarantees for this type of algorithms follow usually from perturbation arguments. A key ingredient is often the Davis-Kahan theorem, informally stated as follows: Consider $\hat{A} = A + \delta A$, for symmetric matrices A and δA . Let $S \subset \mathbb{R}$ be an interval. Denote $\lambda_S(A)$ for the set of all eigenvalues of A contained in S and denote the corresponding set of eigenvectors by V . Denote by $\lambda_S(\hat{A})$ and \hat{V} the analogous quantities for \hat{A} . Define $\Delta(A) = \min\{|\lambda - s| : \lambda \text{ eigenvalue of } A, \lambda \notin S, s \in S\}$ Then the distance between V and \hat{V} , expressed in terms of the *canonical angles* between them, is bounded by $\|\delta A\|/\Delta(A)$, where $\|\cdot\|$ is either the Frobenius norm or two-norm. For the precise definition of canonical angle between two subspaces, we refer to [117], it is a measure for the alignment of orthonormal bases for the respective subspaces.

To see this theorem in work, consider a graph with K communities C_1, \dots, C_K and let \hat{L} be its associated Laplacian and L the Laplacian of the *idealized* graph were the edges in-between communities have been removed. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of L . Put for small $\gamma > 0$, $S = [0, \gamma]$. Then $[0, \lambda_K] \subset S$. Hence, $\Delta(A) \simeq |\lambda_{K+1} - \lambda_K|$ and we deduce that the eigenvectors of L and \hat{L} are close when

$\|\delta A\|$ is small and $|\lambda_{K+1} - \lambda_K|$ large. The latter quantity thus plays an important role and is often called *eigen gap* or *spectral gap*.

Computation of the eigenvectors is usually done with the power method or Krylov subspace methods (such as the Lanczos algorithm). The convergence is faster for larger spectral gaps, when calculating only the first K eigenvectors.

Historically, spectral clustering was initiated in 1973 with the work [38] on a spectral algorithm using the adjacency matrix. In the same year the connection between bi-partitions of graphs and the second eigenvector of the Laplacian was discovered [43].

1.3.2 Modularity optimization

In a graph where the edges are drawn completely at random, one does not expect a cluster-structure to be present. It therefore makes sense to compare the number of edges in a given subgraph with the number of edges expected to be present in a *null-model*, where the edges are drawn regardless of community structure. A commonly used null-model is the configuration model that matches the observed degree sequence, denoted by $\{k_u\}_u$, in the following way: Every vertex u is given k_u half-edges, and every half-edge is merged with another half-edge chosen uniformly at random. The probability for an edge to be present between vertices u and v is thus approximately $\frac{k_u k_v}{2m}$, where m is the number of edges in the graph. With this null-model, the *modularity* [101] of the graph where a vertex u belongs to community σ_u is given by

$$Q = \frac{1}{2m} \sum_{u,v} \left(A_{uv} - \frac{k_u k_v}{2m} \right) \delta(\sigma_u, \sigma_v),$$

where $\delta(x, y)$ is one if $x = y$ and zero otherwise.

Note that Q can be rewritten in terms of $\frac{\ell_i}{m}$, the fraction of edges within community i , and $\left(\frac{d_i}{2m}\right)^2$, the expected fraction of edges in that community in the null-model:

$$Q = \sum_i \left(\frac{\ell_i}{m} - \left(\frac{d_i}{2m} \right)^2 \right).$$

Hence, we expect that a large Q corresponds to good partitions. By definition, $Q \leq 1$, and $Q = 0$ when all vertices are put in a single community. If $Q \leq 0$ for all partitions, then no community structure is present.

In general it is very costly to calculate Q for all possible partitions, and algorithms have been developed to find an approximate solution to this maximization problem.

One such algorithm is proposed in [102]: Starting with the partition where every vertex constitutes a community, we repeatedly obtain partitions by joining communities together by choosing the merge that results in the greatest increase (or smallest decrease) of the modularity. In this way we obtain a collection of partitions of different dimensions, and the partition with largest modularity is the output of the algorithm.

Using the sparsity of the adjacency matrix, this algorithm has been extended in [26] to handle graphs in the order of 10^6 vertices, and to mega-scale graphs with 10^7 vertices in [118].

In [103] a spectral algorithm is proposed based on the modularity for a partition with two communities C_1 and C_2 . Let σ be the vector of community-memberships, $\sigma_u = 1$ if u belongs to C_1 and $\sigma_u = -1$ otherwise. Then,

$$Q = \frac{1}{2m} \sum_{u,v} \left(A_{uv} - \frac{k_u k_v}{2m} \right) \frac{\sigma_u \sigma_v + 1}{2} = \frac{1}{4m} \sigma^* C \sigma,$$

where $C = (A_{uv} - \frac{k_u k_v}{2m})_{u,v}$ is the *modularity matrix*. If we write $\sigma = \sum_u \alpha_u \mathbf{v}_u$, with $\{\mathbf{v}_u\}_u$ the orthonormal eigenvectors with corresponding eigenvalues $\{\gamma_u\}_u$ of the symmetric matrix C (ordered such that $\gamma_1 \geq \gamma_2 \geq \dots$), then

$$Q = \frac{1}{4m} \sum_u (\mathbf{v}_u^* \sigma)^2 \gamma_u.$$

As a first step in maximizing Q as a function of σ , we note that $(\mathbf{v}_1^* \sigma)^2 \gamma_1$ should be one of the dominant factors in the sum, since γ_1 is the largest eigenvalue. Ideally, we would like to choose σ parallel with \mathbf{v}_1 , however this violates the restriction that $\sigma_u \in \{1, \dots, q\}$. Instead we put σ_u equal to the sign of the corresponding element in \mathbf{v}_1 . This thus leads to a simple *spectral* algorithm: for a given network, calculate the leading eigenvector of its corresponding modularity matrix and label vertices according to the sign of elements in this vector.

We refer the reader to Section 2 for an asymptotic characterization of the spectrum of C .

Another method aiming to maximize the modularity is the Louvain Method [14]. The algorithm consists of two phases that are repeated iteratively. The first phase is initialized by putting every vertex in a unique community. Then, iteratively and subsequently, for every node i , and each of its neighbours j , we calculate the change in modularity Δ_{ij} when i would be put in the same community as j . If $\Delta_{ij} \leq 0$ for all neighbours j , we do not change the label of vertex i , otherwise we put i in the community of vertex j that maximizes Δ_{ij} (ties broken uniformly at random). We then move to a new vertex, and so on (possibly treating the same vertex multiple times), until no further improvement can be achieved. In the second phase of the algorithm, we form a graph where the nodes are the communities formed in the previous step, and the weights of edges between nodes are equal to the sum of the weight of the edges in the two corresponding communities. After completion of this phase, we apply the procedure of the first phase to the just constructed network. We continue until there are no more changes and a maximum of modularity is attained. In this way we obtain a hierarchical collection of partitions. The height of the hierarchy depends on the number of passes and is in practice small.

1.3.3 k -means

When data consists of points that can be embedded in a vector space V equipped with a norm $\|\cdot\|$, we can use partitional clustering methods to cluster them. The number of clusters K needs to be preassigned. The objective is to minimize a certain cost-function based on the distance function. A very popular technique is k -means, where the cost-function is given by

$$\sum_{i=1}^K \sum_{\mathbf{x} \in C_i} \|\mathbf{x} - \mathbf{c}_i\|^2,$$

where $C_i \subset V$ is the subset of points in the i -th cluster and $\mathbf{c}_i \in C_i$ is its centroid.

An approximative solution can be found with Lloyd's algorithm [78], an iterative method for which the initial centroids need to be specified. The algorithm has as a drawback that it usually converges to a local minimum, depending on the initial centroids. A possibly remedy is to repeat the algorithm starting at many different initial centroids, which is feasible since the algorithm converges rather fast.

1.3.4 Divisive algorithms

A very popular algorithm for community detection is a divisive algorithm proposed in [101]. We give here a short summary of the latter article.

The idea is to divide iteratively the network into smaller and smaller communities by removing at each step an edge with high *betweenness*. Here, betweenness is some measure expressing how much an edge is in between different communities. An edge with high betweenness should be responsible for connecting many pairs of nodes. Further, if two communities are joint by just a few edges, those edges are expected to have a high betweenness, as many paths between the communities pass through them.

Three measures for betweenness are proposed in [101], but we repeat here only the *shortest-path betweenness*. For an edge it is defined as the number of shortest paths (over all pairs of vertices) running through it.

The proposed algorithm is then as follows: Calculate the betweenness for all edges. Remove the edge with highest betweenness-score. Repeat the previous two steps. This gives a whole collection of partitions and the partition with highest modularity (see above) is chosen as the output of the algorithm.

We remark that calculating the edge betweenness might be costly. Further, the recalculation step seems to add a lot of extra complexity to the algorithm, but turns out to be crucial for accuracy. Indeed, removing an edge might significantly change the betweenness of edges in the obtained graph. The above algorithm works well with graphs upto 10^4 nodes, and faster extensions exist.

1.3.5 Belief propagation

Consider the ordinary SBM (see below for a precise definition) on K communities where an edge is present between vertices from communities i and j with probability p_{ij} (edges are drawn independently). Denote the community-membership of the vertices by $\sigma = \{\sigma_u\}_u$. Given a realization G of this SBM, we could estimate the community-membership as

$$\arg \max_{\hat{\sigma}} \mathbb{P}(\sigma = \hat{\sigma} | G).$$

Unfortunately, apart from computational complexity, it is not guaranteed that the obtained maximizer agrees with the true partition. Indeed $\mathbb{P}(\sigma = \cdot | G)$ possibly has many local maxima and the highest might just be the best fit to the noise¹. Note however, that obtaining the distribution $\mathbb{P}(\sigma = \cdot | G)$ is the most we can learn about σ , since the observation G is the only information we have about σ (G can be considered as a noisy channel between σ and the observer).

Another estimator for σ follows if we know the marginal distribution $\{\psi_i^u\}_i$ for every vertex u , defined as

$$\psi_i^u = \mathbb{P}(\sigma_u = i | G) = \sum_{\hat{\sigma}: \hat{\sigma}_u = i} \mathbb{P}(\sigma = \hat{\sigma} | G), \quad (1.1)$$

where we need to fix the community-membership of a few vertices beforehand (otherwise the marginals are uniform). Indeed, we can then estimate $\sigma_u^* = \arg \max_i \psi_i^u$, and it turns out that this maximizes the fraction of correctly classified vertices [37, 61]. Note that with this approach we take an average over many configurations, rather than the single, best fit as above.

The cavity-method [97] or belief propagation can be used to *approximate* the marginals (1.1) if we make the *assumption* that for any vertex u the community-membership of any of its neighbours are conditionally independent given the community-membership of u . A vertex adjusts its belief of its own community-membership based on messages received from its neighbours: A vertex u sends its neighbour v a message $\psi^{u \rightarrow v} = \{\psi_i^{u \rightarrow v}\}_{i=1}^K$, which is an estimate of its own

¹If the labels $\{\sigma_u\}$ are drawn independently and uniformly from $\{1, \dots, K\}$, then $\mathbb{P}(\sigma | G) \sim \mathbb{P}(G | \sigma)$. It could be that for two almost uncorrelated configurations σ and $\hat{\sigma}$, $\mathbb{P}(G | \sigma) \simeq \mathbb{P}(G | \hat{\sigma})$ are both large. But why would we prefer one configuration over the other?

community-membership in case the edge between u and v were not present². Vertex v then uses this message to update its own belief. The messages (if we ignore the weak-interactions due to the absence of edges) are given by

$$\psi_i^{u \rightarrow v} = \frac{1}{Z} \prod_{w \sim u, w \neq v} \sum_{j=1}^K \psi_j^{w \rightarrow u} p_{ij}, \quad (1.2)$$

where Z is a normalization constant such that $\sum_i \psi_i^{u \rightarrow v} = 1$.

The belief propagation algorithm consists of randomly initializing the messages and updating them according to (1.2) until a fixed point is reached, from which we simply calculate the marginals by summing over all incoming messages to a vertex.

We note that belief propagation is exact on trees. Further, the message sent from u to v does *not* depend on the message from v to u , preventing resonance to occur. Compare this to spectral methods based on the adjacency matrix A , where a leading eigenvector x with eigenvalue λ concentrates on high-degreed vertices due to the resonance in the equation $\sum_v A_{uv} x_v = \lambda x_u$.

In case all vertices have the same expected degree in the SBM, that is $\sum_j p_{ij} = \rho$, for some constant ρ , belief propagation has at least one fixed point, namely $\psi_i^{u \rightarrow v} = \frac{1}{K}$. Reconstruction better than random guess is then impossible if this is the *only* fixed point.

In general, whether we can do better than random guess, depends on the amount of fixed points, their stability and region of attraction. The work [34, 35] uses a fixed point analysis to conjecture phase-transition phenomena in the SBM, see for instance Conjecture 3.2.1 below.

A very readable introduction on some methods from statistical physics (including belief propagation) applied to community detection is [91]. We used the latter article as inspiration for this section. The first work in which belief propagation was applied to community-detection is [53].

1.4 Real datasets and benchmark graphs

We start by listing some common networks, used often to test algorithms.

The first network is known as Zachary's karate club, consisting of 34 club members who split into two separate groups when the instructor left the club after a conflict with the president, see [124].

A similar phenomena is observed in a network of 62 dolphins living in Doubtful Sound (New Zealand). The dolphins are represented as vertices and edges between two of them are present when they met each other more often than expected by chance. The group split into clusters after one of the dolphins left, see [81].

The political blogosphere is a dataset of over 1000 political blogs relating to one another, captured during a period of two months preceding the 2004 US presidential elections. The blogs fall naturally into two categories: democrats and republicans. Detecting their political orientation based solely on the links between blogs is an interesting task for community detection algorithms. See [3].

Many more data has been collected, for instance a co-authorship graph in network science, a graph of astrophysics collaborations, a neural network and even a graph linking the main characters in Les Misérables by Victor Hugo. See for a non-exhaustive overview the Stanford Large Network Dataset Collection at <https://snap.stanford.edu/data/>.

In addition to testing algorithms on real networks, we can also artificially generate them. The study of random graphs was initiated by Erdős and Rényi in [40]. The so-called Erdős-Rényi graph with parameters n and p is a random graph on n vertices where edges are independently present between pairs of vertices with probability p . In this graph, the degrees of vertices are asymptotically Poisson random variables

²more precisely, if the presence of an edge between u and v were *unknown*.

with parameter np , and hence the degree-sequence is homogeneous. An extension that allows for inhomogeneous degree-sequence is the inhomogeneous Erdős-Rényi graph [15], with parameters n and $\{p_{uv}\}_{uv}$: an edge is independently present between vertices u and v with probability p_{uv} .

A particular case of [15], where the graph includes K communities is obtained by giving each vertex u a label $\sigma_u \in \{1, \dots, K\}$, and each probability p_{uv} is a function of the labels σ_u and σ_v . In this way, the ground truth is known, so that we can test algorithms by running them on these benchmark graphs to see if the hidden labels can be retrieved.

The Stochastic Block Model (SBM) [57], for instance, is defined as the random graph where vertices u and v connect with a probability depending only on σ_u and σ_v , i.e., $p_{uv} = B_{\sigma_u \sigma_v}$, for some $K \times K$ positive symmetric matrix B . This model is able to generate a diverse collection of random graphs, while it stays analytically tractable. In practice however, the SBM fails to accurately describe observed data: due to the stochastic inidentifiability of nodes in the same community, it does not allow for degree heterogeneity within blocks. We therefore look at a more general model, the Degree-Corrected Stochastic Block Model (DC-SBM) [69], where the vertices additionally have *weights*. Edges are drawn between vertices with a probability depending on both the weights and the community-membership of the vertices.

We distinguish between exact-, quasi- and weak (or partial) reconstruction: Exact recovery means that an algorithm correctly reconstructs σ_u for all u . Quasi recovery means that only a vanishing fraction of the vertices are misclassified. Weak recovery means that an algorithm does strictly better than random guess, and is also called detection. In particular, for a graph with two equal-sized communities, we speak about weak recovery if an algorithm correctly classifies a fraction of $1/2 + \epsilon$ of the vertices, with $\epsilon > 0$.

Chapter 2

Random matrices

We begin this section with the most celebrated result in random matrix theory: Wigner's semicircle law, motivated by problems from nuclear physics. We state the theorem here for sequences of Hermitian random matrices $(X_N)_{N \in \mathbb{N}}$, where for $N \in \mathbb{N}$,

$$X_N = \frac{1}{\sqrt{N}}(X_{uv})_{u,v=1}^N,$$

with the following conditions on its entries: The entries $(X_{uv})_{1 \leq u < v \leq N}$ are i.i.d. complex (or real) random variables with finite variance σ^2 with their law independent of N . The diagonal entries $(X_{uu})_{1 \leq u \leq N}$ are also i.i.d. random variables with finite variance, but are in addition real-valued and have zero mean. We then have the following spectral statistics, expressed in terms of the empirical spectral distribution

$$\mu_N = \frac{1}{N} \sum_{u=1}^N \delta_{\lambda_u^N},$$

where $\lambda_1^N \geq \lambda_2^N \geq \dots \geq \lambda_N^N$ are the ordered eigenvalues of X_N and $\delta_{\lambda_u^N} : x \mapsto 1_{\lambda_u^N \leq x}$.

Theorem 2.0.1 (Wigner [119, 120]). *Almost surely,*

$$\lim_{N \rightarrow \infty} \mu_N = \sigma_{sc},$$

with σ_{sc} the semi-circle distribution, with respect to the Lebesgue measure defined as

$$\frac{d\sigma_{sc}}{dx} = \frac{1}{2\pi\sigma^2} \sqrt{4\sigma^2 - x^2} \mathbf{1}_{[-2\sigma, 2\sigma]}(x).$$

Note that this result is *universal*, in the sense that the convergence to the *same* semicircle-law takes place independently of the underlying distribution of the elements. In Wigner's original paper [119], he proved the theorem for Bernoulli random variables and extended these results later for general Hermitian matrices in [120]. There are more universality results known in random-matrix theory where the asymptotic behaviour of random matrices is independent of the exact distribution of its entries. We also note that the semi-circle law puts no mass outside the interval $[-2\sigma, 2\sigma]$, i.e., it has *sharp edges*. This implies that with probability $1 - o(1)$ all eigenvalues of X_N , except $o(n)$ lie inside this interval. The behaviour of possible outliers is characterized in [46]: Consider for any $c > 2\sigma$, the interval $I = [-c, c]$. If $\mu := \mathbb{E}[X_{12}] = 0$, then *all* eigenvalues lie inside I with probability $1 - o(1)$. Otherwise, if $\mu > 0$, then only the largest eigenvalue lies outside I and it asymptotically has normal distribution with mean $\sqrt{n}\mu + \sigma^2/(\sqrt{n}\mu)$ and variance

$2\sigma^2/n$. This is an extremely important observation having far-reaching implications for practical applications. **Indeed, the presence of one or more eigenvalues clearly separated from the bulk of eigenvalues of a random matrix could indicate non-random behaviour, for instance an underlying community structure.**

Two methods used to prove Theorem 2.0.1 are the moment method and the Stieltjes transform method. Wigner used the former method, which consists in showing that the moments of the empirical distribution, $\mathbb{E}[\int x^k d\mu_N(x)]$, converge to the moments of the semicircle-law, $\int x^k d\sigma_{sc}$, for large dimension N . Now, since $\text{Tr}\{X_N^k\} = \lambda_1^k + \dots + \lambda_N^k$,

$$\mathbb{E}\left[\int x^k d\mu_N(x)\right] = \frac{1}{N}\mathbb{E}\left[\text{Tr}\{X_N^k\}\right] = \frac{1}{N^{1+k/2}} \sum_{i_1, \dots, i_k=1}^N \mathbb{E}[X_{i_1 i_2} \dots X_{i_{k-1} i_k} X_{i_k i_1}].$$

This gives rise to a combinatorial analysis of the right-hand side. This so-called trace method is often attributed to Füredi and Komlós, after their extension of Wigner’s results for non-centred diagonal entries in [46].

The Stieltjes transform method consists of showing that the Stieltjes transform of μ_N ,

$$s_N : z \mapsto \int_{\mathbb{R}} \frac{1}{x-z} d\mu_N(x) = \frac{1}{N} \text{Tr}\{(X_N - zI)^{-1}\},$$

converges almost surely to the Stieltjes transform of the semicircle-law, $s_{sc} : z \mapsto \int_{\mathbb{R}} \frac{1}{x-z} d\sigma_{sc}$. Analogous to the scalar case, where moment generating functions are used to ”group together” the moments of the scalar variable, the Stieltjes transform method considers all moments of the matrix simultaneously, rather than one at a time as above.

The Stieltjes transform method is used in [98] to *heuristically* characterize the spectrum of both the adjacency A and modularity matrix C (see Section 1.3.2) for random instances of the SBM. The paper thus investigates the limits of spectral methods based on those two matrices. We summarize their result and approach. In particular, the authors of [98] consider the sparse case of the SBM in its most simple form: the vertices are grouped in two clusters of size $n/2$ and edges are present independently with probability a/n within and probability b/n between clusters. See Section 3 for a more complete definition of the SBM.

Write

$$A = \mathbb{E}[A] + A - \mathbb{E}[A],$$

and

$$\mathbb{E}[A] = \frac{a+b}{2} \mathbf{1}\mathbf{1}^T + \frac{a-b}{2} \mathbf{u}\mathbf{u}^T,$$

with $\mathbf{1} = \frac{1}{\sqrt{n}}(1, \dots, 1)$ and $\mathbf{u} = \frac{1}{\sqrt{n}}(1, 1, \dots, -1, -1, \dots)$, where the sign indicates community-membership.

The authors of [98] calculate first the density of $A - \mathbb{E}[A]$ to be

$$\rho_{A-\mathbb{E}[A]} : x \mapsto \frac{n}{\pi} \frac{\sqrt{2(a+b) - x^2}}{a+b},$$

i.e., a modified semicircle-law. Now, $C = \frac{a-b}{2} \mathbf{u}\mathbf{u}^T + A - \mathbb{E}[A]$, i.e., a rank 1 matrix plus a centred perturbation. It turns out that the spectrum of C is asymptotically equal to the spectrum of $A - \mathbb{E}[A]$, except for the eigenvalue $\lambda_C = \frac{a-b}{2} + \frac{a+b}{a-b}$. Similarly, recognizing A as a rank 1 matrix plus C , its spectrum is seen to consist of a semicircular band (possibly plus outliers) plus two eigenvalues, namely λ_C and $\frac{a+b}{2} + 1$. Caution is needed here, as these results (on the density) do not exclude

outliers to be present. In any case, the informative eigenvalue of C is not separated from the bulk when $(a - b)^2 \leq 2(a + b)$.

Note the consequences: If $a = b$, then no community structure is present. However, if $a > b$, but $(a - b)^2 < 2(a + b)$, then the graph contains two communities, but we are unable, using A or C , to detect those. As we will shortly see, no method is able to detect those communities from observations of the graph as $(a - b)^2 > 2(a + b)$ is in fact an information-theoretic threshold with respect to community detection [92].

Chapter 3

The Degree-Corrected Stochastic Block Model

We describe here a suitably normalised version of the Degree-Corrected Stochastic Block Model (DC-SBM) as given in [69]. The DC-SBM is a generalization of the Erdős-Rényi classical model of random graphs. We will consider this model in two regimes, namely a *dense* regime where the degrees grow sufficiently fast with the size of the network and a *sparse* regime where the expected degrees remain constant in the limit of large networks. We assume the number of communities to be arbitrary (i.e., ≥ 2), but constant. Note that the setting with two communities is also called the Extended Planted-Partition Model (EPPM) in [21].

We start with the *dense* regime. Consider a random graph on the set of vertices $V := \{1, \dots, n\}$, partitioned into K communities of $\alpha_k n$ members each: each vertex u is given a label $\sigma_u \in S := \{1, \dots, K\}$. A weight D_u is given to each vertex u to encode its expected degree. Without loss of generality we assume that $D_1 \leq D_2 \leq \dots \leq D_n$. All weights and labels will depend on n , but this is suppressed in the notation here. For each pair (u, v) , we include the edge (u, v) with probability

$$\mathbb{P}(u \sim v) = \begin{cases} \frac{D_u D_v}{n \bar{D}} B_{\sigma_u \sigma_v} & \text{if } u \neq v \\ 0 & \text{if } u = v, \end{cases} \quad (3.1)$$

where $B \in (\mathbb{R}^+)^{K \times K}$ is a symmetric matrix, independent of n and $\bar{D} = 1/n \sum_{l=1}^n D_l$, the average weight. B may be chosen *completely* independent of the weights $\{D_u\}_{u=1}^n$: all information about the community-structure is then captured by B alone.

We make some further assumptions on the parameters of the model: For (3.1) to define a probability, we assume

$$\frac{D_u D_v}{n \bar{D}} B_{\sigma_u \sigma_v} \leq 1, \quad (3.2)$$

for all u, v .

The vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$ is assumed to be constant. Hence, the clusters are well balanced, as the size of each community grows linearly with n . Further, the average weight in a cluster,

$$\bar{D}_i = \frac{1}{\alpha_i n} \sum_{u=1}^n D_u \mathbf{1}_{\sigma_u=i},$$

is assumed to be asymptotically a fraction of the average weight \bar{D} . That is, we assume that there exists non-zero constants d_1, \dots, d_K , such that,

$$\lim_{n \rightarrow \infty} \frac{\bar{D}_i}{\bar{D}} = d_i, \quad (3.3)$$

for all i . In the dense regime, we assume that $\bar{D} \rightarrow \infty$ when $n \rightarrow \infty$. We denote the DC-SBM in this regime by $\mathcal{G}(B, K, \{\sigma_u\}_{u=1}^n, \{D_u\}_{u=1}^n)$.

As an example, we let $\{\sigma_u\}_{u=1}^n$ be any sequence such that $n/2$ of its elements are 1 and the other $n/2$ elements are 2. Then, there are two equally-sized communities: $K = 2$ and $\alpha_1 = \alpha_2 = 1/2$. Let $\{D_u\}_{u=1}^n$ be any non-decreasing sequence with $D_1 > 0$. Put

$$B = \begin{pmatrix} a & b \\ b & a \end{pmatrix},$$

for some constants a and b . Then

$$\mathbb{P}(u \sim v) = \frac{D_u D_v}{n \bar{D}} \begin{cases} a & \text{if } \sigma_u = \sigma_v, \\ b & \text{otherwise.} \end{cases} \quad (3.4)$$

This is exactly the EPPM in [21], as mentioned above.

We now turn to the *sparse* regime of the DC-SBM defined in terms of two positive *constant* parameters a and b : We consider again a random graph on n vertices partitioned into $q \geq 2$ asymptotically equal-sized clusters by giving each vertex v a spin σ_v drawn uniformly from $\{1, \dots, q\}$. The vertices have i.i.d. weights $\{\phi_u\}_{u=1}^n$ governed by some law ν with support in $W \subset [\phi_{\min}, \infty)$, where $0 < \phi_{\min} < \infty$ is a constant *independent of* n . We denote the k -th moment of the weights by $\Phi^{(k)}$, i.e., $\Phi^{(k)} = \int_W x^k d\nu(x)$. Below we state more precise tail-conditions on the weights. An edge is drawn between nodes u and v with probability

$$\mathbb{P}(u \sim v) = \frac{\phi_u \phi_v}{n} \begin{cases} a & \text{if } \sigma_u = \sigma_v, \\ b & \text{otherwise.} \end{cases} \quad (3.5)$$

Note that by the law of large numbers the size of the communities are indeed $\frac{n}{q} + \mathcal{O}(\sqrt{n})$.

We are interested in answering the following question: can we recover the underlying community structure from a single observation of the DC-SBM? We restrict ourselves to the case where we only have topological information, in particular we do not have access to the weights, and other parameters.

A particularly important role for answering this question in the sparse regime will be played by the non-backtracking matrix B . For a given graph $G = (V, E)$, the non-backtracking matrix is indexed by the set of oriented edges $\vec{E} = \{(u, v) : \{u, v\} \in E\}$. For $e = (e_1, e_2), f = (f_1, f_2) \in \vec{E}$, B is defined as

$$B_{ef} = 1_{e_2=f_1} 1_{e_1 \neq f_2}.$$

This matrix was introduced by Hashimoto [60] in 1989. Note that B is non-symmetric and even non-normal, which makes it harder to analyse.

In the dense regime we analyse the normalized adjacency matrix $\hat{H} = \hat{D}^{-1} A \hat{D}^{-1}$, where A is the empirical adjacency matrix and \hat{D} the diagonal matrix containing the degrees.

3.1 Dense regime

3.1.1 Literature on the *ordinary* PPM

The SBM is a special case of the DC-SBM in which all vertices have the same unit weight: $D_u = 1$ for all u . The SBM on two equal-sized communities is known in the computer science community as the Planted Partition Model (PPM). In this model, vertices in the same community connect with probability $p_{\text{in}} := a/n$ and between communities with probability $p_{\text{out}} := b/n$, where we allow a, b to vary (increase) with n .

Standard spectral methods to recovery communities in this model are based on the Adjacency and Laplacian matrices, see the introduction.

Recovering the planted partition (without degree-corrections) often coincides with finding the minimum bisection in the same graph. That is, finding a partition of the graph such that the number of edges between separated components (the *bisection width*) is small. This problem is NP-hard [39].

Graph bisection on random graphs has been studied intensively. For instance, [20] studies the collection of labelled simple random graphs that have $2n$ nodes, node-degree $d \geq 3$ and bisection width $o(n^{1-1/\lfloor (d+1)/2 \rfloor})$. For these graphs the minimum bisection is much smaller than the average bisection, which makes it easier to find. The main result is a polynomial-time algorithm based on the maxflow-mincut theorem, that finds exactly the minimum bisection for almost all graphs.

The work [39] shows that, if in the planted partition model $p_{\text{in}} > p_{\text{out}}$ are fixed (for $n \rightarrow \infty$), then the underlying community structure coincides with the minimum bisection and it can be retrieved in polynomial time. This result is improved in [63].

In [86] the case of non-constant p_{in} and p_{out} is analysed: using a spectral algorithm, one can recover the communities with probability $1 - \delta$ if $\frac{p_{\text{in}} - p_{\text{out}}}{\sqrt{p_{\text{in}}}} = \Omega\left(\sqrt{\frac{\log(n/\delta)}{n}}\right)$.

We refer the reader for a detailed historical overview to the table on page 3 in [1]. The latter work gives precise conditions for *exact* recovery in the dense PPM ($p_{\text{in}} = a \log(n)/n$ and $p_{\text{out}} = b \log(n)/n$, with a, b constants): exact recovery is possible precisely when $\frac{a+b}{2} - \sqrt{ab} > 1$. Compare this to the phase-transition in dense Erdős-Rényi graphs with edge probability $c \log(n)/n$ (c a constant), which are connected if and only if $c > 1$.

3.1.2 Literature on the DC-SBM

Positive results of spectral clustering in the DC-SBM have been obtained by various authors. The work [32] introduces a reconstruction algorithm based on the matrix that is obtained by dividing each element of the adjacency matrix by the geometric mean of its row and column degrees.

The main result in [28] is a polynomial time algorithm that outputs a partitioning that differs from the planted clusters on no more than $n \log(D)/D^{0.98}$ nodes. This recovery succeeds only under certain conditions: the minimum weight should be a fraction of the average weight and the degree of each vertex is $o(n)$.

The article [76] gives an algorithm based on the adjacency matrix of a graph together with performance guarantees. The average degree should be at least of order $\log(n)$. However, since the spectrum of the adjacency matrix is dominated by the top eigenvalues [25], the algorithm does a poor job when the degree-sequence is very irregular.

In Section 4.4, we give a more detailed literature review on spectral methods using the Adjacency and Laplacian matrix, but also on methods as Spectral Clustering on Ratios of Eigenvectors (SCORE) and Regularized Spectral Clustering.

3.1.3 Results in this thesis

We propose in Chapter 4 a spectral clustering algorithm based on a suitably normalized adjacency matrix \hat{H} (defined above). We show that this algorithm consistently recovers the block-membership of all but a vanishing fraction of nodes, in the regime where the lowest degree is of order $\log(n)$ or higher. Recovery succeeds even for very heterogeneous degree-distributions. The used algorithm does not rely on parameters as input. In particular, it does not need to know the number of communities.

An important ingredient in our proof is Lemma A.0.5, which serves as an alternative to the commonly used Davis-Kahan theorem.

We note that \widehat{H} is a natural choice in the degree-corrected setting. Indeed, due to (3.1), we expect that $\mathbb{E}[\widehat{H}_{uv}] \sim B_{\sigma_u\sigma_v}$ up to some scaling. This is in contrast with the normalized Laplacian L and the Adjacency matrix, because $\mathbb{E}[L_{uv}] \sim \sqrt{D_u}\sqrt{D_v}B_{\sigma_u\sigma_v}$ and $\mathbb{E}[A_{uv}] \sim D_uD_vB_{\sigma_u\sigma_v}$. We elaborate on this in Section 4.4, where we also compare \widehat{H} with other spectral methods.

3.2 Sparse regime

The sparse regime is by far the most realistic regime as the expected degree of a node does not grow with the network size. At Facebook for instance, your number of friends is unlikely to increase significantly when new members join. At the same time, this regime is more challenging. Indeed, traditional methods based on the Adjacency or Laplacian matrix working well in the dense case break down when employed in the sparse case. Intuitively, the hardness can be understood as follows: When $a = b$, the graph has no communities. Therefore, if $|a - b|$ is non-zero but very small, we expect that a graph where each vertex has only a small, constant number of neighbours, does not contain enough information to distinguish between the clusters.

3.2.1 Literature on the *ordinary* SBM

We turn first to the *ordinary* SBM, where all vertices have unit weight (*i.e.*, $\phi_u = 1$ for all u). The authors of [36] were the first to conjecture a phase-transition for this setting based on ideas from statistical physics:

Conjecture 3.2.1 ([36]). *Consider a SBM on q balanced communities where edges inside a cluster are present with probability a/n and between clusters with probability b/n . Let M be the matrix with a/q on the diagonal and b/q on all off-diagonal elements. Let λ_1 and λ_2 be its first, respectively, second eigenvalue and let $\text{SNR} = \frac{\lambda_2^2}{\lambda_1} = \frac{(a-b)^2}{q(a+(q-1)b)}$, the signal-to-noise-ratio.*

For any $q \geq 2$, if $\text{SNR} > 1$ (which is generally called the Kesten-Stigum condition), communities can be detected in polynomial time.

For $q \geq 4$, it is theoretically possible to detect communities for some $\text{SNR} < 1$.

It is believed that for $q \geq 4$, a double phase-transition occurs: Detection should be easy (*i.e.*, polynomial time) when $\text{SNR} > 1$, much harder (*i.e.*, exponential time) for $\text{SNR} \in (\tau, 1]$, for some $0 < \tau < 1$, and information-theoretically impossible when $\text{SNR} < \tau$.

The conjecture has been settled in the case of two communities: First in [85] by using a matrix counting the number of self-avoiding paths in the graph, and later, independently, in [94]. Further, [92] shows that for $q = 2$, it is information-theoretically impossible to detect communities for SNR below 1.

In [72] the 'spectral redemption conjecture' was made: detection using the second eigenvalue of the non-backtracking matrix (defined above) would also establish the positive part. This has recently been proved¹ in [17], for any $q \geq 2$ such that λ_q is a simple eigenvalue of M .

More recently, [2] gave an algorithm that detects communities when $\text{SNR} > 1$.

In particular, for the SBM with two asymptotically equal-sized communities we have the following phase-transition result: **When $(a - b)^2 \leq 2(a + b)$, community detection is information-theoretically impossible. Conversely, when $(a - b)^2 > 2(a + b)$, a positively-correlated clustering can be obtain by thresholding the second eigenvalue of the non-backtracking matrix.**

¹Theorems 4 and 5 in [17] are actually a bit more general.

3.2.2 Results in this thesis

Consider the DC-SBM with q asymptotically equal-sized communities. Does this model exhibit a similar behaviour? In particular, does there exist a threshold (in terms of a, b and statistics on ϕ_u) below which community detection is information-theoretically impossible? And above this threshold, can we use again the non-backtracking matrix or do we need to modify it? *A priori especially the latter is unclear, because an algorithm solely based on the non-backtracking matrix cannot use any information on the weights as input.*

Our main results show the following: Firstly, in the setting of $q \geq 2$ communities where the weights are possibly heavy-tailed (with large enough exponent), detection is impossible when $(a - b)^2 \Phi^{(2)} \leq q(a + b)$. To derive this result, we establish a coupling between local neighbourhoods in the graph and multi-type branching processes where the offspring distribution follows a Poisson-mixture (due to the weights). It is crucial that the weights in the graph and the branching process *exactly* coincide. Further, we show that long-range interactions are weak in this model, even if the degrees follow a power-law.

Secondly, we characterize the spectrum of the non-backtracking matrix for DC-SBM's *with bounded weights* on two communities in the regime where the number of vertices is large. In particular, the leading eigenvalue is asymptotically equal to $\rho := \frac{a+b}{2} \Phi^{(2)}$. For the second eigenvalue, we distinguish two regimes: the second eigenvalue is asymptotically equal to $\mu_2 := \frac{a-b}{2} \Phi^{(2)}$ when $\mu_2^2 > \rho$, whereas the second eigenvalue is asymptotically bounded by $\sqrt{\rho}$ when $\mu_2^2 \leq \rho$. All the remaining eigenvalues are asymptotically bounded by $\sqrt{\rho}$. As a consequence, the spectral method based on the non-backtracking matrix successfully detects communities in the regime $(a - b)^2 \Phi^{(2)} > 2(a + b)$. As a corollary we obtain that degree-corrected Degree-Corrected Erdős-Rényi graphs are asymptotically Ramanujan in that they asymptotically satisfy the graph Riemann hypothesis. To derive these results, we carefully analyse multi-type weighted branching processes, leading to an extension of results by Kesten and Stigum [70, 71]. We further derive a weak-law of large numbers for local functionals on DC-SBM's using the coupling-result mentioned above. To derive the bound on the bulk of the spectrum, we decompose powers of the non-backtracking matrix as a sum of products and then bound the individual terms with the trace method.

Surprisingly, no modification of the matrix, nor information about the weights is thus needed, which shows the robustness of the method. We conclude that, as in the standard SBM, the algorithm is optimal in the sense that it works all the way down to the detectability-threshold.

Note that the only relevant statistic of the *random* weight is its second moment. The detectability threshold can be expressed as $\mu_2^2 > \rho$. Both ρ and μ_2 are asymptotically eigenvalues of the expected adjacency matrix conditioned on the weights. Indeed, if A denotes the adjacency matrix, and if ψ_1 and ψ_2 are the vectors defined for $u \in V$ by $\psi_1(u) = \frac{1}{\sqrt{2}} \phi_u$ and $\psi_2(u) = \frac{1}{\sqrt{2}} \sigma_u \phi_u$, then

$$\mathbb{E}[A|\phi_1, \dots, \phi_n] = \frac{a+b}{n} \psi_1 \psi_1^* + \frac{a-b}{n} \psi_2 \psi_2^* - a \frac{1}{n} \text{diag}\{\phi_u^2\}.$$

Put $\hat{\psi}_i = \frac{\psi_i}{\|\psi_i\|_2}$. Then, by the law of large numbers, for $i = 1, 2$,

$$\left\| \mathbb{E}[A|\phi_i, \dots, \phi_n] \hat{\psi}_i - \mu_i \hat{\psi}_i \right\|_2 \rightarrow 0,$$

in probability, as n tends to ∞ .

Chapter 4

A spectral algorithm for the *dense* DC-SBM

We have published parts of this chapter in Advances in Applied Probability [52].

4.1 Introduction

Here we deal with the DC-SBM in its *dense* regime, where the weights grow logarithmically with the size of the network. This ensures that the node-degrees concentrate around their expectations. Consequently, instances of the DC-SBM "look as expected". This facilitates considerably the reconstruction of the community-structure, as the top eigenvectors of carefully chosen matrices associated to the graph should reveal the structure in the expected graph. Indeed, in the *ordinary* SBM, the Adjacency matrix successfully recovers the communities (see Section 3.1.1). However, as we detail below, spectral methods using the Adjacency matrix break down once the degree-sequence becomes too irregular - which is the case in the DC-SBM. We therefore propose here an alternative method where the Adjacency matrix is suitably *normalized*.

The contribution in this chapter is as follows: We demonstrate with a clean analysis that community detection in a moderately sparse DC-SBM is feasible under rather general conditions on the degree-sequence.

The algorithm we propose below uses as input the *normalized adjacency matrix* \widehat{H} with entry (u, v) given by

$$\widehat{H}_{uv} = \begin{cases} \frac{1}{\widehat{D}_u \widehat{D}_v} A_{uv} & \text{if } A_{uv} = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (4.1)$$

Here A is the adjacency matrix of the graph and \widehat{D}_u is the observed degree of vertex u . Note that this matrix is a **natural choice** in the DC-SBM: Indeed, in the model specified in Section 3, vertices u and v (in communities σ_u and σ_v respectively) are connected by an edge with a probability $\sim D_u D_v B_{\sigma_u \sigma_v}$, where B is a symmetric connectivity matrix. After division by the observed degrees we are left with $B_{\sigma_u \sigma_v}$ times a constant (asymptotically depending only on σ_u and σ_v).

We show that \widehat{H} concentrates around a deterministic matrix P of rank L not larger than the number of communities K , when the minimum expected degree is as small as $\log(n)$. To establish this concentration-result, we use Lemma A.0.5 given in the appendix, which could be of independent interest, as a *simple alternative to the commonly used Davis-Kahan theorem*.

Due to the underlying community structure, the matrix that has the first L eigenvectors of P as its columns has the nice property that it has only K different

rows. Hence, due to this fact and the concentration of \widehat{H} around P , the rows of the corresponding eigenvector matrix of \widehat{H} considered as points in an L -dimensional euclidean space, must cluster around K centres. This property indicates that \widehat{H} is the right matrix to analyse when dealing with the DC-SBM. Indeed, associating each vertex with its corresponding row, we show in this chapter that we retrieve the correct community of all but a vanishing fraction of nodes.

Along the way, we point out a natural connection between \widehat{H} and a random walk on the observed graph.

The organization of this chapter is as follows: First we state the necessary conditions on the model. Next we state our main result for community detection in this model, followed by a comparison to other methods (Section 4.4) and a discussion (Section 4.5): the conditions in the main theorem and a connection between \widehat{H} and random walks. Section 4.6 outlines the approach we take to prove the main theorem, which is accompanied by a statement of all auxiliary lemmas. All proofs are deferred to Section 4.7. In the last section we give a suggestion for future research.

4.2 Conditions on the model

Recall the definition of the DC-SBM in its dense regime from Section 3. From assumptions (3.2) and (3.3) on the (average) weight (in a cluster), the following limit exists for all i ,

$$\overline{M}_i = \lim_{n \rightarrow \infty} \frac{M_i}{\sum_{l=1}^n D_l} = \sum_{k=1}^K B_{ik} \alpha_k d_k, \quad (4.2)$$

where $M_i = \sum_l D_l B_{i\sigma_l}$.

If the connection probabilities for two communities are approximately the same, then recovery might still fail. For the *ordinary* SBM with two communities those conditions are precisely characterized in [1]. In the DC-SBM such conditions necessarily depend on statistics of the degree-sequence. We shall see in Section 4.5.2 that we need the following *identifiability conditions*: We assume that for all i, l there exists j such that

$$\frac{B_{ij}}{\overline{M}_i} \neq \frac{B_{lj}}{\overline{M}_l}. \quad (4.3)$$

In the analysis that follows, we will consider the random graph in a dense regime, that is we assume: either

$$\lim_{n \rightarrow \infty} \frac{D_1}{\log(n)} = \infty, \quad (4.4)$$

or, for some constant $c < 1/2$,

$$D_1 \geq C_{B, \mathbf{M}} \cdot \log(n) \text{ and } \lim_{n \rightarrow \infty} \frac{D_n^2}{n^c} \rightarrow 0, \quad (4.5)$$

where $C_{B, \mathbf{M}}$ is some constant depending on B , $\mathbf{M} = (\overline{M}_1, \dots, \overline{M}_K)$ and the convergence rate in (4.2). Further, we assume the following condition on the weights:

$$\frac{D_1^2}{D} = \Omega(\log(n)). \quad (4.6)$$

Note that under those assumptions, D_u represents the expected degree of vertex u upto a multiplicative factor that depends only on the community σ_u . Indeed, if \widehat{D}_u denotes the observed degree of vertex u , then

$$\mathbb{E} \left[\widehat{D}_u \right] = \frac{D_u}{nD} \sum_{l \neq u} D_l B_{\sigma_u \sigma_l} = \frac{D_u}{nD} (M_{\sigma_u} - D_u B_{\sigma_u \sigma_u}) = D_u \overline{M}_{\sigma_u} (1 + \epsilon_n), \quad (4.7)$$

where $\epsilon_n \leq \frac{2}{M_{\sigma_u} D_u} = o_n(1)$.

4.3 Main results

As is the case for *ordinary* SBM's, one can hope for (quasi-) exact recovery¹. Note that [1] shows how one could obtain exact recovery from partial recovery. We therefore focus here on *quasi*-exact recovery.

Our aim is to retrieve the underlying community structure from a single observation of the random graph. We do this by analysing the spectral properties of $\widehat{H} \in \mathbb{R}^{n \times n}$. We shall demonstrate that this matrix is close (in a sense to be specified below) to the matrix P defined for (u, v) as

$$P_{uv} = \frac{1}{n\overline{D}} \frac{B_{\sigma_u \sigma_v}}{M_{\sigma_u} M_{\sigma_v}}. \quad (4.8)$$

Denote the rank of P by L . Due to the community structure, $L \leq K$ (see below for details).

In the regime where (4.4) holds, let f be any function tending to zero, such that

$$f(n) \gg \frac{1}{\widehat{D}_1} + \frac{1}{\sqrt{\log(n)}} + \sqrt{\frac{\log(n)}{\widehat{D}_1}}.$$

For the regime where (4.5) holds, let f be tending to zero in such a way that

$$f(n) \gg \frac{1}{\widehat{D}_1} + \frac{1}{\sqrt{\log(n)}} + \frac{1}{\log^{1/3}(n)}.$$

Further, let $\tau(n) = 1/f(n)^{1/3}$. Algorithm 1 uses \widehat{H} to reconstruct the communities.

Algorithm 1

1. Calculate the average degree in the graph, call it $\widehat{D}_{\text{average}}$. Let \widehat{L} be the number of eigenvalues of \widehat{H} that are in absolute value larger than $f(n)/\widehat{D}_{\text{average}}$.
2. Compute the first \widehat{L} orthonormal eigenvectors of \widehat{H} ordered according to their absolute eigenvalues. Denote these eigenvectors and their corresponding eigenvalues by $\widehat{x}_1, \dots, \widehat{x}_{\widehat{L}}$ and $\widehat{\lambda}_1, \dots, \widehat{\lambda}_{\widehat{L}}$ respectively.
3. Associate to each node $u \in V$ the vector

$$\widehat{z}_u = (\widehat{x}_1(u), \dots, \widehat{x}_{\widehat{L}}(u)). \quad (4.9)$$

Cluster the vectors $(\widehat{z}_u)_{u=1}^n$ as follows: Pick $\tau(n)$ pairs of vertices, label them $(u(1), u'(1)), \dots, (u(\tau(n)), u'(\tau(n)))$. Calculate $\delta(t) = \sqrt{n} \|\widehat{z}_{u(t)} - \widehat{z}_{u'(t)}\|$, and $\epsilon = \min_{t: \delta(t) > f^{2/3}(n)} \delta(t)$. Find a vertex m so that $\{u' : \sqrt{n} \|\widehat{z}_m - \widehat{z}_{u'}\| \leq \epsilon/8\}$ has cardinality larger than $f^{1/3}(n) n$. Form a community consisting of all nodes in $\{u' : \sqrt{n} \|\widehat{z}_m - \widehat{z}_{u'}\| \leq \epsilon/4\}$. Remove those nodes and iterate this procedure.

We have:

Theorem 4.3.1. *Consider a DC-SBM $\mathcal{G}(B, \alpha, \{\sigma_u\}_{u=1}^n, \{D_u\}_{u=1}^n)$. Assume assumptions (3.2), (3.3), (4.3), (4.6) and either (4.4) or (4.5) to hold. Then, Algorithm 1 retrieves the community of all but a vanishing fraction of nodes.*

¹Recall that *exact* means correctly reconstructing the group-membership of *all* vertices, whereas with *quasi-exact* recovery only a *negligible fraction* of the vertices are misclassified.

The first step estimates L . Indeed, by definition there are only L non-zero eigenvalues of P . Those are all of order $1/\bar{D}$ and the corresponding first eigenvalues of \widehat{H} are of the same order. The remaining eigenvalues of \widehat{H} are negligible with respect to $f(n)/\bar{D}$, due to the choice of f (see below for details).

Under the assumptions in Theorem 4.3.1, all but a negligible number of rows of the matrix having the first L eigenvectors of \widehat{H} as its columns, cluster for large n to within negligible distance of block-specific representatives that are separated by some non-vanishing gap (call the corresponding vertices typical). This is exploited in the third step. There, with high probability, all picked vertices are typical. Thus, for a pair t , $\delta(t)$ vanishes in front of $f^{2/3}(n)$ if the vertices in the pair belong to the same community. Hence, by calculating the distance between the other vertices, we obtain ϵ as an estimator for the gap mentioned above. At most $f(n)^{2/3} n$ vertices are not typical. Hence, the chosen ball around m with radius $\epsilon/8$ contains a negligible number of non-typical vertices, the remaining vertices should necessarily be in the same community. By enlarging the radius of the ball around m , we include all vertices of a single community. See the proof of Theorem 4.3.1 below for more details.

Remark 4.3.2. *Note that the only input to the algorithm is the regime (i.e., either $D_1(n) = \Theta(\log(n))$ or $D_1(n) \gg \log(n)$). This information is used to pick the right form of the function f . Alternatively, we could adapt the algorithm so that it requires $L = \text{Rank}(B)$ and α_{\min} instead of the regime: Step 1 can then be skipped, in Step 2 we replace \widehat{L} by L and in Step 3 we chose a vertex m that contains in its $\epsilon/8$ neighbourhood at least $\alpha_{\min}n/2$ vertices.*

4.4 Relation to other spectral algorithms in the literature

Before we prove the main theorem, we make some observations and remarks.

4.4.1 Adjacency matrix

In [76] and [87], the authors use the adjacency matrix A of a graph to recover the underlying community-structure. They consider the matrix having the first K eigenvectors of A as its columns and show that, under appropriate conditions, its rows cluster now in K different *directions*. However, results in [23] and [90] suggest that the algorithms in [76] and [87] fail when the expected degree sequence is too irregular. Intuitively, if the prescribed degree sequence follows a power-law, then so does the spectrum of the adjacency matrix. Further, as we shall demonstrate below, the first K eigenvectors correspond only to the K top-degree nodes, and should therefore not be expected to capture more global features of a graph, such as its underlying block-structure. The following theorem makes this observation more rigorous:

Theorem 4.4.1. *Consider a DC-SBM $\mathcal{G}(B, K, \{\sigma_u\}_{u=1}^n, \{D_u\}_{u=1}^n)$ such that*

$$D_u = \begin{cases} D_1 & \text{if } 1 \leq u < n - k \\ D_1 n^\gamma (u + 1 - (n - k)) & \text{if } u \geq n - k, \end{cases} \quad (4.10)$$

where $k = n^\beta$ and the constants β and γ obey:

$$D_1^2(n) n^{2\gamma + 4\beta - 1} \rightarrow 0 \quad (4.11)$$

and

$$\gamma > 4\beta. \quad (4.12)$$

Further, assume that

$$\sigma_u = \begin{cases} 2 & \text{if } u \leq \frac{n}{2} \\ 1 & \text{if } u > \frac{n}{2}. \end{cases} \quad (4.13)$$

Under these conditions, the first k eigenvectors become for large n indistinguishable from the eigenvectors of a graph that is the disjoint-union of k stars having degrees $D_n + o(1), \dots, D_{n-k} + o(1)$.

For instance, $D_1(n) = n^{1/20}$, $\beta = 1/20$ and $\gamma = 1/5$, meets the assumptions in Theorem 4.4.1. Further, it verifies the conditions in the main theorem (Theorem 4.3.1): Algorithm 1 will successfully return the community membership of all but a vanishing fraction of nodes.

We remark that the above theorem is inspired by the main result in [90]. There, random graphs without community structure are considered and the power-law behaviour of the corresponding spectrum is obtained. To say something about the eigenvectors, we additionally introduce a gap between the top k degreeed-nodes and the remaining $n - k$ nodes. This allows us to use Lemma A.0.5, see the proof of Theorem 4.4.1 below.

4.4.2 Spectral clustering on ratios of eigenvectors

Interestingly, the first eigenvectors of A do contain information about the underlying community structure, but in a hidden way. Indeed, the SCORE method proposed in [65] shows that, under some conditions, using the coordinate-wise ratios of the leading eigenvectors leads to consistent clustering.

Note that we obtain the same random graph model as in [65] by putting $\theta(u) := D_u/\sqrt{n\bar{D}\alpha}$ and $P(i, j) = \alpha B_{ij}$, where $\alpha^{-1} = \max_{i,j} B_{ij}$. We further note that the conditions are more stringent: (2.7) demands that P (or B) is non-singular which is unnecessary here, see Remark 4.5.2 below.

4.4.3 Laplacian

As we pointed out, the adjacency matrix does not capture accurately global properties of a graph. The normalized Laplacian is a more suitable candidate. It is defined by $L = I - D^{-1/2}AD^{-1/2}$, where I is the identity matrix, A is the adjacency graph and D the diagonal matrix with the row sums of A on its diagonal (i.e., the degrees). Object of study in [23] is the Laplacian spectra of random graphs with a given degree sequence (d_1, \dots, d_n) where edges are independently present between each pair of vertices (u, v) with probability $\frac{d_u d_v}{\sum_{i=1}^n d_i}$. In the regime $d_1^2 \gg \bar{D}$, with $\bar{D} = 1/n \sum_{l=1}^n d_l$, the eigenvalues satisfy the semicircle law with respect to the circle of radius $2/\sqrt{\bar{D}}$ centred at 1.

Denote the eigenvalues of the normalized laplacian by $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \leq 2$. It is a well-known fact that all eigenvalues are located in the interval $[0, 2]$ and that the algebraic multiplicity of 0 equals the number of components in the graph. The authors of [23] further study the spectral gap $\lambda = \min\{\lambda_2, 2 - \lambda_n\}$, which reflects global properties of the random graph. According to [23], when $d_1 \gg \log^2(n)$,

$$\lambda \geq 1 - \frac{1 + o(1)}{4/\sqrt{\bar{w}}} - \frac{\log^2(n)}{d_1},$$

thus in this dense regime, all non-zero eigenvalues are close to 1 and thus the spectrum of the Laplacian contains no outliers, in contrast with the adjacency matrix. This bound is improved in [24], to

$$\lambda \geq 1 - 2\sqrt{\frac{6\log(2n)}{d_1}},$$

for $d_1 \gg \log(n)$.

The stochastic block model is a special case of the latent space model [56]. In this model an unknown vector z_u is associated to each node u (in a social network, this vector would represent the unknown social position of person u) and an edge between u and v is present with probability depending only on z_u and z_v . If A is the adjacency matrix of the graph, D the diagonal matrix containing the degrees and $L = D^{-1/2}AD^{-1/2}$, then the population version of these matrices are defined as $\mathcal{A} = \mathbb{E}[A|z_1, \dots, z_n]$, $\mathcal{D} = \text{diag}(\sum_{v=1}^n \mathcal{A}_{1v}, \dots, \sum_{v=1}^n \mathcal{A}_{nv})$, and, $\mathcal{L} = \mathcal{D}^{-1/2}\mathcal{A}\mathcal{D}^{-1/2}$. In [109] convergence of the empirical eigenvectors of L to the population eigenvectors of \mathcal{L} is shown. This follows from their novel result establishing the convergence of L^2 to \mathcal{L}^2 in Frobenius norm. This forms the basis of an algorithm that uses the first k eigenvectors (according to the eigenvalues ordered decreasingly with respect to their absolute value). To recover the hidden communities in the SBM (thus, without degree-corrections). The algorithm is shown to succeed if those first k eigenvalues are sufficiently separated from the rest of the eigenvalues and if the minimum expected degree exceeds $\frac{\sqrt{2n}}{\sqrt{\log n}}$, which is more restrictive than the lower bound of $\log n$.

In [33] the matrix $\mathbb{E}[D]^{-1/2}A\mathbb{E}[D]^{-1/2}$ (reminiscent of the normalized Laplacian) is used to retrieve the underlying community structure in the DC-SBM. Note that this method requires the expected degrees to be known. It succeeds if the minimum degree is of order $\log^6 n$.

To deal with low-degree nodes, the authors in [21] use the degree-corrected random walk laplacian: $I - (D + \tau I)^{-1}A$, where $\tau > 0$ is a constant, to find clusters in the extended planted partition mode (EPPM) where the expected minimum degree is $\Omega(\log n)$. In the EPPM, B is a matrix where an element equals p if it is on the diagonal, and q otherwise; it is thus a special case of the DC-SBM. The algorithm based on the random walk laplacian requires τ as input and the optimal value of τ depends in a complex way on the degree-distribution of the graph. The main theorem in [21] comes with lengthy conditions that are not easy to compare with other results. This theorem restricted to the setting where all d_u 's equal d , assumes q to be a constant, which is more restrictive than our assumptions. It is unclear whether the results for the EPPM can be neatly generalized using the same operator to the DC-SBM, given the complexity of the present conditions.

Although the Laplacian captures global properties of a graph much better than the adjacency matrix, its spectrum is still influenced by the underlying degree-structure. Indeed, consider a DC-SBM with 3000 vertices divided in $K = 3$ equally-sized communities, with

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & 2 \\ 3 & 2 & 5 \end{pmatrix},$$

degree-sequence

$$D_u = \begin{cases} u^{1/3} & \text{if } u = 1, \dots, 1000 \\ (u - 1000)^{1/3} & \text{if } u = 1001, \dots, 2000 \\ (u - 2000)^{1/3} & \text{if } u = 2001, \dots, 3000, \end{cases} \quad (4.14)$$

and community-membership

$$\sigma_u = \begin{cases} 1 & \text{if } u = 1, \dots, 1000 \\ 2 & \text{if } u = 1001, \dots, 2000 \\ 3 & \text{if } u = 2001, \dots, 3000. \end{cases} \quad (4.15)$$

In Figure 4.1, we have plot the eigenvectors corresponding to the first and second largest absolute eigenvalue of $I - \mathbb{E}[D]^{-1/2} \mathbb{E}[A] \mathbb{E}[D]^{-1/2}$, where A is the adjacency matrix and D is the diagonal matrix containing the row sums of A . The Laplacian concentrates around $I - \mathbb{E}[D]^{-1/2} \mathbb{E}[A] \mathbb{E}[D]^{-1/2}$ if the minimum degree is large

enough (see Section 8). The community structure is clearly perturbed by the degree-sequence. In general, an additional step is needed to determine the community-membership of all nodes when using the Laplacian.

Compare this figure to Figure 4.2, containing the first two eigenvectors of $\mathbb{E}[D]^{-1} \mathbb{E}[A] \mathbb{E}[D]^{-1}$. The vertices are seen to be clearly divided into three communities.

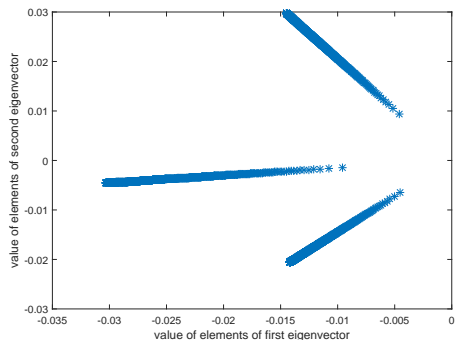


Figure 4.1: Plot of the eigenvectors corresponding to the first and second largest absolute eigenvalue of $I - \mathbb{E}[D]^{-1/2} \mathbb{E}[A] \mathbb{E}[D]^{-1/2}$, where A is the adjacency matrix of a random graph drawn according to the DC-SBM defined at the end of Section 4.2, and D is the diagonal matrix containing the row sums of A . For those eigenvectors, say $(x_1, \dots, x_n)'$ and $(y_1, \dots, y_n)'$, we draw a dot (x_u, y_u) for each element u .

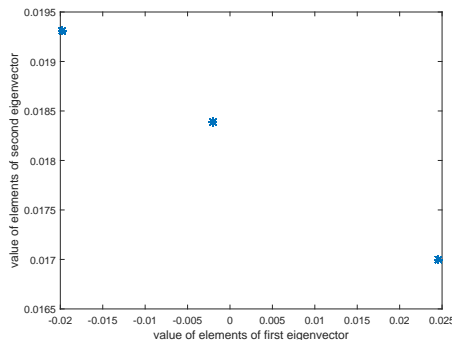


Figure 4.2: Plot of the eigenvectors corresponding to the first and second largest absolute eigenvalue of $\mathbb{E}[D]^{-1} \mathbb{E}[A] \mathbb{E}[D]^{-1}$, where A is the adjacency matrix of a random graph drawn according to the DC-SBM defined at the end of Section 4.2, and D is the diagonal matrix containing the row sums of A . For those eigenvectors, say $(x_1, \dots, x_n)'$ and $(y_1, \dots, y_n)'$, we draw a dot (x_u, y_u) for each element u . Note that many elements are represented by the same dot, clearly reflecting the community structure.

Now consider another two-community DC-SBM on n vertices with

$$B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

degree-sequence

$$D_u = \begin{cases} \log^2(n) & \text{if } u \leq n/2 \\ 100 \log^2(n) & \text{if } u > n/2. \end{cases} \quad (4.16)$$

and community-membership

$$\sigma_u = \begin{cases} 1 & \text{if } u \leq n/2 \\ 2 & \text{if } u > n/2. \end{cases} \quad (4.17)$$

Then, according to Lemma A.0.5, the eigenvectors of H become eventually indistinguishable from the eigenvectors of the $n \times n$ matrix with zero-diagonal and all other elements equal to $\frac{1}{nD}$. Clearly the communities can not be recovered from the latter matrix.

The off-diagonal elements of $\mathbb{E}[D]^{-1/2} \mathbb{E}[A] \mathbb{E}[D]^{-1/2}$ are given by $\frac{1}{nD} \sqrt{D_u} \sqrt{D_v} = \frac{2}{101n} Z_{\sigma_u \sigma_v}$, with $Z = \begin{pmatrix} 1 & 10 \\ 10 & 100 \end{pmatrix}$. Now, Z has eigenvector $\begin{pmatrix} 1 \\ 10 \end{pmatrix}$, corresponding to eigenvalue 101. The other eigenvalue is zero. So that the minimal gap between

different eigenvalues of

$\mathbb{E}[D]^{-1/2} \mathbb{E}[A] \mathbb{E}[D]^{-1/2}$ is $2 - \mathcal{O}(1/n)$. According to [24],
 $\rho\left(D^{-1/2}AD^{-1/2} - \mathbb{E}[D]^{-1/2} \mathbb{E}[A] \mathbb{E}[D]^{-1/2}\right) = o(1)$ w.h.p., where $\rho(X)$ denotes the spectral radius of a matrix X . Consequently, Lemma A.0.5 entails that for large n , clustering according to the eigenvector of $D^{-1/2}AD^{-1/2}$, corresponding to its largest eigenvalue, reveals the community-membership of all but a vanishing fraction of nodes.

Those two examples hint that whether the Laplacian L or the degree-normalized adjacency matrix H should be used depends on the correlation between the degrees and the communities, and the 'signal-strength' of B . The first example shows that if the degrees are uncorrelated, L seems to add some extra noise, whereas H 'filters' the degrees and reflects immediately the underlying communities. In the second example, B gives no information about the communities, but the vertices can be clustered according to their degrees. H ignores this degree-structure and thus fails to detect the communities. L on its turn still reflects the degree-sequence and therefore the communities.

4.4.4 Regularized spectral clustering

The paper [106] deals with the shortcomings of the Laplacian by inflating the degrees: Given a number $\tau > 0$, the regularized graph Laplacian [106, 21] is defined as

$$L_\tau = D_\tau^{-1/2}AD_\tau^{-1/2}, \quad (4.18)$$

where $D_\tau = D + \tau I$.

The regularized spectral clustering algorithm in [106] starts with computing the matrix $X = [X_1, X_2, \dots, X_K]$, where X_1, X_2, \dots, X_K are the eigenvectors corresponding to the K largest eigenvalues. A matrix X^* is then formed by projecting each row of X on the unit sphere. Considering each row of X^* as a point in \mathbb{R}^K , and applying k -means with K centres on these points gives an almost-exact clustering provided some conditions on $\delta + \tau$ (δ is the smallest expected degree) and the smallest strictly positive eigenvalue of L_τ hold. In particular, condition (a) in Theorem 4.2 demands that $\delta + \tau \gg \log(n)$. Since simulation results suggest that τ should be taken as the average degree, it is unclear if this method outperforms the algorithm proposed in the underlying chapter.

We note that [106] is the first work that relates the leverage scores (the euclidean norm of the rows of X) to the quality of the outputted clustering.

4.4.5 Degree-normalized adjacency matrix

The same matrix H is used in [27] to perform community detection on the DC-SBM in the sparse regime (the minimum degree is bounded from below by a constant). The main restriction in their setting is that the minimum degree must be of the same order as the average degree, more precisely there exists $\epsilon > 0$ such that $D_i > \epsilon \bar{D}$ for all i . Hence too much irregularity in the degree sequence is not captured. In this sense our work complements their results.

Spectral clustering is performed in [27] on a minor of \hat{H} where the rows and columns of vertices with a degree smaller than $D_{\text{average}}/\log(n)$ (where D_{average} is the observed average degree in the graph) are put to zero, which is not the same as leaving out completely the nodes with a too low degree. Due to the assumption that all expected degrees are of the same order, most observed degree will exceed the lower bound $D_{\text{average}}/\log(n)$.

There are alternative ways to deal with low degree nodes, see for instance Section 4.8 on future research.

4.5 Discussion

4.5.1 When does the degree-normalized adjacency matrix fail?

Consider a DC-SBM with $K \geq 2$ communities, such that for two different communities $i \neq j$, for all l , $\frac{B_{il}}{M_i} = \frac{B_{jl}}{M_j}$. Then, it can be verified that, for large n , in a dense enough regime, the eigenvectors of H corresponding to non-zero eigenvalues, do not distinguish between communities i and j .

Further, the method breaks down in a too sparse regime. For instance, two low-degreed vertices connected by an edge cause the top eigenvectors to concentrate around them. We observed this when applying \hat{H} on the sparse Political Blogs network [3], see Section 7.

4.5.2 Interpretation of the conditions

Note that, since $\mathbb{E}\hat{D}_u$ is related to D_u according to (4.7), \hat{H} normalizes the tendency of communities to connect by the average degree of their nodes and loses therefore some information about the graph. See the observations and remarks below:

Observation 4.5.1. *If, for some $i, j, l \in S$,*

$$\frac{B_{ij}}{M_i} = \frac{B_{lj}}{M_l},$$

then

$$\frac{\mathbb{E}[\#\text{edges between community } i \text{ and } j]}{\mathbb{E}[\text{total degree of vertices in community } i]} = \frac{\mathbb{E}[\#\text{edges between community } l \text{ and } j]}{\mathbb{E}[\text{total degree of vertices in community } l]}.$$

Remark 4.5.2. *The identifiability condition is violated if there are distinct i and l and there exists some constant $c > 0$ such that*

$$B_{ij} = cB_{lj}$$

for all j . Indeed, in that case, $M_i = cM_l$ and thus

$$\frac{B_{ij}}{M_i} = \frac{cB_{lj}}{cM_l} = \frac{B_{lj}}{M_l}.$$

However, unlike the setting considered in [76, 65], it is not necessary for B to be full rank. Indeed, consider

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & 2 \\ 3 & 2 & 5 \end{pmatrix},$$

which has rank 2. Let $\alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{3}$ and $\sum_{\sigma_u=i} D_u = i\alpha_i n \log^2(n)$ for all $i = 1, 2, 3$. Then it is easily verified that the identifiability condition is met.

Note that $\mathcal{G}(B, K, \{\sigma_u\}_{u=1}^n, \{D_u\}_{u=1}^n)$ and $\mathcal{G}(B^*, K, \{\sigma_u\}_{u=1}^n, \{D_u^*\}_{u=1}^n)$ generate the same ensemble of random graphs whenever

$$\frac{D_u B_{\sigma_u \sigma_v} D_v}{\bar{D}} = \frac{D_u^* B_{\sigma_u \sigma_v}^* D_v^*}{\bar{D}^*}.$$

Hence, the underlying block-matrix B cannot be estimated from a single observation of the graph. Rather, we may estimate

$$n\bar{D} \approx \sum_u \hat{D}_u, \tag{4.19}$$

and, denoting the assigned community-membership (after applying our reconstruction algorithm) of l by τ_l ,

$$\left(\sum_u \widehat{D}_u \right) \frac{\sum_{u:\tau_u=i} \sum_{v:\tau_v=j} \widehat{H}_{uv}}{(\sum_{u:\tau_u=i} 1) (\sum_{v:\tau_v=j} 1)} \approx \frac{B_{ij}}{\overline{M}_i \overline{M}_j}. \quad (4.20)$$

Hence, for a DC-SBM $\mathcal{G}(B, K, \{\sigma_u\}_{u=1}^n, \{D_u\}_{u=1}^n)$, the matrix

$$\left(\frac{B_{ij}}{\overline{M}_i \overline{M}_j} \right)_{i,j=1}^K$$

is identifiable, not B . It is due to this degeneracy of the DC-SBM and the structure of \widehat{H} that condition (4.3) in Theorem 4.3.1 is the best possible:

Lemma 4.5.3. *Consider a DC-SBM $\mathcal{G}(B, K, \{\sigma_u\}_{u=1}^n, \{D_u\}_{u=1}^n)$. Fix i and l , then the following are equivalent:*

1. for all j we have

$$\frac{B_{ij}}{M_i} = \frac{B_{lj}}{M_l},$$

2. there exist a DC-SBM $\mathcal{G}(B^*, K, \{\sigma_u\}_{u=1}^n, \{D_u^*\}_{u=1}^n)$, with the same community-structure $\{\sigma_u\}_u$, such that for all j ,

$$B_{ij}^* = B_{lj}^*$$

and, for all u, v ,

$$\frac{D_u B_{\sigma_u \sigma_v} D_v}{\overline{D}} = \frac{D_u^* B_{\sigma_u \sigma_v}^* D_v^*}{\overline{D}^*}.$$

4.5.3 Random Walk point of view

The matrix \widehat{H} is related to a random walk on an instance of the random graph. Indeed,

$$\widehat{H}_{uv} = \begin{cases} \frac{1}{\widehat{D}_u} \frac{1}{\widehat{D}_v} A_{uv} = \frac{A_{uv}}{\widehat{D}_u} \frac{A_{vu}}{\widehat{D}_v} & \text{if } A_{uv} = 1, \\ 0 & \text{if } A_{uv} = 0, \end{cases}$$

since $A_{uv} = A_{vu}$ is either 1 (in case edge uv is present) or 0 (when u and v are not connected). Now, $\widehat{D}_u = \sum_{l=1}^n A_{lu}$, as it is the observed degree, which we denoted here in increasing order: $\widehat{D}_1 \leq \widehat{D}_2 \leq \dots \leq \widehat{D}_n$. Thus, $\frac{A_{uv}}{\widehat{D}_u}$ is exactly the probability that a random walk (in an undirected graph without weights) jumps from vertex u to v , given that it is currently at vertex u . Denoting the latter probability by $\mathbb{P}_u(u \rightarrow v)$, we see that

$$\widehat{H}_{uv} = \mathbb{P}_u(u \rightarrow v) \mathbb{P}_v(v \rightarrow u) = \mathbb{P}_u(u \rightarrow v \rightarrow u),$$

due to the Markov property of the random walk. In other words, \widehat{H}_{uv} is the probability that a random walk currently at vertex u will consecutively traverse edge uv and back.

Extending this observation to powers of \widehat{H} leads to:

$$(\widehat{H}^k)_{uw} = \sum_{l_1=1, \dots, l_{k-1}=1}^n \mathbb{P}_u(u \rightarrow l_1 \rightarrow \dots \rightarrow l_{k-1} \rightarrow v) \mathbb{P}_v(v \rightarrow l_{k-1} \rightarrow \dots \rightarrow l_1 \rightarrow u),$$

the probability that a random walk, after traversing a path of length k starting at u and ending at v , subsequently traverses that path in the exact opposite direction.

Further, note that

$$(\widehat{D}_1, \dots, \widehat{D}_n)\widehat{H} = (\mathbb{1}_{\widehat{D}_1 \neq 0}, \dots, \mathbb{1}_{\widehat{D}_n \neq 0}),$$

hence if \mathbf{v} is an eigenvector of \widehat{H} with eigenvalue λ , then

$$\sum_{u=1}^n \mathbb{1}_{\widehat{D}_u \neq 0} v_u = \lambda \sum_{u=1}^n \widehat{D}_u v_u.$$

Since it can be easily verified that \widehat{H} is primitive on connected components, the Perron-Frobenius theorem implies that the eigenvector \mathbf{v}_{max} corresponding to the largest eigenvalue λ_{max} (which is positive) has only positive elements. Hence,

$$\lambda_{max} = \frac{\sum_{u=1}^n \mathbb{1}_{\widehat{D}_u \neq 0} v_u}{\sum_{u=1}^n \widehat{D}_u v_u} \geq \frac{\sum_{u=1}^n \mathbb{1}_{\widehat{D}_u \neq 0} v_u}{\widehat{D}_n \sum_{u=1}^n \mathbb{1}_{\widehat{D}_u \neq 0} v_u} = \frac{1}{\widehat{D}_n}.$$

We may derive an upper bound by noting that the spectral radius is bounded from above by the maximal absolute row sum:

$$\lambda_{max} \leq \max_{u=1}^n \left(\sum_{v=1}^n \mathbb{P}_u(u \rightarrow v \rightarrow u) \right).$$

4.6 Outline of proof of main theorem

In this section we consider the setting of Theorem 4.3.1. All lemmas here, except Lemma 4.6.4, assume either (4.4) or (4.5) to hold. Lemma 4.6.4 assumes condition (4.4): the minimum degree should grow faster than $\log(n)$. Lemma 4.6.5 assumes (4.5): the minimum degree is of order $\log(n)$.

Our first objective is to show that \widehat{H} is close to some matrix P , in the sense that their difference $W := \widehat{H} - P$ has negligible spectral radius relatively to that of P . Here, an entry (u, v) of P is defined as

$$P_{uv} = \frac{1}{n\widehat{D}} \frac{B_{\sigma_u \sigma_v}}{\overline{M}_{\sigma_u} \overline{M}_{\sigma_v}}. \quad (4.21)$$

We relate P in turn to Z defined by

$$Z_{ij} = \frac{\alpha_j B_{ij}}{\overline{M}_i \overline{M}_j}, \quad i, j \in S. \quad (4.22)$$

Indeed, we show that if $y = (y(1), \dots, y(K))^T$ is an eigenvector of Z with eigenvalue λ , then $(y(\sigma_1), \dots, y(\sigma_n))^T$ fulfils that role for P with eigenvalue $\frac{1}{\widehat{D}}\lambda$. As a consequence, the eigenvectors of P associated to non-zero eigenvalues are constant on blocks.

Finally, we consider the matrix that has the first L eigenvectors of P as its columns. We show that the rows of this matrix cluster to within vanishing distance of block-specific representatives. We start by inspection of the difference

$$W = \widehat{H} - P = (\widehat{H} - H) + (H - \mathbb{E}[H]) + (\mathbb{E}[H] - P), \quad (4.23)$$

where H is defined as

$$H_{uv} = \begin{cases} \frac{1}{\mathbb{E}\widehat{D}_u\mathbb{E}\widehat{D}_v}A_{uv} & \text{if } A_{uv} = 1, \\ 0 & \text{otherwise .} \end{cases} \quad (4.24)$$

Define

$$\Delta(P) = \min\{|\lambda - \mu| : \lambda \neq \mu, \lambda, \mu \text{ eigenvalue of } P\},$$

i.e., the smallest gap between consecutive eigenvalues. A crucial role will be played by Lemma A.0.5 below, which says that to any eigenvector \widehat{x} of \widehat{H} there exists an eigenvector x of P such that $\|x - \widehat{x}\| \rightarrow 0$ as $n \rightarrow \infty$, whenever

$$\frac{\rho(W)}{\Delta(P)} \rightarrow 0,$$

as $n \rightarrow \infty$, where we recall that $\rho(X)$ denotes the spectral radius of a matrix X . Hence, we need to calculate $\Delta(P)$:

Lemma 4.6.1. *The smallest gap between subsequent eigenvalues of P is given by*

$$\Delta(P) = \Omega\left(\frac{1}{D}\right).$$

All terms in the right hand side of (4.23) have negligible spectral radius with respect to $\Delta(P)$:

Lemma 4.6.2. *The matrix $\mathbb{E}[H]$ is close to P in the following sense:*

$$\rho(\mathbb{E}[H] - P) = \mathcal{O}\left(\frac{1}{D_1}\right) \frac{1}{D} = o_n(1) \frac{1}{D}.$$

Lemma 4.6.3. *The matrix H concentrates with high probability around its expectation, as follows:*

$$\rho(H - \mathbb{E}[H]) = \mathcal{O}\left(\frac{1}{\sqrt{\log(n)}}\right) \frac{1}{D} = o_n(1) \frac{1}{D}.$$

Lemma 4.6.4. *Consider the DC-SBM in the dense regime, where (4.4) holds. Then, for the spectral radius of the difference $\widehat{H} - H$ it holds with high probability that*

$$\rho(\widehat{H} - H) = \mathcal{O}\left(\sqrt{\frac{\log(n)}{D_1(n)}}\right) \frac{1}{D(n)} = o_n(1) \frac{1}{D}.$$

Lemma 4.6.5. *Consider the DC-SBM in the regime where (4.5) holds. Then, for the spectral radius of the difference $\widehat{H} - H$ it holds with high probability that*

$$\rho(\widehat{H} - H) = \mathcal{O}\left(\frac{1}{\log^{1/3}(n)}\right) \frac{1}{D(n)} = o_n(1) \frac{1}{D}.$$

We use these Lemmas in conjunction with Lemma A.0.5 below to prove:

Lemma 4.6.6. *To each normed eigenvector $\widehat{\mathbf{x}}$ of \widehat{H} corresponds a normed eigenvector \mathbf{x} of P such that*

$$\widehat{\mathbf{x}} \cdot \mathbf{x} = 1 - \mathcal{O} \left(\left(\frac{\rho(W)}{\Delta(P)} \right)^2 \right) = 1 - o_n(1),$$

where

$$\rho(W) \leq \rho(\widehat{H} - H) + \rho(H - \mathbb{E}[H]) + \rho(\mathbb{E}[H] - P).$$

Having proved this lemma, we show that Algorithm 1 indeed correctly reconstructs the community of all but a vanishing fraction of vertices.

Recall the definition of \widehat{H} and observe that \widehat{H} is symmetric. Consequently, there exist n eigenvectors of \widehat{H} that form an *orthonormal* basis: thus, we are indeed able to find L *orthonormal* eigenvectors of \widehat{H} corresponding to its first eigenvectors.

Next we show that the $(\widehat{z}_u)_{u \in V}$, defined in (4.9), tend to block-representatives:

Lemma 4.6.7. *There exist K vectors $\{t_k\}_{k \in \mathcal{S}}$, i.e., block-representatives, such that*

$$\|\sqrt{n}\widehat{z}_u - t_{\sigma_u}\| = \mathcal{O} \left(\left(\frac{\rho(W)}{\Delta(P)} \right)^{2/3} \right) = o_n(1)$$

for all but $\mathcal{O} \left(n \left(\frac{\rho(W)}{\Delta(P)} \right)^{2/3} \right)$ nodes.

The remaining and crucial step is to demonstrate that those block-representatives are indeed distinct:

Lemma 4.6.8. *Assume that for all i, j there exists i' such that*

$$\frac{B_{ii'}}{\overline{M}_i \overline{M}_{i'}} \neq \frac{B_{jj'}}{\overline{M}_j \overline{M}_{j'}}, \quad (4.25)$$

then $|t_k - t_l| = \Omega(1)$ for all $k \neq l$.

Proof of Theorem 4.3.1. After proving the above lemmas, it remains to show that \widehat{L} in step (1) of Algorithm 1 with high probability equals L . Further, we should verify that the procedure in step 3 forms the right clusters. For the first step notice the following: In the regime where (4.4) holds,

$$\rho(W) = \mathcal{O} \left(\frac{1}{D_1} + \frac{1}{\sqrt{\log(n)}} + \sqrt{\frac{\log(n)}{D_1}} \right) \frac{1}{\overline{D}},$$

and in the other regime, where (4.5) holds,

$$\rho(W) = \mathcal{O} \left(\frac{1}{D_1} + \frac{1}{\sqrt{\log(n)}} + \frac{1}{\log^{1/3}(n)} \right) \frac{1}{\overline{D}}.$$

Compare this to f as in Algorithm 1: depending on the regime, the term in parentheses goes to zero upon division by $\frac{f(n)}{\overline{D}}$. To see this, note that due to Bernstein's inequality (A.0.7), equation (4.35), $\widehat{D}_u \in (1/2\overline{M}_{\sigma_u}, 3/2\overline{M}_{\sigma_u})D_u$ for $u = 1$ and $u = n$

with high probability. Hence \widehat{D}_1 (\widehat{D}_n) is of the same order of magnitude as D_1 (respective D_n). Now, due to Lemma A.0.5 below, the first L eigenvalues of \widehat{H} are of order $\frac{1}{\widehat{D}} - \mathcal{O}(\rho(W)) \gg \frac{f(n)}{\widehat{D}}$. The remaining eigenvalues are of order $\mathcal{O}(\rho(W)) \ll \frac{f(n)}{\widehat{D}}$. Further D_{average} may be written as twice the sum of $\Omega(n^2)$ independent Bernoulli random variables. It is thus with high probability a constant away from \overline{D} . Hence $\widehat{L} = L$ with high probability.

In step 3, the probability that all picked pairs contain only typical vertices (i.e., whose corresponding rows cluster around K centres) is larger than $(1 - f^{2/3}(n))^{2\tau(n)}$ which tends to one, since $f^{2/3}(n)\tau(n) \rightarrow 0$ as $n \rightarrow \infty$. Thus, with high probability, for a pair t , $\delta(t)$ vanishes in front of $f^{2/3}(n)$ if the vertices in the pair belong to the same community. $\delta(t)$ is of order $\Omega(1)$ otherwise. Hence, ϵ , as defined in step 3 of Algorithm 1, is of order $\Omega(1)$, it thus estimates the separation-distance in Lemma 4.6.8.

Further, at most $f(n)^{2/3} n$ vertices are not typical. Hence, the chosen ball around m with radius $\epsilon/8$ contains at least $(f(n)^{1/3} - f(n)^{2/3})n \gg f(n)^{2/3} n$ typical vertices. Those must necessarily belong to the same community. Since all typical vertices belonging to the same community are at most a distance smaller than $f(n)^{2/3}$ apart, all of them are located in the ball of radius $\epsilon/4$ around m .

We see that the algorithm puts, with high probability, all but a vanishing fraction of nodes in K clusters. \square

4.7 Proofs

4.7.1 Main theorem

In the proofs below, we shall often write

$$D_u = \phi_u \omega(n), \quad (4.26)$$

where $1 = \phi_1 \leq \phi_2 \leq \dots \leq \phi_n$, and

$$\omega(n) = D_1. \quad (4.27)$$

Further, we introduce

$$g(n) = \sum_{l=1}^n \phi_l, \quad (4.28)$$

$$\bar{\phi}(n) = \frac{g(n)}{n}, \quad (4.29)$$

Proof of Lemma 4.6.1. Write

$$P_{uv} = \frac{1}{n\overline{D}} \frac{B_{\sigma_u \sigma_v}}{\overline{M}_{\sigma_u} \overline{M}_{\sigma_v}} = \frac{1}{n\overline{D}} \frac{Z_{\sigma_u \sigma_v}}{\alpha_{\sigma_v}}.$$

Let $\mathbf{y} = (y(1), \dots, y(K))^T$ be an eigenvector of Z with eigenvalue λ , we show that

$\mathbf{w} = (y(\sigma_1), \dots, y(\sigma_n))^T$ is an eigenvector of P with eigenvalue $\frac{1}{D}\lambda$. Indeed,

$$\begin{aligned}
P\mathbf{w} &= \begin{pmatrix} \sum_{l=1}^n P_{1l} \cdot y(\sigma_l) \\ \vdots \\ \sum_{l=1}^n P_{nl} \cdot y(\sigma_l) \end{pmatrix} \\
&= \frac{1}{n\bar{D}} \begin{pmatrix} \sum_{l=1}^n Z_{\sigma_1\sigma_l}/\alpha_{\sigma_l} \cdot y(\sigma_l) \\ \vdots \\ \sum_{l=1}^n Z_{\sigma_n\sigma_l}/\alpha_{\sigma_l} \cdot y(\sigma_l) \end{pmatrix} \\
&= \frac{1}{n\bar{D}} \begin{pmatrix} \sum_{k=1}^K n \alpha_k Z_{\sigma_1 k}/\alpha_k \cdot y(k) \\ \vdots \\ \sum_{k=1}^K n \alpha_k Z_{\sigma_n k}/\alpha_k \cdot y(k) \end{pmatrix} \\
&= \frac{1}{D} \begin{pmatrix} \lambda y(\sigma_1) \\ \vdots \\ \lambda y(\sigma_n) \end{pmatrix} \\
&= \frac{1}{D} \lambda \mathbf{w}.
\end{aligned}$$

Thus $\frac{1}{D}\lambda$ is an eigenvalue of P .

For the other direction, note that if $\sigma_u = \sigma_v$, then row u and row v in P are identical. Hence, if $\mathbf{w} = (w(1), \dots, w(n))^T$ is an eigenvector of P corresponding to a non-zero eigenvalue, then $w(u) = w(v)$. Let $\mathbf{w} = (w(\sigma_1), \dots, w(\sigma_n))^T$ be an eigenvector of P with eigenvalue $\lambda \neq 0$. By carrying out a similar calculation as above, we see that $(w(1), \dots, w(K))^T$ is an eigenvector of Z with eigenvalue $\bar{D}\lambda$.

The statement follows from this one-to-one correspondence between the eigenvectors of both matrices corresponding to non-zero eigenvalues. \square

Proof of Lemma 4.6.2. Note that

$$\mathbb{E}[H] - P = \mathbb{E}[H] - (P - \text{diag}(P_{11}, \dots, P_{nn})) + \text{diag}(P_{11}, \dots, P_{nn}).$$

Now,

$$\rho(\text{diag}(P_{11}, \dots, P_{nn})) = \mathcal{O}\left(\frac{1}{n\bar{D}}\right),$$

as $\text{diag}(P_{11}, \dots, P_{nn})$ contains only K different elements, each of order $\frac{1}{n\bar{D}}$. Further, for $u \neq v$,

$$\mathbb{E}[H_{uv}] = \frac{D_u}{\mathbb{E}[\widehat{D}_u]} \frac{D_v}{\mathbb{E}[\widehat{D}_v]} B_{\sigma_u\sigma_v} \frac{1}{n\bar{D}} = (1 + \delta(n)) \frac{B_{\sigma_u\sigma_v}}{\bar{M}_{\sigma_u} \bar{M}_{\sigma_v}} \frac{1}{n\bar{D}} = P_{uv} + \delta(n)P_{uv},$$

where $\delta(n) = \mathcal{O}(\epsilon_n)$, with, due to (4.7), $\epsilon_n \leq \max_i \frac{2}{\bar{M}_i} \frac{1}{\omega(n)}$ tending to zero uniformly for all u, v . Hence, due to Lemma A.0.4,

$$\rho(\mathbb{E}[H] - (P - \text{diag}(P_{11}, \dots, P_{nn})) + \text{diag}(P_{11}, \dots, P_{nn})) = \mathcal{O}\left(\frac{1}{D_1}\right) \frac{1}{D}.$$

\square

Proof of Lemma 4.6.3. We start by introducing the constants $C_{ij} = \frac{B_{ij}}{\bar{M}_i \bar{M}_j}$, and $\alpha = \max_{ij} \frac{1}{\bar{M}_i \bar{M}_j}$. Put for $u < v$,

$$X_{uv} = X_{vu} = \alpha^{-1} \omega^2(n) (H_{uv} - \mathbb{E}[H_{uv}]),$$

where $\omega(n)$ is defined in (4.27). That is,

$$X_{uv} = \frac{1 + o(1)}{\alpha} \left(\frac{1}{\overline{M}_{\sigma_u} \overline{M}_{\sigma_v}} \frac{1}{\phi_u \phi_v} \text{Ber} \left(\frac{D_u D_v}{n \overline{D}} B_{\sigma_u \sigma_v} \right) - C_{\sigma_u \sigma_v} \frac{\omega(n)}{\overline{\phi}} \frac{1}{n} \right),$$

with ϕ_u and $\overline{\phi}$ defined in (4.26), respectively, (4.29) (the $o(1)$ term is uniform due to (4.7)). Due to our choice of α and the assumption that $\phi_u \geq 1$ for all u ,

$$X_{uv} \in (1 + o(1)) [-p_{uv}, 1 - p_{uv}],$$

where

$$p_{uv} = \frac{C_{\sigma_u \sigma_v} \omega(n)}{\alpha} \frac{1}{\overline{\phi}} \frac{1}{n}.$$

Let $\widehat{X}_{uv} = \frac{X_{uv}}{2}$, i.e., $\widehat{X}_{uv} \in [-p_{uv}, 1 - p_{uv}]$. We shall compare the symmetric zero-diagonal matrix \widehat{X} to the deviation from its expectation of another symmetric zero-diagonal matrix, where elements uv are given by $\text{Ber}(p_{uv})$, for $u \neq v$. Since by assumption (4.6),

$$\frac{\omega(n)}{\overline{\phi}(n)} = \frac{D_1^2(n)}{\overline{D}(n)} = \Omega(\log(n)), \quad (4.30)$$

Lemma A.0.6 applies. Following an argument given in [115], we construct a function Y_{uv} such that Y_{uv} has values only in $\{-p_{uv}, 1 - p_{uv}\}$ and $\mathbb{E}[Y_{uv} | \widehat{X}_{uv}] = \widehat{X}_{uv}$. First, let $\{U_{uv}\}_{u < v}$ be independent uniformly distributed random variables. Fix $u < v$. Define, for $x \in [-p_{uv}, 1 - p_{uv}]$ and $w \in [0, 1]$,

$$F_{uv}(x, w) = 1 - p_{uv} - 1_{x \leq w - p_{uv}},$$

and,

$$Y_{uv} = Y_{vu} = F_{uv}(\widehat{X}_{uv}, U_{uv}).$$

Then,

$$\mathbb{P}\left(F_{uv}(\widehat{X}_{uv}, U_{uv}) = 1 - p_{uv} \mid \widehat{X}_{uv}\right) = \widehat{X}_{uv} + p_{uv},$$

and,

$$\mathbb{P}\left(F_{uv}(\widehat{X}_{uv}, U_{uv}) = -p_{uv} \mid \widehat{X}_{uv}\right) = 1 - p_{uv} - \widehat{X}_{uv},$$

thus,

$$\mathbb{E}\left[Y_{uv} \mid \widehat{X}_{uv}\right] = \widehat{X}_{uv},$$

and,

$$\mathbb{P}(Y_{uv} = 1 - p_{uv}) = \mathbb{E}\left[\widehat{X}_{uv}\right] + p_{uv} = p_{uv}.$$

Hence, indeed, $Y_{uv} = \text{Ber}(p_{uv}) - p_{uv}$.

Let Y be the symmetric zero-diagonal matrix with each element uv given by Y_{uv} , for $u \neq v$. Then, according to Lemma A.0.6,

$$\mathbb{P}\left(\rho(Y) \leq \mathcal{O}\left(\sqrt{\frac{\omega(n)}{\overline{\phi}}}\right)\right) \geq 1 - \mathcal{O}(1/n^2). \quad (4.31)$$

We shall use this observation in the following comparison,

$$\rho(\widehat{X}) = \rho\left(\mathbb{E}[Y | \widehat{X}]\right) \leq \mathbb{E}\left[\rho(Y) \mid \widehat{X}\right],$$

by Jensen's inequality. Put $S = \mathbb{E} \left[\rho(Y) \mid \widehat{X} \right]$, we shall show that it is also upper bounded by $\mathcal{O} \left(\sqrt{\omega(n)/\overline{\phi}} \right)$.

Firstly, note that $|Y|$ is element-wise dominated by the all-one matrix, hence $\rho(Y) \leq n$. Secondly, the sigma-algebra generated by S is contained in the sigma-algebra generated by \widehat{X} . Hence,

$$\mathbb{E} [\rho(Y)|S] = \mathbb{E} \left[\mathbb{E} \left[\rho(Y) \mid \widehat{X} \right] \mid S \right] = \mathbb{E} [S|S] = S.$$

Further, both Y and \widehat{X} take only finitely many different values, and thus $\rho(Y)$ and S take values in a finite space. It therefore makes sense to consider, for $t > 0$, the function

$$\beta(\cdot) = \mathbb{P}(\rho(Y) > t | S = \cdot).$$

We have,

$$S = \mathbb{E} [\rho(Y)|S] \leq \beta(S)n + (1 - \beta(S))t,$$

i.e.,

$$\beta(S) \geq \frac{S - t}{n - t}.$$

Denote $\gamma = \mathbb{P}(S > t + 1)$, then

$$\mathbb{P}(\rho(Y) > t) = \mathbb{E} [\beta(S)] \geq \mathbb{E} [\beta(S) \mathbf{1}_{S > t+1}] \geq \frac{\gamma}{n - t}.$$

As a consequence, for $t = \mathcal{O} \left(\sqrt{\omega(n)/\overline{\phi}} \right)$, by (4.31) one has

$$\mathbb{P}(S > t + 1) = \gamma \leq (n - t) \mathbb{P}(\rho(Y) > t) = (n - t) \mathcal{O}(1/n^2) = \mathcal{O}(1/n).$$

Therefore,

$$\rho(H - \mathbb{E}[H]) = \frac{\alpha}{\omega^2(n)} \rho(X) = 2 \frac{\alpha}{\omega^2(n)} \rho(\widehat{X}) \leq \mathcal{O} \left(\sqrt{\frac{1}{\overline{\phi} \omega^3(n)}} \right),$$

Finally, due to (4.30),

$$\rho(H - \mathbb{E}[H]) = \mathcal{O} \left(\sqrt{\frac{\overline{\phi}}{\omega(n) \overline{\phi} \omega(n)}} \frac{1}{\overline{\phi} \omega(n)} \right) = \mathcal{O} \left(\frac{1}{\sqrt{\log(n)}} \right) \frac{1}{\overline{D}}.$$

□

Proof of Lemma 4.6.4. To prove this theorem we claim that in the present setting, it suffices to show that with high probability,

$$(\widehat{H} - H)_{uv} = \epsilon_{uv} H_{uv},$$

where, for some constant \widehat{C} and all large enough n ,

$$|\epsilon_{uv}| \leq \widehat{C} \epsilon(n), \tag{4.32}$$

with

$$\epsilon(n) := \sqrt{\frac{6}{\min_i \overline{M}_i} \frac{2 \log(n)}{\omega(n)}} = \mathcal{O} \left(\sqrt{\frac{\log(n)}{D_1(n)}} \right) \rightarrow 0, \tag{4.33}$$

by assumption. Consequently, after an appeal to Lemma A.0.4,

$$\rho(\widehat{H} - H) \leq \rho(|\widehat{H} - H|) \leq \widehat{C}\epsilon(n)\rho(H). \quad (4.34)$$

Since, $H = \mathbb{E}[H] + H - \mathbb{E}[H]$, it follows from Lemmas 4.6.2 and 4.6.3 that

$$\rho(H) = \mathcal{O}\left(\frac{1}{\overline{D}}\right).$$

Consider the difference

$$\frac{1}{\widehat{D}_u \widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u \mathbb{E}\widehat{D}_v} = \frac{1}{\mathbb{E}\widehat{D}_u \mathbb{E}\widehat{D}_v} \frac{1}{1 + \frac{\widehat{D}_u - \mathbb{E}\widehat{D}_u}{\mathbb{E}\widehat{D}_u}} - \frac{1}{\mathbb{E}\widehat{D}_u \mathbb{E}\widehat{D}_v} = \frac{1}{\mathbb{E}\widehat{D}_u \mathbb{E}\widehat{D}_v} \epsilon_{uv},$$

thus

$$\epsilon_{uv} = \frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u} + \frac{\mathbb{E}\widehat{D}_v - \widehat{D}_v}{\mathbb{E}\widehat{D}_v} + \mathcal{O}\left(\left(\frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u}\right)^2\right) + \mathcal{O}\left(\left(\frac{\mathbb{E}\widehat{D}_v - \widehat{D}_v}{\mathbb{E}\widehat{D}_v}\right)^2\right).$$

We quantify $\frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u}$. Since \widehat{D}_u is a sum of Bernoulli random variables with mean

$$\mathbb{E}[\widehat{D}_u] = D_u \overline{M}_{\sigma_u} (1 - o(1)),$$

where the $o(1)$ term follows from (4.7), we have for $\epsilon(n)$ as in (4.33), the Bernstein's inequality (see (A.0.7)),

$$\begin{aligned} \mathbb{P}\left(\left|\frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u}\right| > \epsilon(n)\right) &\leq 2\exp\left(-\frac{\epsilon^2(n)}{2 + \epsilon(n)/3} \mathbb{E}[\widehat{D}_u]\right) \\ &= 2\exp\left(-\frac{\epsilon^2(n)}{2 + \epsilon(n)/3} D_u \overline{M}_{\sigma_u} (1 - o(1))\right) \\ &\leq 2\exp\left(-\frac{\epsilon^2(n)}{3} \omega(n) \frac{\overline{M}_{\sigma_u}}{2}\right) \\ &\leq \frac{2}{n^2}. \end{aligned} \quad (4.35)$$

Invoking this we establish the union bound

$$\begin{aligned} \mathbb{P}\left(\left|\frac{\mathbb{E}\widehat{D}_1 - \widehat{D}_1}{\mathbb{E}\widehat{D}_1}\right| \leq \epsilon(n), \dots, \left|\frac{\mathbb{E}\widehat{D}_n - \widehat{D}_n}{\mathbb{E}\widehat{D}_n}\right| \leq \epsilon(n)\right) &\geq 1 - \sum_{u=1}^n \mathbb{P}\left(\left|\frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u}\right| > \epsilon(n)\right) \\ &\geq 1 - \frac{2}{n} \rightarrow 1, \end{aligned}$$

as $n \rightarrow \infty$. Hence,

$$E = \left\{ \left| \frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u} \right| \leq \epsilon(n) \text{ for all } u \in V \right\}$$

holds with high probability. Thus we establish (4.32): $|\epsilon_{uv}| \leq 2\epsilon(n) + \mathcal{O}(\epsilon^2(n)) \leq \widehat{C}\epsilon(n)$, with \widehat{C} a large enough constant.

Henceforth, we condition on E . Then, for $u \neq v$,

$$\widehat{H}_{uv} - H_{uv} = \left(\frac{1}{\widehat{D}_u \widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u \mathbb{E}\widehat{D}_v} \right) A_{uv} = \epsilon_{uv} \frac{1}{\mathbb{E}\widehat{D}_u \mathbb{E}\widehat{D}_v} A_{uv} = \epsilon_{uv} H_{uv}.$$

□

Proof of Lemma 4.6.5. We define

$$\epsilon(n) = \frac{1}{\log^{1/3}(n)} \quad (4.36)$$

and we shall call a vertex u good if $|\mathbb{E}\widehat{D}_u - \widehat{D}_u| \leq \epsilon(n)\mathbb{E}\widehat{D}_u$. We use this definition to split

$$\left| \frac{1}{\widehat{D}_u} \frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v} \right| A_{uv} = M_{uv} + M_{uv}^c + M_{uv}^r - M_{uv}^{\text{cr}}, \quad (4.37)$$

where

$$\begin{aligned} M_{uv} &= \left| \frac{1}{\widehat{D}_u} \frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v} \right| A_{uv} \mathbf{1}_{\{u \text{ and } v \text{ good}\}}, \\ M_{uv}^c &= \left| \frac{1}{\widehat{D}_u} \frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v} \right| A_{uv} \mathbf{1}_{\{v \text{ bad}\}}, \\ M_{uv}^r &= \left| \frac{1}{\widehat{D}_u} \frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v} \right| A_{uv} \mathbf{1}_{\{u \text{ bad}\}}, \\ M_{uv}^{\text{cr}} &= \left| \frac{1}{\widehat{D}_u} \frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v} \right| A_{uv} \mathbf{1}_{\{u \text{ and } v \text{ bad}\}}. \end{aligned}$$

We shall show that all terms in (4.37) have a negligible spectral radius compared to $\Delta(P)$. First note that the difference

$$\frac{1}{\widehat{D}_u} \frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v},$$

may be written as

$$\frac{1}{\mathbb{E}\widehat{D}_u \mathbb{E}\widehat{D}_v} \epsilon_{uv},$$

where

$$\epsilon_{uv} = \frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u} + \frac{\mathbb{E}\widehat{D}_v - \widehat{D}_v}{\mathbb{E}\widehat{D}_v} + \mathcal{O}\left(\left(\frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u}\right)^2\right) + \mathcal{O}\left(\left(\frac{\mathbb{E}\widehat{D}_v - \widehat{D}_v}{\mathbb{E}\widehat{D}_v}\right)^2\right).$$

Now, similarly as in the proof of Lemma 4.6.4, there exists a constant \widehat{C} , such that $\epsilon_{uv} \leq \widehat{C}\epsilon(n)$ if both u and v are good. Consequently, $\rho(M) \leq \widehat{C}\epsilon(n)\rho(\widehat{H})$.

Next we analyse the other terms in (4.37). We start with M^c . The idea is that, although now

$$\left| \frac{1}{\widehat{D}_u} \frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v} \right| = \mathcal{O}\left(\frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v}\right),$$

the total number of non-zero elements in a column of M^c is very small, so that its spectral radius indeed vanishes upon division by $\Delta(P)$. We note that

$$(A_{uv} \mathbf{1}_{\{v \text{ bad}\}})_{u,v} = (A_{uv} \mathbf{1}_{\{u \text{ bad}\}})_{u,v}^T,$$

so that a similar statement holds for the maximal row sum of M^r . Obviously, $M^{\text{cr}} \leq M^c$, and so do their spectral radii.

As a consequence of these observations, it thus suffices to prove our claim for M^c . To do so, we proceed in three steps: First we show that

$$\mathbb{P}(\mathcal{E}_1) = \mathbb{P}\left(\left\{\forall u : \left| \frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u} \right| \leq 1/2\right\}\right) \geq 1 - 2/n. \quad (4.38)$$

From which it follows after a short computation that, with probability larger than $1 - 2/n$, for all u, v ,

$$\left| \frac{1}{\widehat{D}_u} \frac{1}{\widehat{D}_v} - \frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v} \right| \leq 3 \frac{1}{\mathbb{E}\widehat{D}_u} \frac{1}{\mathbb{E}\widehat{D}_v}.$$

Keeping this in mind, it thus suffices to demonstrate that $(A_{uv} \mathbb{1}_{\{v \text{ bad}\}})_{uv}$ has a spectral radius much smaller than the spectral radius of A . The column sum in the former equals the number of bad neighbours a vertex has. That is, the spectral radius is bounded by $\max_u X_u$, where for $u \in V$,

$$X_u = \sum_{v \in \mathcal{N}(u)} Z_v, \quad (4.39)$$

with,

$$Z_v = \mathbb{1}_{\{v \text{ is bad}\}}.$$

Caution is needed here as the indicator functions in (4.39) are not independent.

In the second step we shall show that with high probability the number of edges between vertices in the neighbourhood of u is negligible compared to the expected degree of vertex u . That is,

$$\mathbb{P}(\mathcal{E}_2(u)) = \mathbb{P}\left(\left\{ \sum_{x,y \in \mathcal{N}(u)} A_{xy} \leq \frac{1}{2} \epsilon(n) \omega(n) \right\}\right) \geq 1 - 2/n^2, \quad (4.40)$$

where $\omega(n)$ is defined in (4.27). Hence, except for possibly $\frac{1}{4} \epsilon(n) \omega(n)$ of them, the variables in (4.39) form an independent set (conditional on not having any neighbours among $\mathcal{N}(u)$).

The last step consists in showing that this leads to

$$\mathbb{P}\left(X_u > \epsilon(n) \mathcal{O}\left(\mathbb{E}\widehat{D}_u\right) \mid \mathcal{E}_1, \mathcal{E}_2(u)\right) = o(1/n). \quad (4.41)$$

The assertion follows now straightforwardly: with high probability, we have

$$\begin{aligned} \sum_v M_{uv}^c &\leq 3 \frac{1}{\mathbb{E}\widehat{D}_u} \max_v \frac{1}{\mathbb{E}\widehat{D}_v} X_u \\ &\leq 3 \frac{1}{\mathbb{E}\widehat{D}_u} \max_v \frac{1}{\mathbb{E}\widehat{D}_v} \epsilon(n) \mathcal{O}\left(\mathbb{E}\widehat{D}_u\right) \\ &\leq \mathcal{O}\left(\frac{\epsilon(n)}{\min_v \mathbb{E}\widehat{D}_v}\right) \\ &= \mathcal{O}\left(\frac{\epsilon(n)}{\omega(n)}\right). \end{aligned}$$

Now $\overline{D} = \mathcal{O}(\omega(n))$, since $\frac{D_1^2(n)}{\overline{D}(n)} = \Omega(\log(n))$. Consequently, due to the choice of $\epsilon(n)$ in (4.36),

$$\rho(M^c) = \mathcal{O}\left(\frac{\epsilon(n)}{\omega(n)}\right) = \mathcal{O}\left(\frac{1}{\log^{1/3}(n)}\right) \frac{1}{\overline{D}(n)} = o_n(1) \frac{1}{\overline{D}(n)}.$$

The first step, i.e. demonstrating equation (4.38), is easily carried out: Fix $u \in V$ and use Bernstein's inequality (A.0.7) to verify the bound

$$\mathbb{P}\left(\left|\frac{\mathbb{E}\widehat{D}_u - \widehat{D}_u}{\mathbb{E}\widehat{D}_u}\right| > 1/2\right) \leq 2 \exp\left(-\frac{3}{26} \mathbb{E}\left[\widehat{D}_u\right]\right).$$

Now, for n large enough, $\mathbb{E}\widehat{D}_u \geq \overline{M}C_{B,M}\log n$, and by assumption, $C_{B,M}$ from (4.5) is so large that $\frac{3}{26}\mathbb{E}[\widehat{D}_u] > 2\log(n)$. Hence,

$$\mathbb{P}\left(1/2\mathbb{E}\widehat{D}_u \leq \widehat{D}_u \leq 3/2\mathbb{E}\widehat{D}_u\right) \geq 1 - 2/n^2.$$

We proceed with the second step, i.e., (4.40). Put $\overline{M} = \max_i \overline{M}_i$, $\overline{B} = \max_{i,j} B_{ij}$. Set $C = \max\{1/2\overline{M}, 5\overline{M}^2, \overline{B}\}$. Consider, conditional on $\widehat{D}_u \leq 2\mathbb{E}\widehat{D}_u$,

$$\begin{aligned} \sum_{x,y \in \mathcal{N}(u)} A_{xy} &= \sum_{x,y \in \mathcal{N}(u)} \text{Ber}\left(B_{\sigma_x \sigma_y} \frac{\phi_x \phi_y \omega(n)}{g(n)}\right) \\ &\leq \text{Bin}\left(4(\mathbb{E}\widehat{D}_u)^2, \overline{B} \frac{\phi_x \phi_y \omega(n)}{g(n)}\right) \\ &\leq \text{Bin}\left(5\overline{M}^2 \phi_u^2 \omega^2(n), \overline{B} \frac{\phi_x \phi_y \omega(n)}{g(n)}\right) \\ &\leq \text{Bin}\left(5\overline{M}^2 \phi_n^2 \omega^2(n), \overline{B} \frac{\phi_n^2 \omega(n)}{g(n)}\right) \\ &\leq \text{Bin}\left(C \phi_n^2 \omega^2(n), C \frac{\phi_n^2 \omega(n)}{g(n)}\right), \end{aligned}$$

where ϕ_u and $g(n)$ are defined in (4.26), respectively (4.28). We now show that

$$\mathbb{P}\left(\text{Bin}\left(C \phi_n^2 \omega^2(n), C \frac{\phi_n^2 \omega(n)}{g(n)}\right) \geq \frac{1}{2}\epsilon(n)\omega(n)\right) = o(1/n).$$

First, note that

$$\mathbb{P}\left(\text{Bin}\left(C \phi_n^2 \omega^2(n), C \frac{\phi_n^2 \omega(n)}{g(n)}\right) \geq \frac{1}{2}\epsilon(n)\omega(n)\right) \leq \left(\frac{C \phi_n^2 \omega^2(n)}{\frac{1}{2}\epsilon(n)\omega(n)}\right) \left(C \frac{\phi_n^2 \omega(n)}{g(n)}\right)^{\frac{1}{2}\epsilon(n)\omega(n)}. \quad (4.42)$$

Using that $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$, we have

$$\begin{aligned} \left(\frac{C \phi_n^2 \omega^2(n)}{\frac{1}{2}\epsilon(n)\omega(n)}\right) &\leq \left(2Ce \frac{\phi_n^2 \omega(n)}{\epsilon(n)}\right)^{\frac{1}{2}\epsilon(n)\omega(n)} \\ &= \exp\left(\frac{1}{2}\epsilon(n)\omega(n) \log\left(2Ce \frac{\phi_n^2 \omega(n)}{\epsilon(n)}\right)\right) \\ &\leq \exp\left(\frac{c}{2}\epsilon(n)\omega(n) \log(g(n)) + \frac{1}{2}\epsilon(n)\omega(n) \log(2Ce)\right), \end{aligned}$$

where $c < \frac{1}{2}$ from (4.5) is such that $\frac{\phi_n^2 \omega^2(n)}{\log^{2/3}(n)n^{c^2}} \rightarrow 0$ (and thus $\frac{\phi_n^2 \omega(n)}{g^c(n)} = \frac{o_n(1)}{\log^{1/3}(n)} \leq \epsilon(n)$, since $g(n) = \Theta(n)$ in the particular setting of this lemma). Write

$$\begin{aligned} \left(C \frac{\phi_n^2 \omega(n)}{g(n)}\right)^{\frac{1}{2}\epsilon(n)\omega(n)} &= \exp\left(-\frac{1}{2}\epsilon(n)\omega(n) \log\left((g(n))^{1-c} \frac{(g(n))^c}{C \phi_n^2 \omega(n)}\right)\right) \\ &\leq \exp\left(-\frac{1}{2}\epsilon(n)\omega(n) \log((g(n))^{1-c})\right) \\ &= \exp\left(-\frac{1-c}{2}\epsilon(n)\omega(n) \log(g(n))\right), \end{aligned}$$

if n large enough. Combining these estimates, we see that (4.42) may be bounded from above by

$$\begin{aligned} & \exp\left(-\frac{1-2c}{2}\epsilon(n)\omega(n)\log(g(n)) + \frac{1}{2}\epsilon(n)\omega(n)\log(2Ce)\right) \\ & \leq \left(-\frac{1-2c}{4}\epsilon(n)\omega(n)\log(g(n))\right), \end{aligned}$$

since $g(n) \geq n \gg 2Ce$. Finally, since $\frac{1-2c}{4}\epsilon(n)\omega(n) \geq 2$, for large n ,

$$\mathbb{P}(\mathcal{E}_2) = 1 - \mathbb{P}\left(\sum_{x,y \in \mathcal{N}(u)} A_{xy} \geq \frac{1}{2}\epsilon(n)\omega(n)\right) \geq 1 - e^{-\log(g^2(n))} \geq 1 - 1/n^2, \quad (4.43)$$

that is (4.40).

We proceed with the last step, i.e., establishing (4.41). Write,

$$X_u = \sum_{v \in \mathcal{N}(u): \mathcal{N}(v) \cap \mathcal{N}(u) \neq \emptyset} Z_v + \sum_{v \in \mathcal{N}(u): \mathcal{N}(v) \cap \mathcal{N}(u) = \emptyset} Z_v.$$

We already know from (4.40) that the first sum is smaller than $\frac{1}{2}\epsilon(n)\omega(n)$, with high probability. The variables in the second sum, $\{Z_v\}_{v \in \mathcal{N}(u): \mathcal{N}(v) \cap \mathcal{N}(u) = \emptyset}$, are independent. For such a vertex $v \in \mathcal{N}(u)$ that has no neighbour with u in common, we have $\widehat{D}_v = d'_v + 1$, where

$$d'_v = \sum_{x \notin \mathcal{N}(u), x \neq u} \text{Ber}\left(B_{\sigma_v \sigma_x} \frac{D_v D_x}{nD}\right),$$

the degree of v outside $\mathcal{N}(u) \cup \{u\}$. We show that v is a good vertex with high probability, by proving that d'_v concentrates on its mean which on its turn is close to $\mathbb{E}[\widehat{D}_v]$. Firstly, define

$$\mathbb{E}_*[\cdot] := \mathbb{E}\left[\cdot \mid \mathcal{N}(u), \mathcal{E}_2, \widehat{D}_u \leq 2\mathbb{E}[\widehat{D}_u]\right],$$

then

$$\begin{aligned} \mathbb{E}_*[d'_v] &= \sum_{x \notin \mathcal{N}(u), x \neq u} B_{\sigma_v \sigma_x} \frac{D_v D_x}{nD} \\ &= \sum_{x \neq v} B_{\sigma_v \sigma_x} \frac{D_v D_x}{nD} - \sum_{x \in \mathcal{N}(u) \cup \{u\}, x \neq v} B_{\sigma_v \sigma_x} \frac{D_v D_x}{nD} \\ &\geq \mathbb{E}[\widehat{D}_v] - \mathcal{O}\left(\frac{\phi_n^2 \omega(n)}{g(n)}\right) \mathbb{E}[\widehat{D}_u] \\ &= \mathbb{E}[\widehat{D}_v] - \mathcal{O}\left(\frac{\phi_n^3 \omega^2(n)}{g(n)}\right) \\ &= \mathbb{E}[\widehat{D}_v] - o_n(1). \end{aligned}$$

Secondly, we use Bernstein's inequality (A.0.7) to prove that d'_v concentrates around

$\mathbb{E} \left[\widehat{D}_v \right]$ upto a factor $\epsilon(n)$ as in (4.36):

$$\begin{aligned}
& \mathbb{P} \left(d'_v \geq (1 + \epsilon(n)) \mathbb{E} \widehat{D}_v \mid \mathcal{N}(u), \mathcal{E}_2, \widehat{D}_u \leq 2\mathbb{E} \left[\widehat{D}_u \right] \right) \\
& \leq \exp \left(- \frac{(\epsilon(n) \mathbb{E}_* d'_v + (1 + \epsilon(n)) o_n(1))^2}{2 (\mathbb{E}_* d'_v + 1/3 (\epsilon(n) \mathbb{E}_* d'_v + (1 + \epsilon(n)) o_n(1)))} \right) \\
& \leq \exp \left(- \frac{(\epsilon(n) \mathbb{E}_* d'_v)^2 \left(1 + \frac{o_n(1)}{\epsilon(n) \mathbb{E}_* d'_v} \right)}{4 \mathbb{E}_* d'_v} \right) \\
& \leq \exp \left(-C\epsilon^2(n) \log(n) \right),
\end{aligned}$$

where we redefined $C = \frac{1}{8}$. Similarly,

$$\mathbb{P} \left(d'_v \leq (1 - \epsilon(n)) \mathbb{E} \widehat{D}_v \mid \mathcal{N}(u), \mathcal{E}_2, \widehat{D}_u \leq 2\mathbb{E} \left[\widehat{D}_u \right] \right) \leq \exp \left(-C\epsilon^2(n) \log(n) \right).$$

Hence each vertex $v \in \mathcal{N}(u)$ that has no neighbour with u in common is thus a good vertex with probability $2\exp \left(-C\epsilon^2(n) \log(n) \right)$. Consequently, conditional on $\mathcal{N}(u), \mathcal{E}_2, \widehat{D}_u \leq 2\mathbb{E} \left[\widehat{D}_u \right]$,

$$\sum_{v \in \mathcal{N}(u): \mathcal{N}(v) \cap \mathcal{N}(u) = \emptyset} Z_v \leq \text{Bin} \left(2\mathbb{E} \widehat{D}_u, 2\exp \left(-C\epsilon^2(n) \log(n) \right) \right).$$

We have,

$$\begin{aligned}
& \mathbb{P} \left(\text{Bin} \left(2\mathbb{E} \widehat{D}_u, 2\exp \left(-C\epsilon^2(n) \log(n) \right) \right) \geq \frac{1}{2} \epsilon(n) \mathbb{E} \widehat{D}_u \right) \\
& \leq \left(\frac{2\mathbb{E} \widehat{D}_u}{\frac{1}{2} \epsilon(n) \mathbb{E} \widehat{D}_u} \right) (2\exp \left(-C\epsilon^2(n) \log(n) \right))^{\frac{1}{2} \epsilon(n) \mathbb{E} \widehat{D}_u} \\
& \leq \left(\frac{4e}{\epsilon(n)} \right)^{\frac{1}{2} \epsilon(n) \mathbb{E} \widehat{D}_u} (2\exp \left(-C\epsilon^2(n) \log(n) \right))^{\frac{1}{2} \epsilon(n) \mathbb{E} \widehat{D}_u} \\
& = \exp \left(\frac{1}{2} \epsilon(n) \mathbb{E} \widehat{D}_u \left(\log \frac{8e}{\epsilon(n)} - C\epsilon^2(n) \log(n) \right) \right) \\
& = o(1/n),
\end{aligned}$$

since $\epsilon(n) = 1/\log^{1/3}(n)$. Hence,

$$\mathbb{P} \left(X_u > \frac{1}{2} \epsilon(n) \left(\omega(n) + \mathbb{E} \widehat{D}_u \right) \mid \mathcal{E}_1, \mathcal{E}_2 \right) = o(1/n).$$

The last step ((4.41)) is completed by noting that $\omega(n) = \mathcal{O} \left(\mathbb{E} \widehat{D}_u \right)$. □

Proof of Lemma 4.6.6. All matrices in

$$W = (\widehat{H} - H) + (H - \mathbb{E}[H]) + (\mathbb{E}[H] - P),$$

are real and symmetric, hence, combining Lemmas 4.6.1 - 4.6.5,

$$\rho(W) \leq \rho(\widehat{H} - H) + \rho(H - \mathbb{E}[H]) + \rho(\mathbb{E}[H] - P) = o_n(1) \frac{1}{D(n)}.$$

Employing Lemma A.0.5 gives that to each eigenvector $\widehat{\mathbf{x}}$ of $\widehat{H} = P + W$ corresponds an eigenvector \mathbf{x} of P such that

$$\widehat{\mathbf{x}} \cdot \mathbf{x} \geq \sqrt{1 - \left(\frac{\rho(W)}{\Delta(P)}\right)^2} = 1 - \mathcal{O}\left(\left(\frac{\rho(W)}{\Delta(P)}\right)^2\right) = 1 - o_n(1),$$

since $\Delta(P) = \Omega(1/\overline{D}(n))$. \square

Proof of Lemma 4.6.7. Invoking Lemma 4.6.6, to each $\widehat{\mathbf{x}}_i$ (with eigenvalue $\widehat{\lambda}_i$) there exists a normed eigenvector \mathbf{x}_i (with eigenvalue λ_i) of P such that

$$\widehat{\mathbf{x}}_i \cdot \mathbf{x}_i = 1 - f_i(n), \quad (4.44)$$

with $f_i(n) = o_n(1)$. We claim that all λ_i are larger than zero (note that we refer here to a set of \widehat{L} eigenvalues). This can be seen as follows: From Lemma 4.6.1 we know that the first L eigenvalues of P are of order $1/\overline{D}$ and all other eigenvalues are zero. By Lemma A.0.5, $|\lambda_i - \widehat{\lambda}_i| \leq \rho(W) \ll 1/\overline{D}$, hence the first L eigenvalues of \widehat{H} are also of order $\Omega(1/\overline{D}) - \mathcal{O}(\rho(W)) = \Omega(1/\overline{D})$, and the other $n - L$ are of order $\mathcal{O}(\rho(W))$. Now, the \widehat{L} eigenvalues of \widehat{H} that are picked in Step 1 of Algorithm 1 are precisely those whose absolute eigenvalue exceeds $f(n)/\widehat{D}_{\text{average}} = \Omega(f(n)/\overline{D}) \gg \rho(W)$, by construction of f in Section 3. Hence those eigenvalues must necessarily be of order $\Omega(1/\overline{D})$ (i.e., they are indeed non-zero) and $L = \widehat{L}$ with high probability.

Since \mathbf{x}_i corresponds to a non-zero eigenvalue, it follows from the proof of Lemma 4.6.1 that \mathbf{x}_i is constant on each block, i.e., $\mathbf{x}_i(u) = \mathbf{x}_i(v)$ if $\sigma_u = \sigma_v$. Let $x_i^{(k)}$ be the value of \mathbf{x}_i on block $k \in S$. Put

$$t_k = \sqrt{n}(x_1^{(k)}, \dots, x_L^{(k)}). \quad (4.45)$$

Then,

$$\begin{aligned} 1/n |\{u \in V : \|\sqrt{n}\widehat{z}_u - t_{\sigma_u}\|^2 \geq T^2\}| &\leq \frac{1}{nT^2} \sum_{m=1}^n \|\sqrt{n}\widehat{z}_u - t_{\sigma_u}\|^2 \\ &= 1/T^2 \sum_{u=1}^n \|(\widehat{x}_1^{(u)}, \dots, \widehat{x}_L^{(u)}) - (x_1^{(\sigma_u)}, \dots, x_L^{(\sigma_u)})\|^2 \\ &= 1/T^2 \sum_{k=1}^L \|\widehat{x}_k - x_k\|^2 \\ &= 1/T^2 \sum_{k=1}^L f_k(n), \end{aligned}$$

to finish the proof, let $T = \left(\sum_{k=1}^L f_k(n)\right)^{1/3} = \mathcal{O}\left(\left(\frac{\rho(W)}{\Delta(P)}\right)^{2/3}\right) = o_n(1)$. \square

Proof of Lemma 4.6.8. Below we shall make a spectral decomposition in terms of L orthonormal eigenvectors of Z that span the union of all eigenspaces corresponding to non-zero eigenvalues. Recall from the proof of Lemma 4.6.1 how we can obtain the eigenvectors of Z from the eigenvectors of P .

Recall that by construction $\{\widehat{\mathbf{x}}_i\}_{i=1}^L$ are orthonormal eigenvectors of \widehat{H} corresponding to non-zero eigenvalues spanning an L dimensional space. Recall further from the proof of Lemma 4.6.7 that the corresponding eigenvectors $\{\mathbf{x}_i\}_{i=1}^L$ of P

are associated with non-zero eigenvalues. Lemma A.0.5 (ii) entails that the space spanned by those $\{\mathbf{x}_i\}_{i=1}^L$ has also dimension L . And Lemma 4.6.6 implies that $\{\mathbf{x}_i\}_{i=1}^L$ become an orthonormal set for n tending to infinity (because they become more and more aligned with the orthonormal set $\{\widehat{\mathbf{x}}_i\}_{i=1}^L$).

Let, as in the proof of Lemma 4.6.7, $x_i^{(k)}$ be the value of \mathbf{x}_i on block $k \in S$. Note that $\sum_k n \alpha_k (x_i^{(k)})^2 = 1$ for $i \in \{1, \dots, L\}$. Putting $y_i = \sqrt{n}(x_i^{(1)}, \dots, x_i^{(K)})^T$, we see that each y_i is a *normalized* eigenvector of Z in the sense that $\sum_k \alpha_k (y_i(k))^2 = 1$. Now, assume for a contradiction that $|t_k - t_l| \rightarrow 0$ as $n \rightarrow \infty$:

$$\sum_{i=1}^L |\sqrt{n}x_i^{(k)} - \sqrt{n}x_i^{(l)}|^2 = \sum_{i=1}^L |y_i(l) - y_i(k)|^2 \rightarrow 0. \quad (4.46)$$

We conclude that there exist orthonormal eigenvectors of Z , $\{\bar{y}_1, \dots, \bar{y}_L\}$ (with eigenvalues $\{\lambda_i\}_{i=1}^L$ after a possible relabelling of indices), that span the range of Z , such that

$$\bar{y}_u(k) = \bar{y}_u(l)$$

for all u . The other $K - L$ eigenvectors have zero as an eigenvalue.

To proceed, consider matrix

$$N = \left(\sqrt{\alpha_u} \frac{B_{uv}}{M_u M_v} \sqrt{\alpha_v} \right)_{u,v}.$$

If $(x(1), \dots, x(K))^T$ is an eigenvector of Z then $(\sqrt{\alpha_1}x(1), \dots, \sqrt{\alpha_K}x(K))^T$ is an eigenvector of N , as is easily verified. Hence N has $\{(\sqrt{\alpha_1}\bar{y}_i(1), \dots, \sqrt{\alpha_K}\bar{y}_i(K))^T\}_{i=1}^L$ as eigenvectors corresponding to non-zero eigenvalues and $K - L$ eigenvectors with 0 as eigenvalue (which do not contribute to the spectral decomposition of N). Hence

$$N = \left(\sum_{l=1}^L \sqrt{\alpha_u} \bar{y}_l(u) \lambda_l \sqrt{\alpha_v} \bar{y}_l(v) \right)_{u,v}.$$

Thus, for all u ,

$$\frac{B_{ku}}{M_k M_u} = \sum_m \bar{y}_m(k) \lambda_m \bar{y}_m(u) = \sum_m \bar{y}_m(l) \lambda_m \bar{y}_m(u) = \frac{B_{lu}}{M_l M_u},$$

violating assumption 4.25. \square

4.7.2 Comparison to spectral analysis on the adjacency matrix

Proof of Theorem 4.4.1. This proof leans strongly on ideas borrowed from [90], where graphs *without* a community-structure are considered. Parts of their proof carry through for the DC-SBM considered here. Note that $\lim_{n \rightarrow \infty} g(n)/n = 1$.

By definition, we require without loss of generality $D_1 \leq D_2 \leq \dots \leq D_n$. However, we obtain the same graph (with now a decreasing degree-sequence) by a rearrangement of indices, if we put

$$\phi_u = \begin{cases} \frac{\phi_1}{u} & \text{if } u \leq 1 \leq k = n^\beta \\ 1 & \text{if } u > n^\beta, \end{cases} \quad (4.47)$$

where $\phi_1 = n^{\gamma+\beta}$, and $D_u = \phi_u \omega(n)$ (with ω as in (4.27)).

$$\sigma_u = \begin{cases} 1 & \text{if } u \leq \frac{n}{2} \\ 2 & \text{if } u > \frac{n}{2}. \end{cases} \quad (4.48)$$

Denote a sample of the random graph by G . We decompose G into the following graphs (exactly as in [90]) :

- G_1 , which is a union of vertex disjoint stars S_1, \dots, S_k . Star S_u has as its center node u and as leaves those vertices from among $\{k+1, \dots, n\}$ adjacent to u , but not adjacent to $\{1, \dots, u-1\}$;
- G'_1 is the graph consisting of all edges of G with one endpoint in $\{1, \dots, k\}$ and the other endpoint in $\{k+1, \dots, n\}$, except for those edges in G_1 ;
- G_2 is the subgraph of G , which is induced by $\{1, \dots, k\}$;
- G_3 is the subgraph of G , which is induced by $\{k+1, \dots, n\}$.

Further, let F_u be the subset of vertices in $\{k+1, \dots, n\}$ that are adjacent to $\{1, \dots, u-1\}$ and let C be a constant, independent of n , whose value might change along the course of the proof.

We claim that \widehat{d}_u , the degree of vertex u in G_1 , concentrates around its mean. Indeed, consider

$$\widehat{d}_u = \sum_{l=k+1}^n \text{Ber} \left(\frac{D_u D_l}{g(n)\omega(n)} B_{\sigma_u \sigma_l} \right) - \sum_{l \in F_u} \text{Ber} \left(\frac{D_u D_l}{g(n)\omega(n)} B_{\sigma_u \sigma_l} \right),$$

where g is defined in (4.28). Then,

$$d_u = \mathbb{E} [\widehat{d}_u] \geq \frac{\omega(n)\phi_u}{g(n)} \left(\sum_{l=k+1}^n B_{\sigma_u \sigma_l} - C \mathbb{E} [|F_u|] \right),$$

which we bound from below by estimating $\mathbb{E} [|F_u|]$, for $u \leq k = n^\beta$: For large enough n ,

$$\begin{aligned} \mathbb{E} [|F_u|] &= \sum_{l=k+1}^n \sum_{v=1}^{u-1} \frac{D_l D_v}{\omega(n)g(n)} B_{\sigma_u \sigma_l} \\ &\leq C \frac{\omega(n)\phi_1}{g(n)} \sum_{l=k+1}^n \sum_{v=1}^{u-1} \frac{1}{v} \\ &\leq C \omega(n)\phi_1 \frac{n - n^\beta}{g(n)} n^\beta \\ &\leq C \omega(n) n^{\gamma+2\beta}, \end{aligned}$$

after recalling the special choice for the degree sequence.

Consequently, we have

$$\frac{n}{g(n)} \frac{B_{11} + B_{12}}{2} D_u \geq d_u \geq \frac{B_{11} + B_{12}}{2} D_u \left(\frac{n}{g(n)} - C \frac{\omega(n)n^{\gamma+2\beta}}{g(n)} \right).$$

Invoking large deviation theory on \widehat{d}_u (which is a sum of Bernoulli random variables), we deduce that

$$\mathbb{P} \left(|\widehat{d}_u - d_u| > \sqrt{c' d_u \log n} \right) \leq 2/n^{c'/4}, \quad (4.49)$$

For $c' > 0$ a constant. We take $c' = 8$ to establish (4.49) uniformly over all vertices. We next investigate $\Delta(G_1)$, the smallest gap between different eigenvalues of G_1 . This graph is the union of vertex disjoint stars with degree \widehat{d}_u so that its spectrum is given by

$$\{\pm \sqrt{\widehat{d}_1 - 1}, \dots, \pm \sqrt{\widehat{d}_k - 1}\}.$$

We claim that

$$\Delta(G_1) \geq C\sqrt{\omega(n)}n^{\frac{\gamma-3\beta}{2}} \rightarrow \infty \quad (4.50)$$

with high probability. Indeed, define

$$x_u^\pm = d_u \pm \sqrt{c'd_u \log n},$$

and note that with high probability $\widehat{d}_u \geq x_u^-$ and $\widehat{d}_{u+1} \leq x_{u+1}^+$. To investigate the difference $x_u^- - x_{u+1}^+$, we first bound $d_u - d_{u+1}$ from below:

$$\begin{aligned} d_u - d_{u+1} &\geq \frac{B_{11} + B_{12}}{2} \omega(n) \phi_1 \left(\frac{n/g(n)}{u(u+1)} - C \frac{\omega(n)n^{\gamma+2\beta}}{n} \frac{1}{u} \right) \\ &= \frac{B_{11} + B_{12}}{2} \omega(n) \phi_1 \frac{1}{u} \left(\frac{n/g(n)}{u+1} - C \frac{\omega(n)n^{\gamma+2\beta}}{n} \right) \\ &\geq \frac{B_{11} + B_{12}}{2} \omega(n) \phi_1 \frac{1}{u} \left(\frac{n/g(n)}{n^\beta + 1} - C \frac{\omega(n)n^{\gamma+2\beta}}{n} \right) \\ &\geq \frac{B_{11} + B_{12}}{4} \omega(n) \phi_1 \frac{1}{n^\beta} \frac{1}{n^\beta} \\ &= \frac{B_{11} + B_{12}}{4} \frac{\omega(n)n^{\gamma+\beta}}{n^{2\beta}} \\ &= \frac{B_{11} + B_{12}}{4} \omega(n)n^{\gamma-\beta}. \end{aligned}$$

Next we show that the $\sqrt{d_u \log n}$ terms are negligible:

$$\begin{aligned} \sqrt{d_u \log n} &\leq \sqrt{\frac{B_{11} + B_{12}}{2} n/g(n) D_u \log n} \\ &\leq C \sqrt{\omega(n) \log(n) n^{\gamma+\beta}} \\ &\leq C \omega(n) n^{\frac{\gamma+\beta}{2}} \\ &\ll \omega(n) n^{\gamma-\beta}, \end{aligned}$$

due to (4.12). Hence,

$$x_u^- - x_{u+1}^+ \geq C \omega(n) n^{\gamma-\beta}.$$

As a consequence,

$$\begin{aligned} \Delta(G_1) &\geq \min_{u \in \{1, \dots, k\}} \left(\sqrt{\widehat{d}_u - 1} - \sqrt{\widehat{d}_{u+1} - 1} \right) \\ &\geq \min_{u \in \{1, \dots, k\}} \left(\sqrt{x_u^- - 1} - \sqrt{x_{u+1}^+ - 1} \right) \\ &= \min_{u \in \{1, \dots, k\}} \left(\frac{x_u^- - x_{u+1}^+}{\sqrt{x_u^- - 1} + \sqrt{x_{u+1}^+ - 1}} \right) \\ &\geq C \frac{\omega(n)n^{\gamma-\beta}}{\sqrt{\omega(n)}n^{\frac{\gamma+\beta}{2}}} \\ &= C\sqrt{\omega(n)}n^{\frac{\gamma-3\beta}{2}}, \end{aligned}$$

that is (4.50).

We continue with an inspection of G'_1 , that is, we focus on $\widehat{m}_u = \widehat{D}_u|G'_1$, the degree of vertex u in G'_1 , and show that

$$\widehat{m}_u \leq 2c'\log n, \quad (4.51)$$

with high probability (here, m_u is the expectation of \widehat{m}_u). We shall use this in combination with the fact that the spectral radius of a graph is bounded by its largest degree.

Write

$$\widehat{m}_u = \sum_{l \in F_u} \text{Ber} \left(\frac{\phi_u \omega(n)}{g(n)} B_{1\sigma_l} \right).$$

This expression allows us to deduce an upperbound for m_u ,

$$m_u = \mathbb{E}[\widehat{m}_u] \leq C \mathbb{E}[F_u] \frac{\phi_u \omega(n)}{g(n)} \leq C \omega(n) n^{\gamma+2\beta} n^{\gamma+\beta} \frac{1}{u} \frac{\omega(n)}{g(n)} \leq C \omega^2(n) \frac{n^{2\gamma+3\beta}}{n},$$

which tends to zero due to (4.11). Standard bounds for Bernoulli random variables give

$$\mathbb{P}(|m_u - \widehat{m}_u| \leq c'\log n) \leq 2 \exp \left(-\frac{(c'\log n)^2}{2(m_u + c'\log(n)/3)} \right) \leq 2 \exp \left(-\frac{1}{4} c'\log n \right) = \frac{2}{n^{c'/4}}.$$

We conclude that, with probability at least $1 - \frac{2}{n^2}$,

$$\widehat{m}_u \leq m_u + c'\log n \leq 2c'\log n,$$

i.e., (4.51) holds. An identical estimate holds when $u > k$.

We next bound the number of edges in G_2 , denoted by $E(G_2)$. The square root of $E(G_2)$ is an upper bound for the spectral radius of G_2 .

$$\mathbb{E}[|E(G_2)|] = C \sum_{u=1}^k \sum_{v=1}^k \frac{\phi_u \phi_v \omega(n)}{g(n)} \leq C \frac{n^{\gamma+\beta} n^{\gamma+\beta} \omega(n) n^\beta n^\beta}{g(n)} \leq C \frac{n^{2\gamma+4\beta}}{n},$$

vanishing for large n . Again, upon invoking standard large deviation theory, we have, with probability at least $1 - \frac{2}{n^2}$,

$$\mathbb{E}[|E(G_2)|] \leq 2c'\log n. \quad (4.52)$$

Consider the degree of a vertex $u > k$ in G_3 ,

$$\mathbb{E}[\widehat{D}_u|G_3] = \sum_{v=k}^n \frac{\phi_u \phi_v \omega(n)}{g(n)} B_{\sigma_v \sigma_u} \leq C \frac{\omega(n)}{g(n)} n \leq C \omega(n).$$

Hence,

$$\mathbb{P} \left(\widehat{D}_u|G_3 > C \omega(n) + \sqrt{c' \log(n) C \omega(n)} \right) \leq \frac{2}{n^{c'/4}}. \quad (4.53)$$

Combining these observations leads to our assertion that the first k eigenvectors of A become undistinguishable of those of the k stars, when n tends to infinity. Indeed, split A according to the described graph-composition:

$$A = A|_{G_1} + A|_{G'_1} + A|_{G_2} + A|_{G_3},$$

and note that the spectral radii of $A|_{G'_1}$, $A|_{G_2}$ and $A|_{G_3}$ vanish in the presence of $\Delta(G_1)$. This follows because (as mentioned above) for any graph its spectral radius

is bounded by the minimum of its largest degree and the square root of its number of edges. Hence, due to (4.51) - (4.53),

$$\rho(A|_{G'_1}) \leq 2c'l \log n,$$

$$\rho(A|_{G_2}) \leq \sqrt{2c'l \log n},$$

and

$$\rho(A|_{G_3}) \leq C\omega(n) + \sqrt{c'l \log(n) C\omega(n)},$$

with high probability. All those three bounds vanish indeed upon division by $\Delta(G_1) \geq C\omega(n)n^{\frac{\gamma-3\beta}{2}}$. Lemma A.0.5 finishes the proof. \square

4.7.3 Interpretation of the conditions

Proof of Remark 4.5.1. Assume,

$$\frac{B_{ij}}{M_i} = \frac{B_{lj}}{M_l}$$

then,

$$\frac{B_{ij}}{M_i M_j} = \frac{B_{lj}}{M_l M_j}. \quad (4.54)$$

Now, put $\bar{\phi}_i = \frac{1}{\alpha_i n} \sum_{\sigma_u=i} \phi_u$, then

$$\frac{B_{ij}}{M_i M_j} = \frac{\alpha_i \bar{\phi}_i B_{ij} \alpha_j \bar{\phi}_j}{\alpha_i \bar{\phi}_i M_i M_j \alpha_j \bar{\phi}_j}.$$

We give a probabilistic interpretation to the terms appearing in the denominator:

$$\begin{aligned} n\alpha_i \bar{\phi}_i M_i &= n\alpha_i \bar{\phi}_i \sum_{k=1}^K \sum_{u:\sigma_u=k} \phi_u B_{i\sigma_u} \\ &= n\alpha_i \bar{\phi}_i \sum_{k=1}^K n\alpha_k \bar{\phi}_k B_{ik} \\ &= \sum_{k=1}^K (n\alpha_i)(n\alpha_k) \bar{\phi}_i B_{ik} \bar{\phi}_k \\ &= \sum_{k=1}^K \sum_{u:\sigma_u=i} \phi_u \sum_{v:\sigma_v=k} \phi_v B_{ik} \\ &= \frac{n}{\omega(n)} \sum_{k=1}^K \sum_{u:\sigma_u=i} \sum_{v:\sigma_v=k} \mathbb{P}(l \leftrightarrow m) \\ &= \frac{n}{\omega(n)} \sum_{u:\sigma_u=i} \sum_{m=1}^n \mathbb{P}(l \leftrightarrow m) \\ &= \frac{n}{\omega(n)} \{\text{expected total degree of vertices in community } i\}. \end{aligned} \quad (4.55)$$

An inspection of the numerator reveals

$$\begin{aligned}
n\alpha_i\bar{\phi}_i B_{ij}\bar{\phi}_j\alpha_j n &= \sum_{u:\sigma_u=i} \phi_u \sum_{v:\sigma_v=j} \phi_v B_{ij} \\
&= \frac{n}{\omega(n)} \sum_{u:\sigma_u=i} \sum_{v:\sigma_v=j} \mathbb{P}(u \leftrightarrow v) \\
&= \frac{n}{\omega(n)} \{\text{expected \#edges between community } i \text{ and } j\}
\end{aligned} \tag{4.56}$$

□

Proof of Lemma 4.5.3. Assume first that for some i and l we have for all j

$$\widehat{B}_{ij} = \widehat{B}_{lj}$$

and, for all u, v ,

$$\frac{\phi_u B_{\sigma_u \sigma_v} \phi_v}{g(n)} = \frac{\widehat{\phi}_u \widehat{B}_{\sigma_u \sigma_v} \widehat{\phi}_v}{\widehat{g}(n)},$$

with ϕ_u defined in (4.26) and g in (4.28) ($\widehat{\phi}_u$ and \widehat{g} are defined analogously). Fix j . Let α, β and γ be *any* indices such that $\sigma_\alpha = i, \sigma_\beta = j$ and $\sigma_\gamma = l$. Then,

$$\frac{\widehat{\phi}_\alpha \widehat{B}_{ij} \widehat{\phi}_\beta}{\widehat{g}(n)} = \frac{\phi_\alpha B_{ij} \phi_\beta}{g(n)} \Rightarrow B_{ij} = \frac{\widehat{\phi}_\alpha \widehat{\phi}_\beta g(n)}{\phi_\alpha \phi_\beta \widehat{g}(n)} \widehat{B}_{ij}$$

and

$$\frac{\widehat{\phi}_\gamma \widehat{B}_{lj} \widehat{\phi}_\beta}{\widehat{g}(n)} = \frac{\phi_\gamma B_{lj} \phi_\beta}{g(n)} \Rightarrow B_{lj} = \frac{\widehat{\phi}_\gamma \widehat{\phi}_\beta g(n)}{\phi_\gamma \phi_\beta \widehat{g}(n)} \widehat{B}_{lj},$$

implying that (since $\widehat{B}_{ij} = \widehat{B}_{lj}$)

$$B_{ij} = \frac{\widehat{\phi}_\alpha \widehat{\phi}_\beta g(n)}{\phi_\alpha \phi_\beta \widehat{g}(n)} \widehat{B}_{lj} = \frac{\widehat{\phi}_\alpha \phi_\gamma}{\phi_\alpha \widehat{\phi}_\gamma} B_{lj}.$$

Since j was arbitrary, there exist c such that for all j

$$B_{ij} = cB_{lj},$$

hereby violating the identifiability condition, as pointed out in Remark 4.5.2, i.e.,

$$\frac{B_{ij}}{M_i} = \frac{B_{lj}}{M_l},$$

for all j .

Now assume that (a) holds, that is

$$\frac{B_{ij}}{M_i} = \frac{B_{lj}}{M_l},$$

for all j . Define for $k, l \in S$ and $u \in V$

$$\widehat{B}_{kl} = \frac{1}{M_k} \frac{1}{M_l} B_{kl}$$

and

$$\widehat{\phi}_u = f(n)\phi_u M_{\sigma_u},$$

where

$$f(n) = \frac{\sum_v \phi_v M_{\sigma_v}}{\sum_w \phi_w}.$$

Then,

$$\widehat{B}_{ij} = \frac{1}{M_i} \frac{1}{M_j} B_{ij} = \frac{1}{M_i} \frac{1}{M_j} \frac{M_i}{M_l} B_{lj} = \frac{1}{M_j} \frac{1}{M_l} B_{lj} = \widehat{B}_{lj},$$

and (as above, we define \widehat{g} analogously to g),

$$\begin{aligned} \frac{\widehat{\phi}_u \widehat{B}_{\sigma_u \sigma_v} \widehat{\phi}_v}{\widehat{g}(n)} &= \frac{1}{\sum_w f(n) \phi_w M_{\sigma_w}} \phi_u M_{\sigma_u} f(n) \frac{B_{\sigma_u \sigma_v}}{M_{\sigma_u} M_{\sigma_v}} f(n) M_{\sigma_v} \phi_v \\ &= \frac{\phi_u B_{\sigma_u \sigma_v} \phi_v}{\sum_w \phi_w} \\ &= \frac{\phi_u B_{\sigma_u \sigma_v} \phi_v}{g(n)}. \end{aligned}$$

□

4.8 Directions for future research

4.8.1 Exact recovery

The obtained clustering here is almost-exact: only a vanishing fraction of nodes is miss-classified. It is plausible that an exact clustering could be obtained from this clustering, by using it as input to the "clean-up" algorithm presented in Section 7.2 of [1] or alternatively, Algorithm 2 in [93].

4.8.2 Non-constant B

In the underlying paper we assumed B to be a constant matrix. The current analysis could be extended to a setting where B is allowed to change with n . We need however the existence of a constant $\delta > 0$ such that for all n , $\Delta(Z) \geq \delta$ for \widehat{H} to concentrate. For identifiability we need the existence of some $\epsilon > 0$ such that for all i, j and n , $\max_{i'} \left| \frac{B_{ii'}}{M_i M_{i'}} - \frac{B_{ji'}}{M_j M_{i'}} \right| \geq \epsilon$.

4.8.3 Sparser graphs

The main issue with both the normalized adjacency matrix and the Laplacian is proving when those matrices concentrate around a deterministic matrix. For the Laplacian, if the degrees are of order $\Omega(\log(n))$, matrices concentrate according to [24]. But, if the minimum degree is of order $o(\log(n))$, the graph is seen to have some isolated vertices. Those contribute to multiple zeros in the spectrum: hence the matrix does not concentrate. There are multiple ways to overcome this issue, for instance removing the low-degree vertices or raising all the degrees. The latter strategy is proposed in [75] for the inhomogeneous Erdos-Renyi random graph (where edges are independently present with probabilities $(p_{uv})_{u,v=1}^n$) and also in [106, 21] (see Section 4.4.4) for the DC-SBM. According to [75], for $\tau \sim d$, with $d = n \max_{uv} p_{uv}$, with high probability,

$$\rho \left(L_\tau - \left(\mathbb{E}[D_\tau]^{-1/2} \mathbb{E}[A] \mathbb{E}[D_\tau]^{-1/2} \right) \right) = \mathcal{O} \left(\frac{1}{\sqrt{d}} \right),$$

where L_τ is defined in (4.18).

Based on these observations, it might be fruitful to use \widehat{H} on a graph where the degrees have been artificially inflated.

Chapter 5

Information-theoretical limits for the *sparse* DC-SBM

5.1 Introduction

As we alluded to in the last chapter, spectral methods break down when the graph becomes too sparse. In the sparse *ordinary* SBM (with parameters a and b) it is known [92] that reconstructing the communities from a single observation of the random graph is information-theoretically impossible when $(a - b)^2 \leq 2(a + b)$.

In this chapter we extend results in [92] to the DC-SBM (with parameters a, b and $\Phi^{(2)}$): We prove (using some ideas from [92]) that when $(a - b)^2 \Phi^{(2)} \leq 2(a + b)$, it is information-theoretically impossible to estimate the spins in a way *positively correlated* with the true community structure based only on a single observation of the graph without knowing the weights. Note the *universality* of this result: the threshold depends only on the weight distribution through its second moment.

5.2 Main results

Recall the definition of the DC-SBM (with parameters a, b and weight distribution ν) from Section 3, and denote its i.i.d. spins by $\{\sigma_u\}_u$, its i.i.d. weights by $\{\phi_u\}_u \sim \nu$ and the k -th moment of ν by $\Phi^{(k)}$. We assume that the weights are possibly heavy-tailed with exponent $\beta > 8$: for all large enough k ,

$$\mathbb{P}(\phi_1 \geq k) = \nu([k, \infty)) \leq \frac{1}{k^\beta}.$$

Before we precisely state our results, we define the notion of a positively correlated reconstruction, which is less stringent than quasi-exact recovery. The latter is not feasible since a positive fraction of the nodes are isolated, for which random guess is the only way to reconstruct them.

Definition 5.2.1. *Let G be an observation of the DC-SBM, with true communities $\{\sigma_u\}_{u=1}^n$. Further, let $\{\hat{\sigma}_u\}_{u=1}^n$ be a reconstruction of the communities, based on the observation G . Then, we say that $\{\hat{\sigma}_u\}_{u=1}^n$ is positively correlated with the true partition $\{\sigma_u\}_{u=1}^n$ if there exists $\delta > 0$ such that*

$$\mathbb{P}\left(\frac{1}{n} \sum_{u=1}^n 1_{\{\sigma_u = \hat{\sigma}_u\}} \geq \frac{1}{q} + \delta\right) \rightarrow 1,$$

as $n \rightarrow \infty$.

The main result in this chapter is:

Theorem 5.2.2. *Assume that $(a - b)^2\Phi^{(2)} \leq q(a + b)$. Let G be an instance of the DC-SBM. Let u and v be uniformly chosen vertices in G . Then, for any $s \in \{1, \dots, q\}$,*

$$\mathbb{P}(\sigma_u = s | \sigma_v, G) \xrightarrow{\mathbb{P}} \frac{1}{q}, \quad (5.1)$$

as $n \rightarrow \infty$.

Thus, it is already impossible to estimate the spin of a random vertex given the spin of another vertex, which is an easier problem than reconstructing the group membership of strictly more than a fraction $1/q$ of the vertices (as explained in Lemma 5.8.2):

Theorem 5.2.3. *Let G be an observation of the DC-SBM with $(a - b)^2\Phi^{(2)} \leq q(a + b)$. Then, no reconstruction $\{\hat{\sigma}_u\}_{u=1}^n$ based on G is positively correlated with $\{\sigma_u\}_{u=1}^n$.*

We further introduce the following shorthand notation: $\sigma = (\sigma_1, \dots, \sigma_n)$ and $\phi = (\phi_1, \dots, \phi_n)$. For a subset $U \subset \{1, \dots, n\}$ of the vertices, we define $\sigma_U = \{\sigma_u\}_{u \in U}$ and $\phi_U = \{\phi_u\}_{u \in U}$. For a vertex $\rho \in V$, and integer $r \geq 0$, we denote by $G_r(\rho)$ the r -neighbourhood of ρ .

5.3 Proof heuristics

We argue now heuristically how one could obtain Theorem 5.2.2, its formal proof is deferred to the upcoming sections. For simplicity we restrict to the setting of *two communities*. The idea is to consider an even simpler problem: if we know the spins of all vertices at distance R away from u can we then deduce the sign of u ? By noting that local neighbourhoods are w.h.p. tree-like, we can relate the DC-SBM to a reconstruction problem on trees. In a simplified form (see also [91]) this tree-reconstruction problem is given as follows: Consider a tree, where every node has $d = \frac{a+b}{2}\Phi^{(2)}$ descendants and either a blue or a red color. With probability $p = \frac{a-b}{a+b}$ a descendent inherits its parent's color, and with probability $1 - p$ its color is chosen uniformly at random. Given the colors of vertices at distance R away from the root, can we deduce the root's color? We could use majority vote among the far-away vertices, but when is this informative? On average, $(dp)^R$ vertices have a color that has been copied along the chain from the root, the other vertices have a random (non-informative) color. Among the latter set, roughly half of them are red and half of them are blue, upto some fluctuations of size $\sqrt{d^R}$ (law of large numbers). Hence the fluctuations become so dominant that the information is lost when $\sqrt{d^R} \gtrsim (dp)^R$, that is, $(a - b)^2\Phi^{(2)} \leq 2(a + b)$.

We need however to make the following precise: We should show that long-range interactions are weak, so that the spin of v is indeed "shielded" away from u by the boundary spins at distance R . Further, we need a careful analysis to couple local neighbourhoods to q -type branching processes where the offspring distribution follows a Poisson-mixture (i.e., local neighbourhoods are irregular). We finally need to study broadcasting on these general processes, which can be understood in terms of their *branching number* [41].

5.4 Description of q -type Poisson-mixture branching process

The branching process that we describe here plays a central role both in this chapter and in the chapter on the non-backtracking matrix. Before introducing it, we start with some heuristics.

Consider two vertices u and v connected by an edge (which we denote as $u \sim v$). What can we say about their spins and weights? Assume for simplicity that we consider the setting of only *two* types and that the weights are *discrete* random variables and let $\psi_u, \psi_v \in \bar{W}$, then

$$\begin{aligned}
& \mathbb{P}(\sigma_u = \sigma_v, \phi_u = \psi_u, \phi_v = \psi_v | u \sim v) \\
&= \frac{\mathbb{P}(u \sim v | \sigma_u = \sigma_v, \phi_u = \psi_u, \phi_v = \psi_v) \mathbb{P}(\sigma_u = \sigma_v, \phi_u = \psi_u, \phi_v = \psi_v)}{\mathbb{P}(u \sim v)} \\
&= \frac{\frac{a\psi_u\psi_v}{n} \cdot \frac{1}{2} \cdot d\nu(\psi_u) \cdot d\nu(\psi_v)}{\sum_{\hat{\psi}_u, \hat{\psi}_v} \hat{\psi}_u \hat{\psi}_v \frac{a+b}{2} d\nu(\hat{\psi}_u) d\nu(\hat{\psi}_v)} \\
&= \frac{a}{a+b} \cdot \frac{\psi_u d\nu(\psi_u)}{\Phi^{(1)}} \cdot \frac{\psi_v d\nu(\psi_v)}{\Phi^{(1)}}.
\end{aligned} \tag{5.2}$$

In other words, the conditional distribution of $1_{\{\sigma_u = \sigma_v\}}$, ϕ_u and ϕ_v factorizes into three *independent* parts. Note the size-bias in the weights: knowing that a vertex is not isolated makes it more likely to have a larger weight.

We thus expect the neighbours of a vertex to have their weights distributed according to the biased law $\frac{y d\nu(y)}{\Phi^{(1)}}$. We also expect this to hold approximately through the first so many generations. Indeed if n is large then the vertices *not* included in a neighbourhood have their weights only slightly biased with respect to ν (because w.h.p. a vertex is not included in the neighbourhood anyway, as edge-probabilities scale with $n^{-3/4}$), so that we can apply the above inductively.

This leads us to the introduction of a branching process denoted by T^{Poi} and defined as follows. We begin with a single particle, the root o , having spin $\sigma_o \in \{1, \dots, q\}$ and weight $\phi_o \in W \subset [\phi_{\min}, \infty)$ (which we take random). The root is replaced in generation 1 by $\text{Poi}\left(\frac{a}{q}\Phi^{(1)}\phi_o\right)$ particles of spin σ_o and by $\text{Poi}\left(\frac{b}{q}\Phi^{(1)}\phi_o\right)$ particles of spin s for each $s \in \{1, \dots, q\} \setminus \sigma_o$. Further, the weights of those particles are i.i.d. distributed following law ν^* , the size-biased version of ν , defined for $x \in [\phi_{\min}, \infty)$ by

$$\nu^*([0, x]) = \frac{1}{\Phi^{(1)}} \int_{\phi_{\min}}^x y d\nu(y). \tag{5.3}$$

For generation $t \geq 1$, a particle with spin σ and weight ϕ^* is replaced in the next generation by $\text{Poi}\left(\frac{a}{2}\Phi^{(1)}\phi^*\right)$ particles with the same spin and $\text{Poi}\left(\frac{b}{q}\Phi^{(1)}\phi_o\right)$ particles of each of the remaining $q-1$ spins. Again, the weights of the particles in generation $t+1$ follow in an i.i.d. fashion the law ν^* . The offspring-size of an individual is thus a **Poisson-mixture** with mean $\frac{a+(q-1)b}{q}\Phi^{(2)}$.

5.5 General proof idea and outline

We first note that reconstruction is impossible when $\frac{a+(q-1)b}{q}\Phi^{(2)} \leq 1$, because in this regime there is no giant component¹. Note further that $\frac{a+(q-1)b}{q}\Phi^{(2)} \leq 1$ already implies $(a-b)^2\Phi^{(2)} \leq q(a+b)$.

To establish (5.1) when $\frac{a+(q-1)b}{q}\Phi^{(2)} > 1$ and $(a-b)^2\Phi^{(2)} \leq q(a+b)$, we note that $\text{Var}(\mathbb{E}[\sigma_u | \sigma_{\partial G_R}, \sigma_v, G])$ is asymptotically an upper bound for $\text{Var}(\mathbb{E}[\sigma_u | \sigma_v, G])$, as

¹Indeed, the main result in [15] concerns the existence, size and uniqueness of the giant component. In particular, in the setting considered here, a giant component emerges if and only if $\frac{a+(q-1)b}{q}\Phi^{(2)} > 1$. We shall henceforth assume a giant component to emerge.

conditioning on the boundary spins $\sigma_{\partial G_R}$ of an R -neighbourhood around u is more informative. Now, we can approximate $\text{Var}(\mathbb{E}[\sigma_u | \sigma_v, \sigma_{\partial G_R}, G]) \simeq \text{Var}(\mathbb{E}[\sigma_u | \sigma_{\partial G_R}, G])$, because long-range correlations in this model are weak (Lemma 5.8.1). Further, local neighbourhoods are w.h.p. tree-like, so that calculating the latter variance is equivalent to a certain tree-reconstruction problem discussed in Section 5.6. More specifically, we shall prove (Theorem 5.6.6) that reconstruction of the spin of the root in a q -type tree (with offspring following a Poisson-mixture) based on the spins at depth R (where $R \rightarrow \infty$), is impossible when $(a - b)^2 \Phi^{(2)} \leq q(a + b)$. Hence, $\text{Var}(\mathbb{E}[\sigma_u | \sigma_{\partial G_R}, G]) \rightarrow 0$ as $R \rightarrow \infty$.

Section 5.6 deals with branching processes where the offspring is governed by a Poisson-mixture. The main theorem (i.e., Theorem 5.6.6) deals with a reconstruction problem on these branching processes.

In Section 5.7 we establish a coupling between the local neighbourhood and T^{Poi} . This result does not follow directly from the coupling in [15], because we need the weights in the graph and their counterparts in the branching process to be *exactly* the same.

Finally, in Section 5.8 we show that long-range interactions are weak. The proof of Lemma 5.8.1 is based on an idea in the proof of Lemma 4.7 in [92]. Note however that (besides the presence of weights) the statement of our Lemma 5.8.1 is slightly stronger than Lemma 4.7 in [92], see below for details.

5.6 Broadcasting on the branching process

Here we repeat without changes the definition of a Markov broadcasting process on trees given in [41, 92]. Let \mathcal{T} be an infinite tree with root ρ . Given a number $0 \leq \epsilon < 1/(q-1)$, define a random labelling $\tau \in \{1, \dots, q\}^{\mathcal{T}}$ as follows: First, draw τ_ρ uniformly in $\{1, \dots, q\}$. Then, conditionally independently given τ_ρ , take every child u of ρ and, then with probability $1 - (q-1)\epsilon$ set $\tau_u = \tau_\rho$, and with probability $(q-1)\epsilon$ choose τ_u uniformly from $\{1, \dots, q\} \setminus \tau_\rho$. Continue this construction recursively to obtain a labelling τ for which every vertex, independently, has probability $1 - (q-1)\epsilon$ of having the same label as its parent and probability ϵ for each of the remaining spins.

Suppose that the labels $\tau_{\partial \mathcal{T}_m}$ at depth m in the tree are known (here, $\tau_U = \{\tau_i : i \in U\}$ and $\partial \mathcal{T}_m$ are all vertices at distance m from the root). The paper [41] gives precise conditions in the case of two spins as to when reconstruction of the root label is feasible using the optimal reconstruction strategy (maximum likelihood), i.e., deciding according to the sign of $\mathbb{E}[\tau_\rho | \tau_{\partial \mathcal{T}_m}]$. Interestingly, this is completely decided by the branching number of \mathcal{T} and the flip-probability ϵ . The paper [95] extends the results in [41] to the case of a general number of spins. For completeness we state both theorems here.

Definition 5.6.1. *The branching number of a tree \mathcal{T} , denoted by $\text{Br}(\mathcal{T})$, is defined as follows:*

- *If \mathcal{T} is finite, then $\text{Br}(\mathcal{T}) = 0$;*
- *If \mathcal{T} is infinite, then we define the branching number in terms of percolation. Suppose that we retain each edge in the tree independently with probability p . Then $\text{Br}(\mathcal{T})$ is the unique number such that: If $p < \frac{1}{\text{Br}(\mathcal{T})}$, then all components of the graph are finite a.s., while if $p > \frac{1}{\text{Br}(\mathcal{T})}$, then the graph has infinite components a.s.*

Remark that [41] does not deal with the trivial case of *finite* trees. On such trees, $\text{Br}(\mathcal{T}) = 0$ by convention. This makes sense because, for large m , $\partial \mathcal{T}_m = \emptyset$, and consequently $\mathbb{P}(\tau_\rho = + | \tau_{\partial \mathcal{T}_m}) = 1/q$.

Theorem 1.1 in [41] and Proposition 1.3 in [95] read, tailored to our needs:

Theorem 5.6.2. (Theorem 1.1 in [41]) For $q = 2$, consider the problem of reconstructing τ_ρ from the spins $\tau_{\partial\mathcal{T}_m}$ at the m th level of \mathcal{T} . Define Δ_m as the difference between the probability of correct and incorrect reconstruction given the information at level m :

$$\Delta_m := |\mathbb{P}(\tau_\rho = +|\tau_{\partial\mathcal{T}_m}) - \mathbb{P}(\tau_\rho = -|\tau_{\partial\mathcal{T}_m})|.$$

If $\text{Br}(\mathcal{T})(1 - 2\epsilon)^2 > 1$ then $\lim_{m \rightarrow \infty} \mathbb{E}[\Delta_m] > 0$.

If, however, $\text{Br}(\mathcal{T})(1 - 2\epsilon)^2 < 1$ then $\lim_{m \rightarrow \infty} \mathbb{E}[\Delta_m] = 0$.

Theorem 5.6.3 (Proposition 4.2 in [95]). For general $q \geq 2$, consider the problem of reconstructing τ_ρ from the spins $\tau_{\partial\mathcal{T}_m}$ at the m -th level of \mathcal{T} . Define \mathbf{P}_m^s as the conditional distribution of $\tau_{\partial\mathcal{T}_m}$ given that $\sigma_\rho = s$. Then, $\lim_{m \rightarrow \infty} \|\mathbf{P}_m^i - \mathbf{P}_m^j\|_{\text{TV}} = 0$ if $\text{Br}(\mathcal{T}) \frac{(1-q\epsilon)^2}{1-(q-2)\epsilon} < 1$.

Remark 5.6.4. Note that if $\text{Br}(\mathcal{T}) \frac{(1-q\epsilon)^2}{1-(q-2)\epsilon} < 1$, then

$$\begin{aligned} & \mathbb{E}[|\mathbb{P}(\tau_\rho = i|\tau_{\partial\mathcal{T}_m}) - \mathbb{P}(\tau_\rho = j|\tau_{\partial\mathcal{T}_m})|] \\ &= \sum_A \mathbb{P}(\tau_{\partial\mathcal{T}_m} = A) |\mathbb{P}(\tau_\rho = i|\tau_{\partial\mathcal{T}_m} = A) - \mathbb{P}(\tau_\rho = j|\tau_{\partial\mathcal{T}_m} = A)| \\ &= \frac{1}{q} \sum_A |\mathbf{P}_m^i(A) - \mathbf{P}_m^j(A)| \rightarrow 0, \end{aligned} \tag{5.4}$$

as $m \rightarrow \infty$. Thus Theorem 5.6.3 implies Theorem 5.6.2.

Note that in these theorems the tree is fixed, compared to the setting in this paper where the multi-type branching process of Section 5.4 is considered. But, it can be easily seen that the spins on a fixed instance \mathcal{T} of T^{Poi} are distributed according to the above broadcasting process.

We thus need to calculate the branching number of a typical instance \mathcal{T} :

Proposition 5.6.5. Consider the multi-type branching process T^{Poi} , where the root has spin drawn uniformly from $\{1, \dots, q\}$ and weight governed by ν . Then, given the event that the branching process does not go extinct, $\text{Br}(T^{\text{Poi}}) \leq \frac{a+(q-1)b}{q} \Phi^{(2)}$ almost surely.

Note that it can in fact be easily proved that $\text{Br}(T^{\text{Poi}}) = \frac{a+(q-1)b}{q} \Phi^{(2)}$ almost surely, given that the process survives.

We conclude with the main theorem of this section.

Theorem 5.6.6. Consider the multi-type branching process T^{Poi} , where the root has spin drawn uniformly from $\{1, \dots, q\}$ and weight governed by ν . Denote the branching process by T and its spins by τ^n . Further, let R be an unbounded non-decreasing function. Assume that $(a-b)^2 \Phi^{(2)} < q(a+b)$, then, for any $s \in \{1, \dots, q\}$,

$$\mathbb{P}(\tau_\rho = s | T_{R(n)}, \tau_{\partial T_{R(n)}}) \xrightarrow{\mathbb{P}} \frac{1}{q},$$

as $n \rightarrow \infty$.

Proof. Since $\epsilon = \frac{b}{a+(q-1)b}$, Proposition 5.6.5 gives that $\text{Br}(\mathcal{T}) \frac{(1-q\epsilon)^2}{1-(q-2)\epsilon} < 1$ almost surely. Theorem 5.6.3 (and Remark 5.6.4) then completes the proof. \square

Remark 5.6.7. In (5.15) we use a coupling between the Poisson tree and the local neighbourhood around a fixed vertex u , while we condition on the spins of all vertices exactly distance $R(n)$ away from u . If there are no such vertices, i.e., when the neighbourhood 'dies out', then this does not entail extra information. Hence the convention that $\text{Br}(\mathcal{T}) = 0$ for a finite tree \mathcal{T} .

5.7 Coupling of local neighbourhood

This section has as its objective to establish a coupling between the local neighbourhood of an arbitrary fixed vertex in the DC-SBM and T^{Poi} . The main result is the following theorem, where we let T , τ , and ψ be random instances of T^{Poi} , its spins and its weights, respectively.

Theorem 5.7.1. *Let ρ be a uniformly picked vertex in $V(G)$, where for each n , $G = G(n)$ is an instance of the DC-SBM. There exists an unbounded non-decreasing function $R : \mathbb{N} \rightarrow \mathbb{N}$ such that*

$$\| (G_{R(n)}(\rho), \sigma_{G_{R(n)}}, \phi_{G_{R(n)}}) - (T_{R(n)}, \tau_{T_{R(n)}}, \psi_{T_{R(n)}}) \|_{\text{TV}} = o_n(1),$$

and,

$$\mathbb{P} \left(|G_{R(n)}| \leq n^{1/9} \right) = 1 - o_n(1).$$

Further, in case ν has support in $[\phi_{\min}, \phi_{\max}]$, with $\phi_{\max} < \infty$, we have explicitly for $R(n) = C \log(n)$, with $C < \frac{1 - \log(4/e)}{3 \log(2 \cdot \phi_{\max}^2 \cdot (a \vee b))}$,

$$\| (G_R(\rho), \sigma_{G_R}, \phi_{G_R}) - (T_R, \tau_{T_R}, \psi_{T_R}) \|_{\text{TV}} \leq n^{-\frac{1}{2} \log(4/e)}.$$

We defer its proof to the end of this section. It uses an alternative description of the branching process in Section 5.6.

5.7.1 Alternative description of branching process

For notational convenience, we restrict ourselves here to the case of two communities only. The proof for a general number of communities follows then analogously.

We obtain an alternative description of the graph by considering a particle u with spin σ_u and weight ϕ_u to be of type $x_u = \phi_u \sigma_u \in S = -W \cup W$. We denote the law of x_u by μ , i.e., for $A \subset S$, $\mu(A) = \int_A \frac{1}{2} d\nu(|x|)$. Two distinct vertices u and v are then joined by an edge with probability $\frac{\kappa(x_u, x_v)}{n}$, where $\kappa : S \times S \rightarrow \mathbb{R}$ is defined for $(x, y) \in S \times S$ by

$$\kappa(x, y) = |xy| (1_{\{xy > 0\}} a + 1_{\{xy < 0\}} b). \quad (5.5)$$

Analogously, we obtain the following equivalent description of the branching process: We begin with a single particle o of type x_o governed by μ , giving birth to $\text{Poi}(\lambda_{x_o}(S))$ children, where for $x \in S$, and $A \subset S$,

$$\lambda_x(A) = \int_A \kappa(x, y) d\mu(y). \quad (5.6)$$

conditioned on x_o the children have i.i.d. types governed by $\mu_{x_o}^*$ ², where for $x \in S$, and $A \subset S$,

$$\mu_x^*(A) = \frac{\lambda_x(A)}{\lambda_x(S)} = \int_A \left(\frac{a}{a+b} 1_{xy > 0} + \frac{b}{a+b} 1_{xy < 0} \right) |y| \frac{d\nu(|y|)}{\Phi(1)}. \quad (5.7)$$

For generation $t \geq 1$, all particles give birth independently in the following way: A particle with type x^* is replaced in the next generation by $\text{Poi}(\lambda_{x^*}(S))$ children, again with i.i.d. types governed by $\mu_{x^*}^*$.

²Note that if y has law μ_x^* , then for any $A \subset W$, $\mathbb{P}(\text{sign}(y) = \text{sign}(x), |y| \in A) = \frac{a}{a+b} \int_A z \frac{d\nu(z)}{\Phi(1)} = \mathbb{P}(\text{sign}(y) = \text{sign}(x)) \mathbb{P}(|y| \in A)$. Hence, we can identify $\text{sign}(y)$ with the particle's spin and $|y|$ with its independent weight.

In case of a general number of communities, we let μ be the product measure of the uniform measure on $\{1, \dots, q\}$ with the measure ν . I.e., for $s \in \{1, \dots, q\}$ and $A \subset [\phi_{\min}, \infty)$, we have $\mu(\{s\} \times A) = \frac{1}{q} \cdot \nu(A)$.

In [15] it is shown that local neighbourhoods of the graph are described by the above branching process, if we *ignore* the types. (To be precise: the equivalent description used in [15] is that a particle of type x gives birth to $\text{Poi}(\lambda_x(A))$ children with type in A , for any $A \subset S$. Those numbers are independent for different sets A and different particles.)

The coupling-technique in [15] uses a discretization of κ as an intermediate step, thereby losing some information: types in the tree deviate slightly from their counterparts in the graph. We shall therefore use another coupling method, presented below, so that **the types in graph and branching process coincide exactly**.

5.7.2 Coupling

We use the following exploration process: At time $m = 0$, choose a vertex ρ uniformly in $V(G)$, where G is an instance of the DC-SBM. Initially, it is the only active vertex: $\mathcal{A}(0) = \{\rho\}$. All other vertices are neutral at start: $\mathcal{U}(0) = V(G) \setminus \{\rho\}$. No vertex has been explored yet: $\mathcal{E}(0) = \emptyset$. At each time $m \geq 0$ we arbitrarily pick an active vertex u in $\mathcal{A}(m)$ that has shortest distance to ρ , and explore all its neighbours in $\mathcal{U}(m)$, the set of unexplored vertices. If $uv \in E(G)$ for $v \in \mathcal{U}(m)$, then we set v active in step $m+1$, otherwise it remains neutral. At the end of step m , we designate u to be explored. Thus,

$$\begin{aligned}\mathcal{E}(m+1) &= \mathcal{E}(m) \cup \{u\}, \\ \mathcal{A}(m+1) &= (\mathcal{A}(m) \setminus \{u\}) \cup (\mathcal{N}(u) \cap \mathcal{U}(m)),\end{aligned}$$

and,

$$\mathcal{U}(m+1) = \mathcal{U}(m) \setminus \mathcal{N}(u).$$

Our aim in this section is to show that the exploration process and the branching process are equal upto depth $R(n)$ (defined in Theorem 5.7.1) with probability tending to one for large n . We do this in two steps:

Firstly, we establish that the types of the vertices in $\mathcal{U}(m)$ are i.i.d. with law $\mu^{(m)}$ (defined in (5.8) below) such that

$$\left\| \mu^{(m)} - \mu \right\|_{TV} = \mathcal{O}\left(n^{-\beta/8} + mn^{-3/4}\right).$$

This is the content of the following:

Lemma 5.7.2. *The following holds conditioned that all the weights are smaller than n^α , with $\alpha = 1/8$: Let $1, \dots, m$ be the vertices in $\mathcal{E}(m)$, with types $X_1 = x_1, \dots, X_m = x_m$. Then, the vertices in $\mathcal{U}(m)$ have i.i.d. types with law $\mu^{(m)} = \mu_{x_1, \dots, x_m}^{(m)}$, where*

$$d\mu^{(m)}(\cdot) = \frac{g(\cdot)d\mu_\alpha(\cdot)}{\int_S g(z)d\mu_\alpha(z)}, \quad (5.8)$$

with μ_α denoting the measure of the types conditioned that all weights are bounded by n^α , and where,

$$g(\cdot) = \prod_{i=1}^m \left(1 - \frac{\kappa(x_i, \cdot)}{n}\right). \quad (5.9)$$

Further, for all (x_1, \dots, x_m) :

$$\left\| \mu_{x_1, \dots, x_m}^{(m)} - \mu \right\|_{TV} = \mathcal{O}\left(n^{-\alpha\beta} + mn^{2\alpha-1}\right).$$

Secondly, if u has type $X = x \in S$, then its D neighbours in $\mathcal{U}(m)$ (i.e., those vertices that will be added to $\mathcal{A}(m+1)$) have i.i.d. types with a law $\mu_x^{*(m+1)}$ (defined in (5.10) below), which is $\mathcal{O}(n^{-3/8})$ away from μ_x^* in total variation distance. Further, the total variation distance between the number of neighbours D and $\text{Poi}(\lambda_x(S))$ is $\mathcal{O}(n^{-1/4})$:

Lemma 5.7.3. *The following holds conditioned that all the weights are smaller than n^α , with $\alpha = 1/8$: Assume u has type $X = x$. Let D be the number of neighbours u has in $\mathcal{U}(m)$. Then, the types of those neighbours are i.i.d. with law $\mu_x^{*(m)}$, where*

$$d\mu_x^{*(m)}(\cdot) = \frac{\kappa(x, \cdot) d\mu^{(m)}(\cdot)}{\int_S \kappa(x, y) d\mu^{(m)}(y)}. \quad (5.10)$$

For large n and $m = o(n^{1/4})$,

$$\left\| \mu_x^{*(m)} - \mu_x^* \right\|_{TV} = \mathcal{O}\left(n^{\alpha(1-\beta)} + mn^{3\alpha-1} + n^{-\alpha\beta/2}\right) = \mathcal{O}\left(n^{-3/8}\right). \quad (5.11)$$

Further,

$$\|D - \text{Poi}(\lambda_x(S))\|_{TV} = \mathcal{O}\left(n^{(1-\beta/2)\cdot 1/8} + n^{-1/4}\right) = \mathcal{O}\left(n^{-1/4}\right). \quad (5.12)$$

To establish the desired coupling, we need to show that certain events happen with high probability. To define those events, we need some notation: For $u \in \partial G_r$ (we identify $\partial G_r = \{1, \dots, |\partial G_r|\}$), put

$$D_u = |\mathcal{N}(u) \cap \mathcal{U}(|G_{r-1}| + u - 1)|.$$

Conditioned that u has type $X_u = x_u$, let

$$\widehat{D}_u = \text{Poi}(\lambda_{x_u}(S)).$$

Further, for $v \in \{1, \dots, D_u\}$, let U_{uv} denote the type of child v of vertex u and let \widehat{U}_{uv} be a random variable with law $\mu_{x_u}^*$. We assume that $\{\widehat{U}_{uv}\}_v$ are independent conditioned on $X_u = x_u$.

We put the function $g : s \mapsto 2^s - 1$ and define the events

$$A_{r+1} = \{\forall u \in \partial G_r : D_u = \widehat{D}_u\},$$

$$B_{r+1} = \{\forall u \in \partial G_r, v \in \{1, \dots, D_u\} : U_{uv} = \widehat{U}_{uv}\},$$

$$C_r = \{|\partial G_s| \leq \log^{g(s)}(n) \ \forall s \leq r\},$$

and their intersection

$$E_r = \bigcap_{s=1}^r \{A_s \cap B_s \cap C_s\}.$$

Further, we let K_r be the event that no vertex outside G_r has more than one neighbour in G_r and that there are no edges in ∂G_r (this implies that the neighbourhood is indeed a tree).

The events E_r and K_r happen with high probability:

Lemma 5.7.4. *The following holds conditioned that all the weights are smaller than n^α , with $\alpha = 1/8$: Fix $R \geq 0$. Then, for $r \leq R$,*

$$\mathbb{P}(E_{r+1}|E_r) = 1 - o_n(1).$$

Lemma 5.7.5. *The following holds conditioned that all the weights are smaller than n^α , with $\alpha = 1/8$: Fix $R \geq 0$. Then, for $r \leq R$,*

$$\mathbb{P}(K_r | C_R) = 1 - o_n(1).$$

The proofs of Lemma's 5.7.2 - 5.7.5 can be found in Section 5.9.2. The main theorem then follows:

Proof of Theorem 5.7.1. We start with the general case, where we can assume that all weights are bounded by n^α . Indeed, by a union bound over all vertices, this happens with probability $1 - \mathcal{O}(n^{1-\alpha\beta}) = 1 - o_n(1)$. For a fixed integer $R > 0$, we have

$$\mathbb{P}\left(\bigcap_{s=1}^R K_s, E_R\right) = 1 - o_n(1).$$

We construct a sequence $\{N_k\}_{k=0}^\infty$ inductively as follows: Put $N_0 = 0$ and for each k , $N_k > N_{k-1}$ as the smallest number such that

$$\mathbb{P}\left(\bigcap_{s=1}^k K_s, E_k\right) \geq 1 - \frac{1}{k}, \text{ and } \log^{2^k-1}(n)k \leq n^{1/9},$$

for all $n \geq N_k$. Put for $N_k \leq n < N_{k+1}$, $R(n) = k$. Then, for $n \geq N_k$,

$$\mathbb{P}\left(\bigcap_{s=1}^{R(n)} K_s, E_{R(n)}, |G_{R(n)}| \leq n^{1/9}\right) \geq 1 - \frac{1}{k}.$$

In case all weights are bounded by a constant $\phi_{\max} < \infty$, we can replace the errors in Lemma's 5.7.2 and 5.7.3 by $c_1 \frac{m}{n}$, respectively $c_2 \frac{m}{n}$, where c_1 and c_2 are constants depending on ϕ_{\max} . Further, the growth-condition can now be replaced by $g(s) \leq c_3^s \log(n)$, for some constant $c_3 > 1$ depending on ϕ_{\max} . This implies in particular that the neighbourhood size is smaller than n^β for a small constant $\beta > 0$. We have made this calculations explicit in the version of September 2016 of [50] on Arxiv. \square

5.8 No long-range correlation in DC-SBM

In this section we establish the main Theorem 5.2.2, from which Theorem 5.2.3 then follows. To this end, we first condition on both the spins of $\partial G_{R(n)}$ and all weights in G . Lemma 5.8.1 below shows that we then can remove the conditioning on σ_u and the graph structure outside the R -neighbourhood (including the weights):

$$\mathbb{P}(\sigma_u = + | \sigma_{\partial G_R}, \sigma_v, G, \phi) = \mathbb{P}(\sigma_u = + | \sigma_{\partial G_R}, G_R, \phi_{G_R}) + o_n(1). \quad (5.13)$$

We established in the previous section that a neighbourhood in G looks like a T^{Poi} tree with a Markov broadcasting process on it. Hence, the right-hand side of (5.13) converges to $1/q$ in probability, establishing (5.1). We show in Lemma 5.8.2 below that this contradicts the existence of a reconstruction that is positively correlated with the true type-assignment.

We begin by preparing an auxiliary lemma to prove (5.1), it establishes that long-range interactions are sufficiently weak. Its proof is inspired by Lemma 4.7 in [92]. However (besides the additional complication of weights) the result stated here is stronger in the sense that the $o_n(1)$ terms converge uniformly to 0 and that "conditioning on G " may now be replaced with "conditioning on $G_{A \cup B}$ ".

Lemma 5.8.1. *The following holds conditioned that all the weights are smaller than n^α , with $\alpha = 1/8$: Let G be an instance of the DC-SBM. Let $s \in \{1, \dots, q\}$. Let u be an uniformly picked vertex in $V(G)$. Let $A = A(G)$, $B = B(G)$, $C = C(G) \subset V$ be a (random) partition of $V(G)$, with $u \in A$, such that B separates A and C in G .*

Assume that $|A \cup B| \leq n^{1/9}$ for asymptotically almost every realization of G . Then there exists a sequence of events $(\Omega_n)_n$ and a sequence of non-negative real numbers $(\epsilon_n)_n$, such that $\mathbb{P}(\Omega_n) = 1 - o_n(1)$, and $\epsilon(n) = o_n(1)$, and further, for each n ,

$$|\mathbb{P}(\sigma_u = s | \sigma_{B \cup C}, G, \phi) - \mathbb{P}(\sigma_u = s | \sigma_B, G_{A \cup B}, \phi_{A \cup B})| \leq \epsilon(n), \quad (5.14)$$

on Ω_n .

The proof of Lemma 5.8.1 can be found in Section 5.9.3. The main Theorem 5.2.2 then follows:

Proof of Theorem 5.2.2. Put $A = G_{R-1}$, $B = \partial G_R$ and $C = G \setminus G_R$. We use the monotonicity property of conditional variance³ to obtain that, for any $s \in \{1, \dots, q\}$,

$$0 \leq \text{Var}(\mathbb{E}[1_{\{\sigma_u=s\}} | \sigma_v, G]) \leq \text{Var}(\mathbb{E}[1_{\{\sigma_u=s\}} | \sigma_{B \cup C}, G, \phi]) + o_n(1)$$

since $v \in B \cup C$ w.h.p. It suffices to show that the right-hand side tends to 0, because this implies that $\mathbb{P}(\sigma_u = s | \sigma_v, G) \xrightarrow{\mathbb{P}} 1/q$.

To show that the right-hand side tends indeed to 0, it suffices that

$$\mathbb{P}(\sigma_u = s | \sigma_{B \cup C}, G, \phi) \xrightarrow{\mathbb{P}} 1/q.$$

Now, by using the partition $A \cup B \cup C$ of $V(G)$ in Lemma 5.8.1, we have, since $G_R \leq n^{1/9}$ w.h.p., and all weights are bounded by n^α w.h.p. (this follows from a union bounded over all vertices),

$$\mathbb{P}(\sigma_u = s | \sigma_{B \cup C}, G, \phi) \stackrel{w.h.p.}{=} \mathbb{P}(\sigma_u = s | \sigma_{\partial G_R}, G_R, \phi_{G_R}) + o_n(1).$$

Theorem 5.7.1 entails that the local neighbourhood is w.h.p. equal to T^{Poi} . Let T^n be an independent copy of T^{Poi} with root ρ , spins τ^n and weights ψ^n . Note that we stress the dependence on n , because the Poisson-tree is sampled again for each n .

$$\begin{aligned} \mathbb{P}(\sigma_u = s | \sigma_{\partial G_R}, G_R, \phi_{G_R}) + o_n(1) &\stackrel{w.h.p.}{=} \mathbb{P}(\tau_\rho^n = s | \tau_{\partial T_R^n}^n, T_R^n, \psi_{T_R^n}^n) + o_n(1) \\ &= \mathbb{P}(\tau_\rho^n = s | \tau_{\partial T_R^n}^n, T_R^n) + o_n(1), \end{aligned} \quad (5.15)$$

due to the coupling from Theorem 5.7.1. By Theorem 5.6.6, the right-hand side of (5.15) tends to $1/q$ in probability. \square

Using the following auxiliary lemma (whose proof can be found in Section 5.9.3), Theorem 5.2.3 follows from Theorem 5.2.2:

Lemma 5.8.2. *Assume that $(a - b)^2 \Phi^{(2)} \leq q(a + b)$. Let G be an observation of the DC-SBM, with true communities $\{\sigma_i\}_{i=1}^n$. Let u and v be two uniformly picked vertices. Let $\{\hat{\sigma}_i\}_{i=1}^n$ be a reconstruction of the communities, based on the observation G . Assume that there exists $\delta > 0$ such that*

$$f(n) := \frac{1}{n} \sum_{i=1}^n 1_{\{\sigma_i = \hat{\sigma}_i\}} \geq \frac{1}{q} + \delta,$$

with high probability. Then, there exists $s \in \{1, \dots, q\}$, such that $\mathbb{P}(\sigma_u = s | \sigma_v, G)$ does not converge in probability to $1/q$.

We summarize these results in Theorem 5.2.3:

Proof of Theorem 5.2.3. Combine Theorem 5.2.2 and Lemma 5.8.2. \square

³For random variables X, Y, Z , we have $\text{Var}(\mathbb{E}[X|Y]) \leq \text{Var}(\mathbb{E}[X|Y, Z])$. Indeed, put $z = \mathbb{E}[X|Y, Z]$, then by Jensen's inequality $\mathbb{E}[z|Y]^2 \leq \mathbb{E}[z^2|Y]$. So that, after taking expectations on both sides, $\mathbb{E}[\mathbb{E}[X|Y]^2] \leq \mathbb{E}[\mathbb{E}[X|Y, Z]^2]$. Writing out the definition of the variance then establishes the claim.

5.9 Proofs

5.9.1 Broadcasting on the branching process

Proof of Proposition 5.6.5. Denote the multi-type branching process by T . Assume w.l.o.g. that the root has $D \geq 1$ children denoted as $1, \dots, D$. Denote by T_u^* the subtree of all particles with common ancestor u . We observe that if $\text{Br}(T_u^*) < c$ for all u , then $\text{Br}(T) < c$.

Now, conditioned on the spin of the root, $(T_u^*)_{u=1}^D$ are i.i.d. copies of T^{Poi} with weight governed by the biased law ν^* . The latter is a Galton-Watson process with offspring mean $\frac{a+(q-1)b}{q}\Phi^{(2)}$. If it dies out, then $\text{Br}(T_u^*) = 0$ by definition. Hence, given that the process survives (and thus necessarily $\frac{a+(q-1)b}{q}\Phi^{(2)} > 1$), Proposition 6.4 in [82] entails that $\text{Br}(T_u^*) = \frac{a+(q-1)b}{q}\Phi^{(2)}$ a.s. \square

5.9.2 Coupling of local neighbourhood

Proof of Lemma 5.7.2. Recall that we assume that all weights are bounded by n^α . Consider vertex $v \in \mathcal{U}(m)$ with type Y . We show first that, conditioned on $v \notin \mathcal{N}(1, \dots, m)$ and $X_1 = x_1, \dots, X_m = x_m$, Y has law $\mu_{x_1, \dots, x_m}^{(m)}$. From Bayes theorem we have, for $y \in S$,

$$\begin{aligned} & \mathbb{P}(Y \leq y | v \notin \mathcal{N}(1, \dots, m), X_1 = x_1, \dots, X_m = x_m) \\ &= \frac{\mathbb{P}(Y \leq y) \mathbb{P}(v \notin \mathcal{N}(1, \dots, m) | Y \leq y, X_1 = x_1, \dots, X_m = x_m)}{\mathbb{P}(v \notin \mathcal{N}(1, \dots, m) | X_1 = x_1, \dots, X_m = x_m)}, \end{aligned} \quad (5.16)$$

since $\mathbb{P}(Y \leq y | X_1 = x_1, \dots, X_m = x_m) = \mathbb{P}(Y \leq y)$. Recall (5.9) and observe that

$$g(\cdot) = \mathbb{P}(v \notin \mathcal{N}(1, \dots, m) | Y = \cdot, X_1 = x_1, \dots, X_m = x_m).$$

Hence, the denominator in (5.16) is just $\int_S g(z) d\mu(z)$ and evaluating the numerator yields $\int_{-\infty}^y g(z) d\mu(z)$. We thus obtain (5.8).

Since for $|y| \leq \mathcal{O}(n^\alpha)$, $d\mu_\alpha(y) = \frac{d\mu(y)}{\mathbb{P}(\phi \leq n^\alpha)}$, it follows that $\|\mu_\alpha - \mu\|_{\text{TV}} = \mathcal{O}(n^{-\alpha\beta})$.

To bound $\|\mu_\alpha - \mu^{(m)}\|_{\text{TV}}$, note that (in view of (5.5)) $g(y) = 1 - \mathcal{O}(mn^{2\alpha-1})$, for $|y| \leq \mathcal{O}(n^\alpha)$. Thus, $I := \int_S g(z) d\mu_\alpha(z) = 1 - \mathcal{O}(mn^{2\alpha-1})$. Therefore,

$$\left\| \mu^{(m)} - \mu_\alpha \right\|_{\text{TV}} \leq \int_S \left| \frac{g(y)}{I} - 1 \right| d\mu_\alpha(y) = \mathcal{O}(mn^{2\alpha-1}).$$

We finish by invoking the triangle inequality. \square

Proof of Lemma 5.7.3. Put $n_m = |\mathcal{U}(m)|$ and let Y_1, \dots, Y_D denote the types of the neighbours of u .

Let f_1, \dots, f_n be arbitrary measurable functions. The first claim follows if we prove that

$$\mathbb{E} \left[e^{-\sum_{j=1}^D f_j(Y_j)} \mid D = d \right] = \prod_{j=1}^d \left(\int_S e^{-f_j(y)} d\mu_x^{*(m)}(y) \right). \quad (5.17)$$

Now, abbreviating conditioning on $\mathcal{N}(u) \cap \mathcal{U}(m) = F$ by F , we have,

$$\begin{aligned} & \mathbb{E} \left[e^{-\sum_{j=1}^D f_j(Y_j)} 1_{D=d} \right] \\ &= \sum_{F \subset [n_m], |F|=d} \mathbb{E} \left[e^{-\sum_{j \in F} f_j(Y_j)} \middle| F \right] \cdot \left(1 - \frac{1}{n} \int_S \kappa(x, y) d\mu^{(m)}(y) \right)^{n_m-d} \\ & \quad \cdot \left(\frac{1}{n} \int_S \kappa(x, y) d\mu^{(m)}(y) \right)^d. \end{aligned}$$

We have,

$$\begin{aligned} \mathbb{P}(D = d) &= \binom{n_m}{d} \left(1 - \frac{1}{n} \int_S \kappa(x, y) d\mu^{(m)}(y) \right)^{n_m-d} \\ & \quad \cdot \left(\frac{1}{n} \int_S \kappa(x, y) d\mu^{(m)}(y) \right)^d. \end{aligned}$$

Hence,

$$\mathbb{E} \left[e^{-\sum_{j=1}^D f_j(Y_j)} \middle| D = d \right] = \frac{1}{\binom{n_m}{d}} \sum_{F \subset [n_m], |F|=d} \mathbb{E} \left[e^{-\sum_{j \in F} f_j(Y_j)} \middle| F \right].$$

Conditioned on $F \subset [n_m]$, the types $(Y_j)_{j \in F}$ are i.i.d., thus

$$\mathbb{E} \left[e^{-\sum_{j \in F} f_j(Y_j)} \middle| F \right] = \prod_{j=1}^d \left(\frac{\int_S e^{-f_j(y)} \frac{\kappa(x, y)}{n} d\mu^{(m)}(y)}{\int_S \frac{\kappa(x, y)}{n} d\mu^{(m)}(y)} \right),$$

which combined with (5.10) gives (5.17), our first claim.

Further,

$$\begin{aligned} \|\mu_x^{*(m)} - \mu_x^*\|_{\text{TV}} &\leq \int_S f_x(y) \left| \frac{d\mu^{(m)}(y)}{I_x^{(m)}} - \frac{d\mu(y)}{I_x} \right| \\ &= \frac{1}{I_x} \int_S f_x(y) \left| d\mu^{(m)}(y) \left(1 + \mathcal{O} \left(I_x^{(m)} - I_x \right) \right) - d\mu(y) \right|, \end{aligned} \tag{5.18}$$

where $f_x(y) = (1_{\{xy > 0\}}a + 1_{\{xy < 0\}}b) |y|$, $I_x^{(m)} = \int_S f_x(z) d\mu^{(m)}(z)$ and $I_x = \int_S f_x(z) d\mu(z)$. Now,

$$\begin{aligned} |I_x^{(m)} - I_x| &\leq \mathcal{O}(n^\alpha) \int_{|z| \leq n^\alpha} |d\mu^{(m)}(z) - d\mu(z)| + \int_{|z| > n^\alpha} |z| d\mu(z) \\ &= \mathcal{O} \left(n^{\alpha-\alpha\beta} + mn^{3\alpha-1} + n^{-\alpha\beta/2} \right), \end{aligned} \tag{5.19}$$

where we used the proof of the previous lemma to bound the first term and Cauchy-Schwartz inequality for the second term. Now, the right-hand side in (5.18) is thus of the same order (since the weights have bounded expectation).

For the last claim, observe that $D = \text{Bin}(n_m, p)$, where $p = \frac{1}{n} \int_S \kappa(x, y) d\mu^{(m)}(y)$. Hence, since the weights have bounded first moment,

$$\|\text{Bin}(n_m, p) - \text{Poi}(n_m p)\|_{\text{TV}} \leq \sum_{i=1}^{n_m} p^2 = \mathcal{O} \left(n^{-3/4} \right).$$

Standard bounds for Poisson random variables entail the existence of a constant $C_{\text{Poi}} \geq 1$ such that $\|\text{Poi}(\mu) - \text{Poi}(\lambda)\|_{\text{TV}} \leq C_{\text{Poi}}|\mu - \lambda|$. Consequently,

$$\begin{aligned} \frac{1}{C_{\text{Poi}}} \|\text{Poi}(n_m p) - \text{Poi}(\lambda_x(S))\|_{\text{TV}} &\leq |n_m - n|p + |x| |I_x^{(m)} - I_x| \\ &\leq \frac{|n_m - n|}{n} n^\alpha \\ &\quad + \mathcal{O}\left(n^{2\alpha - \alpha\beta} + mn^{4\alpha - 1} + n^{\alpha - \alpha\beta/2}\right). \end{aligned}$$

Thus, by the triangle inequality,

$$\|\text{Bin}(n_m, p) - \text{Poi}(\lambda_x(S))\|_{\text{TV}} = \mathcal{O}\left(n^{(1-\beta/2)\cdot 1/8} + n^{-1/4}\right).$$

□

Proof of Lemma 5.7.4. Write $n_r = |\partial G_r|$. We have

$$\mathbb{P}(E_{r+1}|E_r) \geq \mathbb{P}(B_{r+1}|E_r) - \mathbb{P}(\neg A_{r+1}|E_r) - \mathbb{P}(\neg C_{r+1}|E_r).$$

Now,

$$\mathbb{P}(B_{r+1}|E_r, n_r) \geq 1 - \sum_{u=1}^{n_r} \mathbb{P}\left(\neg B_{r+1}^{(u)} \mid \bigcap_{v=1}^{u-1} B_{r+1}^{(v)}, E_r\right), \quad (5.20)$$

where $B_{r+1}^{(u)} = \{\forall w \in \{1, \dots, D_u\} : U_{uw} = \hat{U}_{uw}\}$. Denote the already explored vertices by $1, \dots, m$ (where $m = |G_{r-1}| + u - 1$) and their types as X_1, \dots, X_m . Conditioned on those types, the vertices in $\mathcal{U}(m)$ are i.i.d. with distribution $\mu^{(m)}$. Hence:

$$\begin{aligned} \mathbb{P}\left(B_{r+1}^{(u)} \mid \bigcap_{v=1}^{u-1} B_{r+1}^{(v)}, E_r, n_r, X_1, \dots, X_m\right) &= \mathbb{P}\left(B_{r+1}^{(u)} \mid X_1, \dots, X_m\right) \\ &\geq \mathbb{P}\left(B_{r+1}^{(u)} \mid D_u \leq \log(n) \log^{g(r)}(n), X_1, \dots, X_m\right) \\ &\quad \cdot \mathbb{P}\left(D_u \leq \log(n) \log^{g(r)}(n) \mid X_1, \dots, X_m\right). \end{aligned} \quad (5.21)$$

Now, $D_u \leq \sum_{i=1}^d \text{Ber}\left((a+b)\frac{\phi_i^*}{n}\right)$, where ϕ^* is governed by the size-biased law ν^* and $\{\phi_i\}_i$ are i.i.d. and bounded by n^α . Hoeffding's inequality gives that $\frac{1}{n} \sum_{i=1}^n \phi_i \leq 2\Phi^{(1)}$ w.p. at least $1 - \exp(-n^{1-2\alpha})$, and $\phi^* \leq \log^{g(r)}(n)$ w.p. at least $1 - \mathcal{O}\left(\left(\log^{g(r)}(n)\right)^{1-\beta}\right)$ (note the exponent $\beta - 1$ of the size-biased power-law).

Conditioned on those events, we use a multiplicative Chernoff bound to obtain,

$$\mathbb{P}\left(D_u \leq \log(n) \log^{g(r)}(n) \mid X_1, \dots, X_m\right) \geq 1 - \mathcal{O}\left(\left(\log^{g(r)}(n)\right)^{1-\beta}\right). \quad (5.22)$$

Lemma 5.7.3 entails, since $m = o(n^{1/4})$,

$$\mathbb{P}\left(B_{r+1}^{(u)} \mid D_u \leq \log^{g(r)+1}(n), X_1, \dots, X_m\right) \geq 1 - \mathcal{O}\left(\frac{\log^{g(r)+1}(n)}{n^{3/8}}\right). \quad (5.23)$$

Then, (5.21) - (5.23) together give

$$\mathbb{P} \left(B_{r+1}^{(u)} \left| \bigcap_{v=1}^{u-1} B_{r+1}^{(v)}, E_r, X_1, \dots, X_m \right. \right) \geq 1 - \mathcal{O} \left(\left(\log^{g(r)}(n) \right)^{1-\beta} \right).$$

Now, since conditioned on E_r , $n_r \leq \log^{g(r)}(n)$, (5.20) gives

$$\mathbb{P}(B_{r+1}|E_r) \geq 1 - \mathcal{O} \left(\left(\log^{g(r)}(n) \right)^{2-\beta} \right).$$

The growth condition (C_r) follows also from (5.22).

We take a similar approach to quantify

$$\mathbb{P}(A_{r+1}|E_r, n_r) \geq 1 - \sum_{u=1}^{n_r} \mathbb{P} \left(\neg A_{r+1}^{(u)} \left| \bigcap_{v=1}^{u-1} A_{r+1}^{(v)}, E_r, n_r \right. \right), \quad (5.24)$$

where, $A_{r+1}^{(u)} = \{D_u = \widehat{D}_u, D_u \leq \log^{g(r)+1}(n)\}$. Now,

$$\mathbb{P} \left(A_{r+1}^{(u)} \left| \bigcap_{v=1}^{u-1} A_{r+1}^{(v)}, E_r \right. \right) \geq 1 - \mathcal{O} \left(n^{(1-\beta/2)1/8} + n^{-1/4} + \log^{g(r)(1-\beta)}(n) \right), \quad (5.25)$$

due to Lemma 5.7.3, since $n - |\mathcal{U}(m)| = o(n^{1/4})$ when r is fixed. Thus, (5.24) gives

$$\mathbb{P}(A_{r+1}|E_r) \geq 1 - \mathcal{O} \left(\log^{g(r)}(n) n^{(1-\beta/2)1/8} + n^{-1/4} + \log^{g(r)(2-\beta)}(n) \right).$$

□

Proof of Lemma 5.7.5. Fix $u, v \in \partial G_r$. The probability of having an edge between u and v is smaller than $\mathcal{O}(n^{2\alpha-1})$. For any $w \in V(G \setminus G_r)$, the probability that (u, w) and (v, w) both appear is smaller than $\mathcal{O}(n^{4\alpha-2})$. Now, Lemma 5.7.4 implies that

$$|G_r| \leq \log(n)^{g(R)} R = \log^{2^R-1}(n) R.$$

Hence, the result follows from a union bound over all triples u, v, w . □

5.9.3 No long-range correlation in DC-SBM

Proof of Lemma 5.8.1. For a fixed graph g , spin-configuration τ and degree-configuration ψ , we make a factorization of $\mathbb{P}(G = g, \sigma = \tau | \phi = \psi)$ into parts depending on A, B and C . We claim that the part that measures the interaction between A and C is asymptotically independent of τ . Put

$$\Psi_{uv}(g, \tau, \psi) = \begin{cases} a \frac{\psi_u \psi_v}{n} & \text{if } (u, v) \in E(g) \text{ and } \tau_u = \tau_v \\ b \frac{\psi_u \psi_v}{n} & \text{if } (u, v) \in E(g) \text{ and } \tau_u \neq \tau_v \\ 1 - a \frac{\psi_u \psi_v}{n} & \text{if } (u, v) \notin E(g) \text{ and } \tau_u = \tau_v \\ 1 - b \frac{\psi_u \psi_v}{n} & \text{if } (u, v) \notin E(g) \text{ and } \tau_u \neq \tau_v. \end{cases}$$

We define for arbitrary sets $U_1, U_2 \subset V$,

$$\begin{aligned} Q_{U_1, U_2} &= Q_{U_1, U_2}(g, \tau, \psi) = Q_{U_1, U_2}(g_{U_1 \cup U_2}, \tau_{U_1 \cup U_2}, \psi_{U_1 \cup U_2}) \\ &= \prod_{u \in U_1, v \in U_2} \Psi_{uv}(g, \tau, \psi), \end{aligned}$$

where the subscript indicates restriction of the corresponding quantities to $U_1 \cup U_2$. Then, we have,

$$\mathbb{P}(G = g | \sigma = \tau, \phi = \psi) = Q_{A \cup B, A \cup B} Q_{B \cup C, C} Q_{A, C}. \quad (5.26)$$

We begin by demonstrating that $Q_{A, C}$ is asymptotically independent of τ : Write,

$$Q_{A, C}(g, \tau, \psi) = \prod_{u \in A, v \in C: \tau_u = \tau_v} \left(1 - a \frac{\psi_u \psi_v}{n}\right) \prod_{u \in A, v \in C: \tau_u \neq \tau_v} \left(1 - b \frac{\psi_u \psi_v}{n}\right),$$

since A and C are separated by B (there are thus no edges between A and C). The first product may be rewritten as,

$$\begin{aligned} \prod_{u \in A, v \in C: \tau_u = \tau_v} \left(1 - a \frac{\psi_u \psi_v}{n}\right) &= \exp \left(\sum_{u \in A, v \in C: \tau_u = \tau_v} \log \left(1 - a \frac{\psi_u \psi_v}{n}\right) \right) \\ &= \exp \left(\sum_{u \in A, v \in C: \tau_u = \tau_v} \left(-a \frac{\psi_u \psi_v}{n} + \mathcal{O}(n^{4\alpha-2}) \right) \right) \\ &= \exp \left(-\frac{a}{n} \sum_{u \in A, v \in C: \tau_u = \tau_v} \psi_u \psi_v + \mathcal{O}(n_A n^{4\alpha-1}) \right). \end{aligned}$$

Now, the sum $\frac{1}{n} \sum_{u \in A, v \in C: \tau_u = \tau_v} \psi_u \psi_v$ tends to $\frac{\|A\| \Phi^{(1)}}{q}$, if $(\tau, \psi) \in \Omega(n)$, where

$$\|A\| = \sum_{u \in A} \psi_u,$$

and where,

$$\Omega(n) = \left\{ (\tau', \psi') : \left| \frac{1}{n} \sum_{\tau_u = k, u \in V} \psi_u - \frac{\Phi^{(1)}}{q} \right| \leq n^{-\frac{1}{4}}, \forall k \in \{1, \dots, q\} \right\}. \quad (5.27)$$

Indeed,

$$\begin{aligned} \frac{1}{n} \sum_{u \in A, v \in C: \tau_u = \tau_v} \psi_u \psi_v &= \sum_{k=1}^q \sum_{u \in A} 1_{\{\tau_u = k\}} \psi_u \frac{1}{n} \sum_{v \in C} 1_{\{\tau_v = k\}} \psi_v \\ &= \frac{\|A\| \Phi^{(1)}}{q} + \mathcal{O}\left(n^{-\frac{1}{72}}\right), \end{aligned} \quad (5.28)$$

since $|V| - |C| \leq n^{1/9}$ and $\psi_u \leq n^{1/8}$.

As a consequence,

$$\begin{aligned} \prod_{u \in A, v \in C: \tau_u = \tau_v} \left(1 - a \frac{\psi_u \psi_v}{n}\right) &= \exp \left(\mathcal{O}\left(n^{-\frac{1}{72}}\right) \right) \cdot \exp \left(-a \frac{\|A\| \Phi^{(1)}}{q} \right) \\ &= (1 + o_n(1)) \exp \left(-a \frac{\|A\| \Phi^{(1)}}{q} \right), \end{aligned}$$

where the o_n term is uniform for all $(\tau, \psi) \in \Omega(n)$. We carry out a similar calculation for the other product. Together we obtain

$$Q_{A, C}(g, \tau, \psi) = (1 + o_n(1)) \exp \left(-\frac{a + (q-1)b}{q} \|A\| \Phi^{(1)} \right), \quad (5.29)$$

uniformly for all $(\tau, \psi) \in \Omega(n)$. This proves that $Q_{A,C}(g, \tau, \psi)$ is indeed essentially independent of τ for most pairs (τ, ψ) .

We use the above to prove that, for $u \in V$,

$$\begin{aligned} & \mathbb{P}(\sigma_u = \tau_u | \sigma_{B \cup C} = \tau_{B \cup C}, G = g, \phi = \psi, (\phi, \sigma) \in \Omega(n)) \\ &= (1 + o_n(1)) \mathbb{P}(\sigma_u = \tau_u | \sigma_B = \tau_B, G_{A \cup B} = g_{A \cup B}, \phi_{A \cup B} = \psi_{A \cup B}, (\phi, \sigma) \in \Omega(n)) \\ & \quad + o_n(1). \end{aligned} \tag{5.30}$$

Fix $(\tau, \psi) \in \Omega(n)$. Then,

$$\mathbb{P}(G = g, \sigma = \tau | \phi = \psi, (\phi, \sigma) \in \Omega(n)) = \mathbb{P}(G = g | \sigma = \tau, \phi = \psi) f(\psi, n), \tag{5.31}$$

where $f(\psi, n) = \mathbb{P}(\sigma = \tau | \phi = \psi, (\phi, \sigma) \in \Omega(n)) = \frac{q^{-n}}{\mathbb{P}((\phi, \sigma) \in \Omega(n) | \phi = \psi)}$. Hence, plugging (5.26) and (5.29) in (5.31),

$$\begin{aligned} & \mathbb{P}(G = g, \sigma = \tau | \phi = \psi, (\phi, \sigma) \in \Omega(n)) \\ &= Q_{A \cup B, A \cup B}(g, \tau, \psi) Q_{B \cup C, C}(g, \tau, \psi) \\ & \quad \cdot (1 + o_n(1)) \exp\left(-\frac{a + (q-1)b}{q} \|A\| \Phi^{(1)}\right) f(\psi, n). \end{aligned} \tag{5.32}$$

Put, for $U \subset V$,

$$\Omega_U(n) = \Omega_U(\psi, \tau_U, n) = \{\tau' : \tau'_U = \tau_U, (\tau', \psi) \in \Omega(n)\},$$

then, invoking (5.32),

$$\begin{aligned} & \mathbb{P}(G = g, \sigma_U = \tau_U | \phi = \psi, (\phi, \sigma) \in \Omega(n)) \\ &= \sum_{\tau' \in \Omega_U(n)} \mathbb{P}(G = g, \sigma = \tau' | \phi = \psi, (\phi, \sigma) \in \Omega(n)) \\ &= \sum_{\tau' \in \Omega_U(n)} Q_{A \cup B, A \cup B}(g, \tau', \psi) Q_{B \cup C, C}(g, \tau', \psi) \\ & \quad \cdot (1 + o_n(1)) \exp\left(-\frac{a + (q-1)b}{q} \|A\| \Phi^{(1)}\right) f(\psi, n) \\ &= (1 + o_n(1)) \exp\left(-\frac{a + (q-1)b}{q} \|A\| \Phi^{(1)}\right) f(\psi, n) \\ & \quad \cdot \sum_{\tau' \in \Omega_U(n)} Q_{A \cup B, A \cup B}(g, \tau', \psi) Q_{B \cup C, C}(g, \tau', \psi), \end{aligned} \tag{5.33}$$

where we could interchange the order $o_n(1)$ term and the sum because the former holds *uniformly* for all $(\phi, \sigma) \in \Omega(n)$.

We apply (5.33) with $U = A$ and $U = A \cup B$, to rewrite the right hand side of

$$\begin{aligned} & \mathbb{P}(\sigma_A = \tau_A | \sigma_B = \tau_B, G = g, \phi = \psi, (\phi, \sigma) \in \Omega(n)) \\ &= \frac{\mathbb{P}(G = g, \sigma_{A \cup B} = \tau_{A \cup B} | \phi = \psi, (\phi, \sigma) \in \Omega(n))}{\mathbb{P}(G = g, \sigma_B = \tau_B | \phi = \psi, (\phi, \sigma) \in \Omega(n))} \end{aligned} \tag{5.34}$$

as

$$\begin{aligned} & (1 + o_n(1)) \frac{\sum_{\tau' \in \Omega_{A \cup B}(n)} Q_{A \cup B, A \cup B}(g, \tau', \psi) Q_{B \cup C, C}(g, \tau', \psi)}{\sum_{\tau' \in \Omega_B(n)} Q_{A \cup B, A \cup B}(g, \tau', \psi) Q_{B \cup C, C}(g, \tau', \psi)} \\ &= (1 + o_n(1)) \frac{Q_{A \cup B, A \cup B}(g, \tau, \psi) \sum_{\tau' \in \Omega_{A \cup B}(n)} Q_{B \cup C, C}(g, \tau', \psi)}{\sum_{\tau''' \in \Omega_{B \cup C}(n)} Q_{A \cup B, A \cup B}(g, \tau''', \psi) \sum_{\tau'' \in \Omega_{A \cup B}(n)} Q_{B \cup C, C}(g, \tau'', \psi)}, \end{aligned}$$

where we used that $Q_{U_1, U_2}(\tau')$ depends on τ' only through $\tau'_{U_1 \cup U_2}$ to rewrite the numerator. Factorization of the denominator is justified as follows: For an arbitrary $\tau' \in \Omega_B(n)$, put $\tau'' = (\tau_{AUB}, \tau'_C) \in \Omega_{AUB}(n)$ and $\tau''' = (\tau'_A, \tau_{BUC}) \in \Omega_{BUC}(n)$. Then,

$$Q_{AUB, AUB}(g, \tau', \psi) Q_{BUC, C}(g, \tau', \psi) = Q_{AUB, AUB}(g, \tau''', \psi) Q_{BUC, C}(g, \tau'', \psi). \quad (5.35)$$

This proves that the double summation is at least as large as the single sum. Equality follows upon putting $\tau' = (\tau''', \tau_B, \tau''_C)$ for arbitrary $\tau'' \in \Omega_{AUB}(n)$ and $\tau''' \in \Omega_{BUC}(n)$: (5.35) is then again satisfied. Hence, (5.34) is equivalent to

$$\begin{aligned} & \mathbb{P}(\sigma_A = \tau_A | \sigma_B = \tau_B, G = g, \phi = \psi, (\phi, \sigma) \in \Omega(n)) \\ &= (1 + o_n(1)) \frac{Q_{AUB, AUB}(g, \tau, \psi)}{\sum_{\tau''' \in \Omega_{BUC}(n)} Q_{AUB, AUB}(g, \tau''', \psi)}. \end{aligned} \quad (5.36)$$

We shall rewrite the right hand side of (5.36) to obtain on the one hand:

$$\begin{aligned} & \mathbb{P}(\sigma_u = \tau_u | \sigma_B = \tau_B, G = g, \phi = \psi, (\phi, \sigma) \in \Omega(n)) \\ &= (1 + o_n(1)) \widehat{F}(g_{AUB}, \tau_{u \cup B}, \psi_{AUB}), \end{aligned} \quad (5.37)$$

for some function $\widehat{F}(\cdot) \leq 1$. And, on the other hand:

$$\begin{aligned} & \mathbb{P}(\sigma_u = \tau_u | \sigma_B = \tau_B, G = g, \phi = \psi, (\phi, \sigma) \in \Omega(n)) \\ &= (1 + o_n(1)) \mathbb{P}(\sigma_u = \tau_u | \sigma_{BUC} = \tau_{BUC}, G = g, \phi = \psi, (\phi, \sigma) \in \Omega(n)). \end{aligned} \quad (5.38)$$

To do so, note that

$$\sum_{\tau''' \in \Omega_{BUC}(n)} Q_{AUB, AUB}(g, \tau''', \psi) = \sum_{\tau'''_A \in \{1, \dots, q\}^A} Q_{AUB, AUB}(g_{AUB}, (\tau'''_A, \tau_B), \psi_{AUB}),$$

Therefore, (5.36) is equivalent to

$$\mathbb{P}(\sigma_A = \tau_A | \sigma_B = \tau_B, G = g, \phi = \psi, (\phi, \sigma) \in \Omega(n)) = (1 + o_n(1)) F(g_{AUB}, \tau_{AUB}, \psi_{AUB}),$$

for some function $F(\cdot) \leq 1$. If we fix $u \in A$ and integrate over all possible values of $\tau_{A \setminus u}$ while keeping τ_{BUC} and ψ constant, we obtain (5.37).

To establish (5.38), we multiply both denominator and numerator of (5.36) by $Q_{BUC, C}(g, \tau, \psi)$:

$$\begin{aligned} & \mathbb{P}(\sigma_A = \tau_A | \sigma_B = \tau_B, G = g, \phi = \psi, (\phi, \sigma) \in \Omega(n)) \\ &= (1 + o_n(1)) \frac{Q_{AUB, AUB}(g, \tau, \psi) Q_{BUC, C}(g, \tau, \psi)}{\sum_{\tau' \in \Omega_{BUC}(n)} Q_{AUB, AUB}(g, \tau', \psi) Q_{BUC, C}(g, \tau', \psi)} \\ &= (1 + o_n(1)) \frac{\mathbb{P}(G = g, \sigma = \tau | \phi = \psi, (\phi, \sigma) \in \Omega(n))}{\mathbb{P}(G = g, \sigma_{BUC} = \tau_{BUC} | \phi = \psi, (\phi, \sigma) \in \Omega(n))} \\ &= (1 + o_n(1)) \mathbb{P}(\sigma_A = \tau_A | \sigma_{BUC} = \tau_{BUC}, G = g, \phi = \psi, (\phi, \sigma) \in \Omega(n)). \end{aligned}$$

Integrating again over $\tau_{A \setminus u}$ gives (5.38).

We use (5.37) to obtain

$$\begin{aligned} & \mathbb{P}(\sigma_u = \tau_u | \sigma_B = \tau_B, G_{AUB} = g_{AUB}, \phi_{AUB} = \psi_{AUB}, (\phi, \sigma) \in \Omega(n)) \\ &= \sum_{\widehat{g}, \psi_C} \mathbb{P}(\sigma_u = \tau_u | \sigma_B = \tau_B, G = \widehat{g}, \phi = (\psi_{AUB}, \psi_C), (\phi, \sigma) \in \Omega(n)) \\ &\quad \cdot \mathbb{P}(G = \widehat{g}, \phi_C = \psi_C | \sigma_B = \tau_B, G_{AUB} = g_{AUB}, \phi_{AUB} = \psi_{AUB}, (\phi, \sigma) \in \Omega(n)) \\ &= (1 + o_n(1)) \widehat{F}(g_{AUB}, \tau_{u \cup B}, \psi_{AUB}) + o_n(1) \\ &= (1 + o_n(1)) \mathbb{P}(\sigma_u = \tau_u | \sigma_B = \tau_B, G = g, \phi = \psi, (\phi, \sigma) \in \Omega(n)) + o_n(1). \end{aligned} \quad (5.39)$$

Combining (5.38) and (5.39) gives

$$\begin{aligned} & \mathbb{P}(\sigma_u = \tau_u | \sigma_{B \cup C} = \tau_{B \cup C}, G = g, \phi = \psi, (\phi, \sigma) \in \Omega(n)) \\ &= (1 + o_n(1)) \mathbb{P}(\sigma_u = \tau_u | \sigma_B = \tau_B, G = g, \phi = \psi, (\phi, \sigma) \in \Omega(n)) \\ &= (1 + o_n(1)) \mathbb{P}(\sigma_u = \tau_u | \sigma_B = \tau_B, G_{A \cup B} = g_{A \cup B}, \phi_{A \cup B} = \psi_{A \cup B}, (\phi, \sigma) \in \Omega(n)), \end{aligned}$$

i.e., the claim (5.30).

Our last step consists in removing the condition $(\sigma, \phi) \in \Omega(n)$: Put $\epsilon(n) = 1 - \mathbb{P}((\sigma, \phi) \in \Omega(n))$, then $\lim_{n \rightarrow \infty} \epsilon(n) = 0$. Indeed,

$\sum_{u \in C} 1_{\{\sigma_u = k\}} \phi_u = \sum_{u \in V} 1_{\{\sigma_u = k\}} \phi_u + \mathcal{O}(n^{17/72})$, where the sum over V has $n \frac{\Phi(1)}{q}$ as a mean. The claim thus follows upon applying Hoeffding's inequality (the weights are assumed to be bounded by n^α).

Consider the random variable

$\mathbb{P}((\phi, \sigma) \in \Omega(n) | \sigma_B, G_{A \cup B}, \phi_{A \cup B}) = \mathbb{E}[1_{(\phi, \sigma) \in \Omega(n)} | \sigma_B, G_{A \cup B}, \phi_{A \cup B}]$. It has expectation $1 - \epsilon(n)$, so that

$$\mathbb{P}\left(\mathbb{E}[1_{(\phi, \sigma) \in \Omega(n)} | \sigma_B, G_{A \cup B}, \phi_{A \cup B}] \geq 1 - \sqrt{\epsilon(n)}\right) \geq 1 - 2\sqrt{\epsilon(n)}. \quad (5.40)$$

Indeed, if contrary to our claim $f := \mathbb{E}[1_{(\phi, \sigma) \in \Omega(n)} | \sigma_B, G_{A \cup B}, \phi_{A \cup B}] \geq 1 - \sqrt{\epsilon(n)}$ with probability at most $1 - 2\sqrt{\epsilon(n)}$, then

$$\mathbb{E}[f] \leq 1 \cdot (1 - 2\sqrt{\epsilon(n)}) + (1 - \sqrt{\epsilon(n)}) \cdot 2\sqrt{\epsilon(n)} < 1 - \epsilon(n).$$

Similarly, for $B \cup C$,

$$\mathbb{P}\left(\mathbb{E}[1_{(\phi, \sigma) \in \Omega(n)} | \sigma_{B \cup C}, G, \phi] \geq 1 - \sqrt{\epsilon(n)}\right) \geq 1 - 2\sqrt{\epsilon(n)}. \quad (5.41)$$

It follows that, with probability at least $1 - \mathcal{O}(\sqrt{\epsilon(n)})$,

$$\begin{aligned} & \mathbb{P}(\sigma_u = + | \sigma_B, G_{A \cup B}, \phi_{A \cup B}) \\ &= \left(1 - \mathcal{O}(\sqrt{\epsilon(n)})\right) \mathbb{P}(\sigma_u = + | \sigma_B, G_{A \cup B}, \phi_{A \cup B}, (\phi, \sigma) \in \Omega(n)) \\ &\quad + \mathcal{O}(\sqrt{\epsilon(n)}) \mathbb{P}(\sigma_u = + | \sigma_B, G_{A \cup B}, \phi_{A \cup B}, (\phi, \sigma) \notin \Omega(n)) \\ &= (1 + o_n(1)) \mathbb{P}(\sigma_u = + | \sigma_{B \cup C}, G, \phi, (\phi, \sigma) \in \Omega(n)) + o_n(1) \\ &= (1 + o_n(1)) \mathbb{P}(\sigma_u = + | \sigma_{B \cup C}, G, \phi) + o_n(1), \end{aligned}$$

where we used (5.40), (5.30) and (5.41) in the first, second, respectively last equality. \square

Proof of Lemma 5.8.2. Assume for a contradiction that for every s , $\mathbb{P}(\sigma_u = s | \sigma_v, G)$ tends to $1/q$ in probability. Since $\hat{\sigma}_u$ depends on σ_u only through G , we have for any $s \in \{1, \dots, q\}$,

$$\begin{aligned} \text{Var}\left(\mathbb{E}[1_{\{\sigma_u = s\}} | \sigma_v, G]\right) &= \text{Var}\left(\mathbb{E}[1_{\{\sigma_u = s\}} | \hat{\sigma}_u, \sigma_v, G]\right) \\ &\geq \text{Var}\left(\mathbb{E}[1_{\{\sigma_u = s\}} | \hat{\sigma}_u]\right), \end{aligned} \quad (5.42)$$

where the term on the left tends to zero by assumption. By definition of $f(n)$,

$$1/q + \delta + o(1) \leq \sum_s \mathbb{P}(\sigma_u = \hat{\sigma}_u | \hat{\sigma}_u = s) \mathbb{P}(\hat{\sigma}_u = s).$$

Hence, for large enough n , there must be an s such that $\mathbb{P}(\sigma_u = \hat{\sigma}_u | \hat{\sigma}_u = s) \geq 1/q + \delta/2$ and $\mathbb{P}(\hat{\sigma}_u = s) \geq \frac{\delta}{3q}$. As a consequence, the term on the right of (5.42) does not tend to zero. \square

Chapter 6

Non-backtracking spectrum of the *sparse* DC-SBM

We have published parts of this chapter in the conference proceedings of ITCS 2017 [51].

6.1 Introduction

In the previous chapter we established that in the *sparse* DC-SBM there is a detectability-threshold below which detecting the communities is impossible. Traditional methods based on the Adjacency and Laplacian matrix break down long before this threshold is reached. This is due to localization of the eigenvectors; the phenomena where the "mass" in eigenvectors concentrates around just a few nodes. Taking the Adjacency matrix A as an example, we can understand this as follows: Since A is symmetric, its eigenvectors can be obtained iteratively by applying powers of A on some initial vector, i.e., $x_{k+1} = Ax_k$. Let us start with the all-one vector and assume for simplicity that our network is a star where the central node has high degree d . Then $x_2 = (d, 1, \dots, 1)^T$, $x_3 = (d, d, \dots, d)^T$, $x_4 = (d^2, d, \dots, d)^T$, etcetera. This hints that high-degreed vertices accumulate an "unfair" large part of the mass in eigenvectors. In Section 4.4.1 we made this more precise by showing that the first eigenvectors of a power-law graph become indistinguishable from eigenvectors corresponding to high-degreed stars (this is a modest extension of the results in [90] for graphs without communities). Note that the degree-heterogeneity is caused here by the sparsity, as outliers of order $\log(n)/\log\log(n)$ now occur. Further, in Chapter 2 we explained that in a sufficiently dense regime the bulk of eigenvalues of A should be confined to a finite interval (semi-circle law) and the few outliers should correspond to the community-structure. The presence of relatively high-degreed nodes in the sparse regime "smears" those sharp edges so that the informative eigenvectors get lost in the bulk, this is clearly illustrated in Figure 1 of [72]. Similar observations can be made for the Laplacian matrix [54].

We thus need a method that does not suffer from resonance. Belief-propagation is conjectured to work all the way down to the detectability-threshold, see Section 1.3.5 and references there, in particular [34, 35]. In the belief-propagation algorithm a vertex informs its neighbours about its own belief concerning its spin *without* taking into account its neighbours presence. The non-backtracking matrix appears naturally when we linearise the belief-propagation algorithm, see below.

The non-backtracking matrix B of a graph $G = (V, E)$ is indexed by the set of its oriented edges $\vec{E} = \{(u, v) : \{u, v\} \in E\}$. For $e = (e_1, e_2)$, $f = (f_1, f_2) \in \vec{E}$, B is defined as

$$B_{ef} = 1_{e_2=f_1} 1_{e_1 \neq f_2}.$$

This matrix was introduced by Hashimoto [60] in 1989.

We study here the spectrum of B when G is a random graph generated according to the *sparse* DC-SBM (with parameters a, b and weight distribution with 2-nd moment $\Phi^{(2)}$), see Section 3. We characterise its leading eigenvalues and corresponding eigenvectors when the number of vertices in G tends to infinity.

Results in [72] show that the spectral method based on the non-backtracking matrix performs well in community detection on real datasets. A first step in testing the robustness of this method has been done in [17]: In the *ordinary* SBM, positively-correlated reconstruction can be obtained by thresholding the second-eigenvector of B when $(a - b)^2 > 2(a + b)$.

We extend this robustness result to the degree-corrected setting. Informally, we have the following results: With high probability, the leading eigenvalue of the non-backtracking matrix B is asymptotic to $\rho = \frac{a+b}{2}\Phi^{(2)}$. The second eigenvalue is asymptotic to $\mu_2 = \frac{a-b}{2}\Phi^{(2)}$ when $\mu_2^2 > \rho$, but asymptotically bounded by $\sqrt{\rho}$ when $\mu_2^2 \leq \rho$. All the remaining eigenvalues are asymptotically bounded by $\sqrt{\rho}$. Further, a clustering positively-correlated with the true communities can be obtained based on the second eigenvector of B in the regime where $\mu_2^2 > \rho$ (i.e., precisely when $(a - b)^2\Phi^{(2)} > 2(a + b)$).

A side-result is that Degree-Corrected Erdős-Rényi graphs asymptotically satisfy the graph Riemann hypothesis, a quasi-Ramanujan property.

In our proof we derive and use a weak law of large numbers for local-functionals on Degree-Corrected Stochastic Block Models, which could be of independent interest.

6.1.1 Linearisation of belief propagation

Following [91], we explain here how the non-backtracking matrix is naturally obtained by linearising belief propagation (Section 1.3.5). We restrict here to the *ordinary* SBM (with parameters a and b) on two equal-sized communities.

Consider (1.2), with initial messages

$$\psi_i^{u \rightarrow v} = \frac{1}{2} + \epsilon_i^{u \rightarrow v}, \quad (6.1)$$

where $\epsilon_i^{u \rightarrow v}$ are small random perturbation, forming vector ϵ . Linearising (1.2) gives the substitution scheme

$$\epsilon \mapsto M\epsilon, \quad (6.2)$$

with $M_{((u,v),i),((w,x),j)} = \frac{\partial \psi_i^{u \rightarrow v}}{\partial \psi_j^{w \rightarrow x}}$.

It turns out that M can be written as $M = B \otimes T$, where \otimes denotes the tensor product, B the non-backtracking matrix, and T the stochastic transition matrix given element-wise by

$$T_{ij} = \frac{p_{ij}}{\sum_k p_{ik}}.$$

The matrix T takes into account the contribution of a single edge coming in to u , say (w, u) , to the message transmitted from u to v . If (w, u) were the only incoming edge (we consider (u, v) here as the outgoing edge), the message u transmitted to v would be the belief based on $\psi^{w \rightarrow u}$ and the presence of edge (u, v) , i.e., $\psi^{u \rightarrow v} = T\psi^{w \rightarrow u}$. We note that we can rewrite $T = \lambda I + (1 - \lambda)\frac{J}{2}$ in terms of the identity matrix I , the all-one matrix J , and its second eigenvalue $\lambda = \frac{a-b}{a+b}$.

In general, there are multiple edges coming in, which explains the role of B in $M = B \otimes T$.

The updates (6.2) make sure that belief propagation drifts away from the trivial fixed point when the latter is unstable, i.e., when the eigenvalues of M exceed 1.

But, the eigenvalues of M are products of eigenvalues of B and T . As we shall later see, the leading eigenvalues of B are given by $\rho = \frac{a+b}{2}$, and $\mu = \frac{a-b}{2}$ when

$\mu^2 > \rho$. As reasoned in [91], we can disregard its first eigenvalue/eigenvector pair (since the messages should be normalized). The matrix T has leading eigenvalues 1 and λ , and again we can ignore 1. For the fixed-point to be unstable it is thus necessary that $\mu\lambda > 1$, that is $\mu^2 > \rho$.

6.1.2 Proof heuristics

We now give a heuristic proof for the above informal result (inspired also by [72]). Let ψ_1 and ψ_2 be the vectors on \vec{E} defined element-wise as $\psi_1(e) = 1$ and $\psi_2(e) = \sigma(e_2)$ and let $\ell \sim \log n$. We claim that asymptotically when $n \rightarrow \infty$, $B^\ell \psi_1$ and $B^\ell \psi_2$ are eigenvectors of B with corresponding eigenvalues ρ , respectively μ_2 .

We give an heuristic derivation. Firstly, for oriented edges e and f , B_{ef}^ℓ counts the number of non-backtracking paths of length ℓ from e to f . Now, the ℓ -neighbourhood of vertex e_2 looks like a tree with forward degree on average equal to ρ , see Section 5.7. Hence there are roughly ρ^ℓ vertices at distance ℓ from e_2 . Therefore,

$$(B^\ell \psi_1)(e) = \sum_f B_{ef}^\ell \cdot 1 \simeq \rho^\ell \psi_1(e), \quad (6.3)$$

since $\psi_1(e) = 1$. Consequently,

$$B(B^\ell \psi_1) \simeq \rho(B^\ell \psi_1).$$

Similarly,

$$(B^\ell \psi_2)(e) = \sum_f B_{ef}^\ell \cdot \sigma(f_2) \simeq Z_\ell^+ - Z_\ell^-, \quad (6.4)$$

where Z_ℓ^\pm denotes the number of vertices with spin \pm at distance ℓ from e_2 . From Section 5.7, we know that a particle has on average ρ children, each of them inheriting its spin with probability $\frac{a}{a+b}$. Hence,

$$\begin{aligned} \mathbb{E}[Z_\ell^+ - Z_\ell^- | Z_{\ell-1}^+, Z_{\ell-1}^-] &= \rho \left(\left(\frac{a}{a+b} - \frac{b}{a+b} \right) Z_{\ell-1}^+ - \left(\frac{a}{a+b} - \frac{b}{a+b} \right) Z_{\ell-1}^- \right) \\ &= \mu_2 (Z_{\ell-1}^+ - Z_{\ell-1}^-). \end{aligned} \quad (6.5)$$

Iteration gives,

$$\mathbb{E}[Z_\ell^+ - Z_\ell^- | \sigma(e_2)] = \mu_2^\ell \cdot \sigma(e_2).$$

Plugging this into (6.4), we obtain

$$(B^\ell \psi_2)(e) \simeq \mu_2^\ell \sigma(e_2) + \mathcal{O}(\rho^{\ell/2}) = \mu_2^\ell \left(\sigma(e_2) + \mathcal{O}\left(\left(\frac{\rho}{\mu_2^2}\right)^{\ell/2}\right) \right),$$

since fluctuations are of order $\sqrt{Z_\ell^+ + Z_\ell^-}$. Since we assume $\mu_2^2 > \rho$,

$$(B^\ell \psi_2)(e) \simeq \mu_2^\ell (\sigma(e_2) + o(1)).$$

That is,

$$B(B^\ell \psi_2) \simeq \mu_2(B^\ell \psi_2).$$

Consequently, if ξ_2 denotes the empirical normalised second eigenvector of B , then

$$\sum_{v \in \mathcal{N}(u)} \xi_2(v \rightarrow u) \simeq |\mathcal{N}(u)| \sigma(u).$$

Thus we can indeed reconstruct vertices by simply labelling vertex u according to the sign of $\sum_{v \in \mathcal{N}(u)} \xi_2(v \rightarrow u)$.

We also find, for the eigenvalues $\mu_1, \mu_2, \dots, \mu_m$ (with m equal to the number of edges), for arbitrary integer $l \geq 0$,

$$\sum_{i=1}^m |\mu_i|^l \leq \text{Tr} \left(B^l (B^l)^* \right) \simeq m \rho^l.$$

We conclude that *the bulk of eigenvalues* should be confined to the disk with radius $\sqrt{\rho}$ in the complex plane.

Our work consists in making the above precise, in particular bounding the norm of the third eigenvalue.

6.2 Main Results

We consider random graphs on n nodes $V = \{1, \dots, n\}$ drawn according to the *sparse* DC-SBM as in Section 3. Out of technical convenience we partition the vertices into *two* clusters of sizes n_+ and n_- , such that for some constant $\gamma \in (0, 1]$,

$$n_{\pm} = \frac{n}{2} + \mathcal{O}(n^{1-\gamma}). \quad (6.6)$$

Note that this is more general than labelling vertices uniformly at random.

The community-membership of a vertex v is again denoted by its spin $\sigma(v)$ from $\{+, -\}$ and its weight by ϕ_v . We assume in this chapter that the weights of vertices are bounded, i.e., there exists $0 < \phi_{\min} \leq \phi_{\max} < \infty$ such that ν has support in $[\phi_{\min}, \phi_{\max}]$. An edge is drawn between nodes u and v with probability $\frac{\phi_u \phi_v}{n} a$ when u and v have the same spin and with probability $\frac{\phi_u \phi_v}{n} b$ otherwise.

Local neighbourhoods in the sparse graphs under consideration are tree-like with high probability. In Section 5.7 we showed that these trees are distributed according to a Poisson-mixture two-type branching process. We denote the mean progeny matrix of the branching process by

$$M = \frac{\Phi^{(2)}}{2} \begin{pmatrix} a & b \\ b & a \end{pmatrix}. \quad (6.7)$$

We introduce the orthonormal vectors

$$g_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \text{ and } g_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad (6.8)$$

together with the scalars

$$\rho = \mu_1 = \frac{a+b}{2} \Phi^{(2)}, \text{ and } \mu_2 = \frac{a-b}{2} \Phi^{(2)}. \quad (6.9)$$

Then, g_k ($k = 1, 2$) are the left-eigenvectors of M associated to eigenvalues μ_k :

$$g_k^* M = \mu_k g_k^*, \quad k = 1, 2. \quad (6.10)$$

Finally, we define for $k \in \{1, 2\}$,

$$\chi_k(e) = g_k(\sigma(e_2)) \phi_{e_2}, \quad \text{for } e \in \vec{E}. \quad (6.11)$$

We show that the *candidate eigenvectors*

$$\zeta_k = \frac{B^l B^{*l} \check{\chi}_k}{\|B^l B^{*l} \check{\chi}_k\|} \quad (6.12)$$

are then, for $\ell \sim \log(n)$, asymptotically aligned with the first two eigenvectors of B . Note the weight in (6.11), which is *not* present in the ordinary SBM.

Theorem 6.2.1 (Degree-Corrected Extension of Theorem 4 in [17]). *Let G be drawn according to the DC-SBM such that assumption (6.6) holds. Assume that $\ell = C_{\min} \log(n)$, with $C_{\min} > 0$ a small constant defined in (6.14).*

If $\mu_2^2 > \rho$, then, with high probability, the eigenvalues λ_i of B satisfy

$$|\lambda_1 - \rho| = o(1), |\lambda_2 - \mu_2| = o(1), \quad \text{and, for } i \geq 3, \quad |\lambda_i| \leq \sqrt{\rho} + o(1).$$

Further, if, for $k \in \{1, 2\}$, ξ_k is a normalized eigenvector associated to λ_k , then ξ_k is asymptotically aligned with ζ_k . The vectors ξ_1 and ξ_2 are asymptotically orthogonal.

If $\rho > 1$, and $\mu_2^2 \leq \rho$, then, with high probability, the eigenvalues λ_i of B satisfy

$$|\lambda_1 - \rho| = o(1), \quad \text{and, for } i \geq 2, \quad |\lambda_i| \leq \sqrt{\rho} + o(1).$$

Further, ξ_1 is asymptotically aligned with ζ_1 .

Note that $\mu_2^2 > \rho$ implies $\rho > 1$, so that we consider the DC-SBM precisely in the regime where a giant component emerges, see [15].

In Theorem 6.2.2 we show that positively correlated clustering is possible based on the second eigenvector of B when above the feasibility threshold. More precisely, let $\hat{\sigma} = \{\hat{\sigma}(v)\}_{v \in V}$ be estimators for the spins of the vertices. Following [34], we say that $\hat{\sigma}$ has positive overlap with the true spin configuration $\sigma = \{\sigma(v)\}_{v \in V}$ if for some $\delta > 0$, with high probability,

$$\max_p \frac{1}{n} \sum_{v=1}^n 1_{\hat{\sigma}(v)=p \circ \sigma(v)} > \frac{1}{2} + \delta,$$

where p runs over the identity mapping on $\{+, -\}$ and the permutation that swaps $+$ and $-$.

Theorem 6.2.2 (Degree-Corrected Extension of Theorem 5 in [17]). *Let G be drawn according to the DC-SBM such that assumption (6.6) holds and such that $\mu_2^2 > \rho$. Let ξ_2 be the second normalized eigenvector of B .*

Then, there exists a deterministic threshold $\tau \in \mathbb{R}$, such that the following procedure yields asymptotically positive overlap: Put for vertex $v \in V$ its estimator $\hat{\sigma}(v) = +$ if $\sum_{e: e_2=v} \xi_2(e) > \frac{\tau}{\sqrt{n}}$ and put $\hat{\sigma}(v) = -$ otherwise.

6.2.1 Quasi Ramanujan property

Following the definition introduced in [79], a k -regular graph is Ramanujan if its second largest absolute eigenvalue is no larger than $2\sqrt{k-1}$. In [59], a graph is said to satisfy the graph Riemann hypothesis if B has no eigenvalues λ such that $|\lambda| \in (\sqrt{\rho_B}, \rho_B)$, where ρ_B is the Perron-Frobenius eigenvalue of B . The graph Riemann hypothesis can be seen as a generalization of the Ramanujan property, because a regular graph satisfies the graph Riemann hypothesis if and only if it has the Ramanujan property [59, 96].

Now, put $a = b = 1$ to obtain a Degree-Corrected Erdős-Rényi graph where vertices u and v are connected by an edge with probability $\frac{\phi_u \phi_v}{n}$. Our results imply that, with high probability, $\rho_B = \Phi^{(2)} + o(1)$, while all other eigenvalues are in absolute value smaller than $\sqrt{\Phi^{(2)}} + o(1)$. Consequently, these Degree-Corrected Erdős-Rényi graphs asymptotically satisfy the graph Riemann hypothesis.

6.2.2 Notation

We say that a sequence $(E_n)_n$ of events happens with high probability (w.h.p.) if $\lim_{n \rightarrow \infty} \mathbb{P}(E_n) = 1$.

We denote by $\|\cdot\|$ both the euclidean norm for vectors and the operator norm of matrices. I.e., for vectors $x = (x_1, \dots, x_m)$, and a matrix A , $\|x\| = \sqrt{\sum_{u=1}^m x_u^2}$, and $\|A\| = \sup_{x, \|x\|=1} \|Ax\|$.

Below we use that the neighbourhoods with a radius no larger than $C_{\text{coupling}} \log_\rho(n)$ can be coupled w.h.p. to certain branching processes, where

$$C_{\text{coupling}} := \frac{\left(\frac{1}{3} - \frac{1}{9} \log(4/e)\right) \wedge \left(\frac{1}{80} \wedge \frac{\gamma}{4}\right)}{\log_\rho(2(a+b)\phi_{\max}^2)}. \quad (6.13)$$

We put,

$$C_{\min} = \frac{1}{10} C_{\text{coupling}} \quad (6.14)$$

and consider often neighbourhoods of radius $C_{\min} \log_\rho(n)$.

We denote the k -th moment of the weight distribution ν by $\Phi^{(k)}$. I.e., $\mathbb{E}[\phi_1^k] = \Phi^{(k)}$.

The non-backtracking property for oriented edges $e, f \in \vec{E}$ is denoted by $e \rightarrow f$, i.e., $e_2 = f_1$ and $f_2 \neq e_1$.

In proofs, we often use the symbols c_1, c_2, \dots for suitably chosen constants.

6.3 Outline and proof strategy

We follow the same general approach as in [17]. We focus primarily on the differences and complications here: we often omit or shorten the proof of a statement if it may be proven in a very similar way.

We detail first our proof-strategy. Since B is neither symmetric nor a-priori normal, we cannot use standard tools such as the Bauer-Fike theorem.

Denote the top right eigenvectors of B by v_1 and v_2 , and the corresponding eigenvalues by λ_1 , respectively λ_2 . Assume that we can write

$$B = \lambda_1 v_1 w_1^* + \lambda_2 v_2 w_2^* + R,$$

with R a matrix with norm bounded by $\sqrt{\lambda_1}$, $w_i^* v_i = 1$, for all i , $w_i^* v_j = w_i^* w_j = v_i^* v_j = 0$ for $i \neq j$ and $w_i^* R = 0$ for all i . Then,

$$B^\ell = \lambda_1^\ell v_1 w_1^* + \lambda_2^\ell v_2 w_2^* + R^\ell. \quad (6.15)$$

Now, if $\lambda_2^2 > \lambda_1$, then $\|R^\ell\| = \mathcal{O}\left(\lambda_1^{\ell/2}\right)$, and hence R^ℓ is a small perturbation in (6.15). This intuition leads to an extension of the Bauer-Fike theorem, given in Proposition 6.4.1: If

$$B^\ell = \rho^\ell x_1 y_1^* + \mu_2^\ell x_2 y_2^* + R_\ell, \quad (6.16)$$

with $x_i = x_i(\ell)$, $y_i = y_i(\ell)$, R_ℓ as in the conditions of Proposition 6.4.1, then

$$\lambda_1 - \rho = o(1), \quad \lambda_2 - \mu_2 = o(1), \quad \text{and,} \quad \left\| \frac{x_i}{\|x_i\|} - v_i \right\| = o(1).$$

How should we choose x_i and y_i so that (6.16) is true? Firstly, to meet the bound on R_ℓ we have to control $\|B^\ell x\|$ when $x \perp y_i$. We do this by decomposing powers of B as a sum of products, using a technique that first appeared in [85] and [17]. In particular, we obtain Proposition 6.10.1, from which it follows that we have to bound

$\langle \check{\chi}_j, B^{\ell-t-1}x \rangle$ for $x \perp y_i$ and $t \leq \ell$. Now we show that this term is small if $\langle \check{\varphi}_i, x \rangle = 0$, where $\varphi_i = \frac{B^\ell \chi_i}{\|B^\ell \chi_i\|}$. It is therefore natural to choose $y_i = \check{\varphi}_i$. But then, if $c_i := y_i y_i^*$ is bounded away from zero and $y_1^* y_2 \rightarrow 0$,

$$B^\ell \check{\varphi}_1 = \rho^\ell c_1 x_1 + \mathcal{O}(\rho^{\ell/2}),$$

and,

$$B^\ell \check{\varphi}_2 = \mu_2^\ell c_2 x_2 + \mathcal{O}(\rho^{\ell/2}),$$

where the big "O" term denotes a vector of norm $\mathcal{O}(\rho^{\ell/2})$. This implies that x_i should be parallel with $B^\ell \check{\varphi}_i$.

We therefore need a careful analysis of $\varphi_1, \varphi_2, B^\ell \check{\varphi}_1$, and $B^\ell \check{\varphi}_2$. We analyse them first on branching processes, as *most* neighbourhoods are locally tree-like. We should be careful here, because with non-zero probability certain neighbourhoods cannot be accurately described with a branching process. However, to calculate, for instance, $\varphi_1^* \varphi_2 = \sum_{e \in \vec{E}} \varphi_1(e) \varphi_2(e)$, we could circumvent this issue by establishing a law of large numbers for local functionals in order to replace $\frac{1}{n} \sum_{e \in \vec{E}} \varphi_1(e) \varphi_2(e)$ by $\mathbb{E} \left[\frac{B^\ell \chi_1(o)}{\|B^\ell \chi_1\|} \cdot \frac{B^\ell \chi_2(o)}{\|B^\ell \chi_2\|} \right]$ on a branching process with root o having spin uniformly picked from \pm . From the heuristics above, we expect $\mathbb{E} \left[\frac{B^\ell \chi_1(o)}{\|B^\ell \chi_1\|} \cdot \frac{B^\ell \chi_2(o)}{\|B^\ell \chi_2\|} \right] \simeq \mathbb{E}[\sigma_o] = 0$.

We detail now the outline of the rest of this chapter.

In Section 6.4, we give the necessary background on non-backtracking matrices. Further, we give an extension of the Bauer-Fike Theorem, that first appeared in [17].

In Section 6.5 we give the proof of Theorem 6.2.1. It builds on Propositions 6.5.1 and 6.5.2. Their proofs are deferred to later sections.

In Section 6.6 we consider two-type branching process where the offspring distribution is governed by a Poisson mixture to capture the weights of the vertices. We associate two martingales to this process and extend limiting results by Kesten and Stigum [70, 71]. Hoeffding's inequality plays an important role here to prove concentrations results for the weights. Further, we define a cross-generational functional on these branching processes that is correlated with the spin of the root.

In Section 5.7 we state a coupling between local neighbourhoods and the branching process with weights in Section 6.6. This coupling is slightly different from Section 5.7. It is technically more involved than the ordinary coupling on graphs with unit weight. It is crucial that the weights in the graph and the branching process *exactly* coincide. We further establish a growth condition on the local neighbourhoods, using a stochastic domination argument that is more involved than its analogue in unweighed graphs.

In Section 6.8 we define local functionals that map graphs, together with their spins *and* weights to the real numbers. We establish, using Efron-Stein's inequality, a weak law of large numbers for those functionals, which could be of independent interest. Part of the work here is again hidden in the coupling from Section 5.7.

In Section 6.9 we apply those local functionals to establish Proposition 6.5.1.

In Section 6.10 we decompose powers of the matrix B as a sum of products. This technique appeared first in [85] for matrices counting self-avoiding paths and was elaborated in [17]. To bound the norm of the individual matrices occurring in the decomposition, we use the trace method initiated in [46]. In doing so, we need to bound the expectation of products of higher moments of the weights over certain paths. This is a significant complication with respect to the ordinary SBM, see Section 6.10.2 for a comparison.

In Section 6.11 we prove that positively correlated clustering is possible based on the second eigenvector of B , i.e., Theorem 6.2.2. We use the symmetry present in the two-communities setting here, which gets in general broken in models with more than two communities.

Detailed proofs of the statements in Sections 6.6, 5.7, 6.8, 6.10 and 6.11 can be found in Sections 6.12 - 6.16.

In each section we give a detailed comparison with the ordinary SBM.

6.4 Preliminaries

6.4.1 Background on non-backtracking matrix

We repeat here the most important observations made in [17].

Firstly, for any $k \geq 1$, B_{ef}^k counts the number of non-backtracking paths between oriented edges e and f . A non-backtracking path is defined as an oriented path between two oriented edges such that no edge is the inverse of its preceding edge, i.e., the path makes no backtrack.

Another important observation is that $(B^*)_{ef} = B_{fe} = B_{e^{-1}f^{-1}}$, where for oriented edge $e = (e_1, e_2)$, we set $e^{-1} = (e_2, e_1)$. If we introduce the swap notation, for $x \in \mathbb{R}^{\vec{E}}$,

$$\check{x}_e = x_{e^{-1}}, \quad e \in \vec{E},$$

then for any $x, y \in \mathbb{R}^{\vec{E}}$, and integer $k \geq 0$,

$$\langle y, B^k x \rangle = \langle B^k \check{y}, \check{x} \rangle.$$

Denote by P the matrix on $\mathbb{R}^{\vec{E} \times \vec{E}}$, defined on oriented edges e, f as

$$P_{ef} = 1_{f=e^{-1}}.$$

Then, $Px = \check{x}$, $P^* = P$ and $P^{-1} = P$. Further,

$$(B^k P)^* = P(B^*)^k = B^k P,$$

so that we can write the symmetric matrix $B^k P$ in diagonal form: Let $(\sigma_{k,j})_j$ be eigenvalues of $B^k P$ ordered in decreasing order of absolute value, and let $(x_{k,j})_j$ be the corresponding orthonormal eigenvectors. Then,

$$B^k = (B^k P)P = \sum_j \sigma_{k,j} x_{k,j} x_{k,j}^* P = \sum_j \sigma_{k,j} x_{k,j} \check{x}_{k,j}^* = \sum_j s_{k,j} x_{k,j} y_{k,j}^*, \quad (6.17)$$

where $s_{k,j} = |\sigma_{k,j}|$ and $y_{k,j} = \text{sign}(\sigma_{k,j}) \check{x}_{k,j}$. Since P is an orthogonal matrix, $(\check{x}_{k,j})_j$ form an orthonormal base for $\mathbb{R}^{\vec{E}}$ and the term furthest on the right of (6.17) is thus the spectral value decomposition of B^k . Now, if B is irreducible and if ξ denotes the normalized Perron eigenvector of B with eigenvalue $\lambda_1(B) > 0$, we have $\lambda_1(B) = \lim_{k \rightarrow \infty} (\sigma_{k,1})^{1/k}$, and $\lim_{k \rightarrow \infty} \|x_{k,1} - \xi\| = 0$.

In [17], the Bauer-Fike Theorem is extended to prove the spectral claims we make here.

6.4.2 Extension of Bauer-Fike Theorem

Tailored to our needs, we use the following proposition from [17]:

Proposition 6.4.1 (Special case of Proposition 8 in [17]). *Let $\ell = C \log_\rho n$, with $C > 0$. Let $A \in M_n(\mathbb{R})$, such that for some vectors $x_1 = x_{\ell,1}, y_1 = y_{\ell,1}, x_2 = x_{\ell,2}, y_2 = y_{\ell,2} \in \mathbb{R}$, some matrix $R_\ell \in M_n(\mathbb{R})$, and some non-zero constants $\rho > \mu_2$ with $\mu_2^2 > \rho$,*

$$A^\ell = \rho^\ell x_1 y_1^* + \mu_2^\ell x_2 y_2^* + R_\ell. \quad (6.18)$$

Assume there exist $c_0, c_1 > 0$ such that for all $i \in \{1, 2\}$, $\langle y_i, x_i \rangle \geq c_0$, $\|x_i\| \|y_i\| \leq c_1$. Assume further that $\langle x_1, y_2 \rangle = \langle x_2, y_1 \rangle = \langle x_1, x_2 \rangle = \langle y_1, y_2 \rangle = 0$ and for some $c > 0$

$$\|R_\ell\| < \rho^{\ell/2} \log^c(n).$$

Let $(\lambda_i)_{1 \leq i \leq n}$, be the eigenvalues of A with $|\lambda_n| \leq \dots \leq |\lambda_1|$. Then,

$$|\lambda_1 - \rho| = o(1), |\lambda_2 - \mu_2| = o(1), \quad \text{and, for } i \geq 3, \quad |\lambda_i| \leq \sqrt{\rho} + o(1).$$

Further, there exist unit eigenvectors ψ_1, ψ_2 of A with eigenvalues λ_1 , respectively λ_2 such that

$$\|\psi_i - \frac{x_i}{\|x_i\|}\| = o(1).$$

Proof. This is a special case of Proposition 8 in [17]. In the notation of the latter, we have $\ell' = \ell - 2$, $\theta_1 = \rho$, $\theta_2 = \mu_2$, $\theta = \mu_2$, $\gamma \geq \frac{a+b}{|a-b|} > 1$. Further $\frac{c_0(c_0\gamma^k - c_1)_+}{4c_1} \wedge \frac{c_0^2}{2(\ell\sqrt{\ell'})c_1} \sim \frac{1}{\log_\rho n}$, and thus

$$\|R_\ell\| \leq \log^c(n) \left(\frac{\sqrt{\rho}}{|\mu_2|} \right)^\ell |\mu_2|^\ell = o(1) \frac{1}{\log_\rho n} |\theta|^\ell.$$

□

To prove the case $\mu_2^2 > \rho$ of Theorem 6.2.1, we thus need to find candidate vectors x_1, x_2, y_1 and y_2 that meet the conditions in Proposition 6.4.1 and further verify that the remainder R_ℓ has small norm. Note that the last condition is true whenever $\|B^\ell x\| \leq \rho^{\ell/2} \log^c(n)$ for all normalized x in $\text{span}\{y_1, y_2\}^\perp$.

To address the case $\mu_2^2 \leq \rho$ of Theorem 6.2.1, we appeal to Proposition 7 in [17], which is very similar in spirit to Proposition 6.4.1.

6.5 Proof of Theorem 6.2.1

6.5.1 The case $\mu_2^2 > \rho$.

We start with the case $\mu_2^2 > \rho$. We decompose, for some vectors x_1, y_1, x_2 and y_2 and matrix R_ℓ ,

$$B^\ell = \rho^\ell x_1 y_1^* + \mu_2^\ell x_2 y_2^* + R_\ell,$$

and we show that the assumptions of Proposition 6.4.1 are met.

Let ℓ be as in Theorem 6.2.1 and recall χ_k and ζ_k from (6.11) and (6.12). For ease of notation, we introduce for $k \in \{1, 2\}$,

$$\varphi_k = \frac{B^\ell \chi_k}{\|B^\ell \chi_k\|}, \quad \text{and } \theta_k = \|B^\ell \check{\varphi}_k\|. \quad (6.19)$$

Then, $\zeta_k = \frac{B^\ell \check{\varphi}_k}{\theta_k}$.

To prove the main theorem, we need the following two propositions. The proofs are deferred to Section 6.9 and 6.10.1. The material in Section 6.9 builds on ingredients from Sections 5.7 - 6.8, where we assume that $\mu_2^2 > \rho$, unless stated otherwise.

Proposition 6.5.1 (Degree-Corrected Extension of Proposition 19 in [17]). *Assume that $\mu_2^2 > \rho$. Let $\ell = C \log_\rho n$ with $0 < C < C_{\min}$. For some $b, c > 0$, with high probability,*

$$(i) \quad b |\mu_k^\ell| \leq \theta_k \leq c |\mu_k^\ell| \text{ if } k \in \{1, 2\},$$

(ii) $\text{sign}(\mu_k^\ell) \langle \zeta_k, \check{\varphi}_k \rangle \geq b$ if $k \in \{1, 2\}$,

(iii) $|\langle \varphi_1, \varphi_2 \rangle| \leq (\log n)^3 n^{C - (\frac{\gamma}{2} \wedge \frac{1}{40})}$,

(iv) $|\langle \zeta_j, \check{\varphi}_k \rangle| \leq (\log n)^3 n^{\frac{3}{2}C - (\frac{\gamma}{2} \wedge \frac{1}{40})}$ if $k \neq j \in \{1, 2\}$.

(v) $|\langle \zeta_1, \zeta_2 \rangle| \leq (\log n)^8 n^{2C - (\frac{\gamma}{2} \wedge \frac{1}{40})}$.

Put $H = \text{span}\{\check{\varphi}_1, \check{\varphi}_2\}$, then

Proposition 6.5.2 (Degree-Corrected Extension of Proposition 20 in [17]). *Let $\ell = C \log_\rho n$ with $0 < C < C_{\min}$. For some $c > 0$, with high probability,*

$$\sup_{x \in H^\perp, \|x\|=1} \|B^\ell x\| \leq (\log n)^c \rho^{\ell/2}. \quad (6.20)$$

Put $\bar{\varphi}_1 = \check{\varphi}_1$, and $\bar{\varphi}_2 = \frac{\check{\varphi}_2 - \langle \check{\varphi}_1, \check{\varphi}_2 \rangle \check{\varphi}_1}{\|\check{\varphi}_2 - \langle \check{\varphi}_1, \check{\varphi}_2 \rangle \check{\varphi}_1\|}$, then $\bar{\varphi}_1$ and $\bar{\varphi}_2$ are orthonormal and $\|\bar{\varphi}_2 - \check{\varphi}_2\| = o(\rho^{-\ell/2})$, due to Proposition 6.5.1 (iii).

Let $\bar{\zeta}_1$ be the normalized orthogonal projection of ζ_1 on $\text{span}\{\bar{\varphi}_2\}^\perp$. Similarly, let $\bar{\zeta}_2$ be the normalized orthogonal projection of ζ_2 on $\text{span}\{\bar{\zeta}_1, \bar{\varphi}_1\}^\perp$.

Then $\langle \bar{\zeta}_1, \bar{\zeta}_2 \rangle = 0$ and for $i = 1, 2$, $\|\bar{\zeta}_i - \zeta_i\| = o(\rho^{-\ell/2})$, as follows from Proposition 6.5.1 (iv) and (v).

We set

$$D = \theta_1 \bar{\zeta}_1 \bar{\varphi}_1^* + \theta_2 \bar{\zeta}_2 \bar{\varphi}_2^* = \rho^\ell \left(\frac{\theta_1}{\rho^\ell} \bar{\zeta}_1 \right) \bar{\varphi}_1^* + \mu_2^\ell \left(\frac{\theta_2}{\mu_2^\ell} \bar{\zeta}_2 \right) \bar{\varphi}_2^*.$$

Note that,

$$\|B^\ell \bar{\varphi}_1\| = \theta_1 = O(\rho^\ell),$$

and

$$\|B^\ell \bar{\varphi}_2\| = \|B^\ell ((1 + o(1))\check{\varphi}_2 + o(1)\bar{\varphi}_1)\| = O(\rho^\ell).$$

As a consequence, from Proposition 6.5.2,

$$\|B^\ell\| = O(\rho^\ell).$$

Since $D\bar{\varphi}_i = B^\ell \check{\varphi}_i + \theta_i(\bar{\zeta}_i - \zeta_i)$,

$$\|B^\ell \bar{\varphi}_i - D\bar{\varphi}_i\| \leq \|B^\ell\| \|\bar{\varphi}_i - \check{\varphi}_i\| + \theta_i \|\bar{\zeta}_i - \zeta_i\| = \mathcal{O}(\rho^{\ell/2}).$$

Let P be the orthogonal projection on $H = \text{span}\{\bar{\varphi}_1, \bar{\varphi}_2\} = \text{span}\{\check{\varphi}_1, \check{\varphi}_2\}$, then $\|B^\ell P - D\| = \mathcal{O}(\rho^{\ell/2})$.

Put $R_\ell = B^\ell - D$. Write for $y \in \mathbb{R}^{\bar{E}}$ with unit norm, $y = h + h^\perp$, with $h \in H$ and $h^\perp \in H^\perp$, then

$$\begin{aligned} \|R_\ell y\| &= \|B^\ell h^\perp + (B^\ell - D)h\| \\ &\leq \sup_{x \in H^\perp, \|x\|=1} \|B^\ell x\| + \|B^\ell P - D\| \\ &= \mathcal{O}(\log^c(n) \rho^{\ell/2}), \end{aligned} \quad (6.21)$$

as follows from Proposition 6.5.2.

We finish by applying Proposition 6.4.1 with $x_1 = \frac{\theta_1}{\rho^\ell} \bar{\zeta}_1$, $y_1 = \bar{\varphi}_1$, $x_2 = \frac{\theta_2}{\mu_2^\ell} \bar{\zeta}_2$, and, $y_2 = \bar{\varphi}_2$.

6.5.2 The case $\mu_2^2 \leq \rho$.

In case $\mu_2^2 \leq \rho$, Proposition 6.5.1 (i) and (ii) continue to hold for $k = 1$. Further, Proposition 6.5.1 (iii) as well as Proposition 6.5.2 continue to hold. We need however the following bound for $k = 2$:

Proposition 6.5.3. *Assume that $\mu_2^2 \leq \rho$. Let $\ell = C \log_\rho n$ with $0 < C < C_{\min}$. For some $c > 0$, with high probability,*

$$\theta_2 \leq (\log n)^c \rho^{\ell/2}.$$

Using this proposition and $\|\bar{\varphi}_2 - \check{\varphi}_2\| = o(\rho^{-\ell/2})$, we get

$$\|B^\ell \bar{\varphi}_2\| \leq (\log n)^{c+1} \rho^{\ell/2}.$$

It remains to apply Proposition 7 from [17].

6.6 Poisson-mixture two-type branching processes

The proofs of the statements in this section are deferred to Section 6.12.

6.6.1 A theorem of Kesten and Stigum

We consider the two-type Poisson-mixture branching process defined in Section 5.4.

We use the notation $Z_t = \begin{pmatrix} Z_t^{(+)} \\ Z_t^{(-)} \end{pmatrix}$ for the population at generation $t \geq 1$, where $Z_t(\pm)$ is the number of type \pm particles in generation t . We let $(\mathcal{F}_t)_{t \geq 1}$ denote the natural filtration associated to $(Z_t)_{t \geq 1}$.

We associate two matrices to the branching process, namely M defined in (6.7), and, for a root with weight ϕ_o ,

$$M_{\phi_o} = \frac{\Phi^{(1)} \phi_o}{\Phi^{(2)}} M. \quad (6.22)$$

Then, M is the *transition* matrix for generations $t \geq 1$ and later:

$$\mathbb{E}[Z_{t+1}|Z_t] = M Z_t, \quad \text{for all } t \geq 1, \quad (6.23)$$

and M_{ϕ_o} describes the transition from the root to the first generation:

$$\mathbb{E}[Z_1|Z_0, \phi_o] = M_{\phi_o} Z_0, \quad (6.24)$$

where, by assumption $Z_0 = \begin{pmatrix} 1_{\sigma_o=+} \\ 1_{\sigma_o=-} \end{pmatrix}$. Note that the difference between the root and later generations stems from the fact that the root's weight is deterministic in the conditional expectation, whereas the weight of a particle in any later generation has expectation $\frac{\Phi^{(2)}}{\Phi^{(1)}}$.

Recall from (6.10) that g_k ($k = 1, 2$) are the left-eigenvectors of M associated to eigenvalues μ_k :

$$g_k^* M = \mu_k g_k^*, \quad k = 1, 2. \quad (6.25)$$

Note that M_{ϕ_o} has the same left-eigenvectors as M , while the corresponding eigenvalues are given by

$$\mu_{k, \phi_o} = \frac{\Phi^{(1)} \phi_o}{\Phi^{(2)}} \mu_k, \quad k = 1, 2. \quad (6.26)$$

Theorem 6.6.1 shows that a Kesten-Stigum theorem applies to the "classical" branching process obtained after restricting the above process to generations 1 and later: We expect

$$\frac{\langle g_k, Z_t \rangle}{\mu_k^{t-1}} = \frac{\mathbb{E}[\langle g_k, Z_t \rangle] + \mathcal{O}(\sqrt{Z_t})}{\mu_k^{t-1}} = \langle g_k, Z_1 \rangle + o(1),$$

since $\mu_2^2 > \mu_1 = \rho > 1$. Corollary 6.6.2, then, joins this classical branching process to the transition from the root to generation 1.

We further consider the vector $\Psi_t = (\Psi_t(+), \Psi_t(-))$, containing sums of the weights,

$$\Psi_t(\pm) = \sum_{u \in Y_t} 1_{\sigma_u = \pm} \phi_u, \quad (6.27)$$

where Y_t is the set of particles at distance t from the root, and where ϕ_u and σ_u denote the weight respectively spin of a particle u . Note that $\Psi_t = Z_t$ in case of unit weights.

The martingale Theorem 6.6.3 is not present in [17]. We need it to bound the variance of the cross-generational functional defined in Section 6.6.3.

Theorem 6.6.1 (Degree-Corrected Extension of Theorem 21 in [17]). *Assume that $\mu_2^2 > \rho$. Put $\mathcal{F}_t = \{Z_s\}_{s \leq t}$. For any $k = 1, 2$,*

$$\left(X_k(t) := \frac{\langle g_k, Z_t \rangle}{\mu_k^{t-1}} - \langle g_k, Z_1 \rangle \right)_{t \geq 1},$$

is an \mathcal{F}_t -martingale converging a.s. and in L^2 such that for some $C > 0$ and all $t \geq 1$, $\mathbb{E}[X_k(t)] = 0$ and $\mathbb{E}[X_k^2(t)|Z_1] \leq C\|Z_1\|_1$.

Corollary 6.6.2. *Assume that $\mu_2^2 > \rho$. For $k = 1, 2$, with the weight $\phi_o = \psi_o$ of the root fixed, the sequence of random variables $(Y_{k, \psi_o}(t))_{t \geq 1} = \left(\frac{\langle g_k, Z_t \rangle}{\mu_k^{t-1} \mu_{k, \psi_o}} \right)_{t \geq 1}$ converges almost surely and in L^2 to a random variable $Y_{k, \psi_o}(\infty)$ with $\mathbb{E}[Y_{k, \psi_o}(\infty)|\sigma_o] = g_k(\sigma_o)$. Further, the L^2 -convergence takes place uniformly over all ψ_o .*

Theorem 6.6.3. *Assume that $\mu_2^2 > \rho$. Put $\mathcal{G}_t = \{\Psi_s\}_{s \leq t}$. For any $k = 1, 2$,*

$$\left(X_k(t) := \frac{\langle g_k, \Psi_t \rangle}{\mu_k^{t-1}} - \langle g_k, \Psi_1 \rangle \right)_{t \geq 1},$$

is an \mathcal{G}_t -martingale converging a.s. and in L^2 such that for some $C > 0$ and all $t \geq 1$, $\mathbb{E}[X_k(t)] = 0$ and $\mathbb{E}[X_k^2(t)|Z_1] \leq C\|Z_1\|_1$.

6.6.2 Quantitative version of the Kesten-Stigum theorem

We now quantify the growth of the population size. The latter is defined as

$$S_t = \|Z_t\|_1, \quad t \geq 0,$$

i.e., the number of individuals in generation $t \geq 0$. Given S_t , for $t \geq 1$ we have

$$S_{t+1} = \text{Poi} \left(\sum_{l=1}^{S_t} X_t^{(l)} \right), \quad (6.28)$$

where $(X_t^{(l)})_l$ are i.i.d. copies of $\frac{a+b}{2}\Phi^{(1)}\phi^*$, where ϕ^* follows law ν^* .

Note that in the *ordinary* Stochastic Block Model (i.e., when all vertices have unit weight), the argument of the Poisson random variables in (6.28) is deterministic, contrary to the general case under consideration here. Using (6.23) recursively in conjunction with (6.24), it follows that

$$\mathbb{E}[S_t|\phi_o] = \frac{\Phi^{(1)}\phi_o}{\Phi^{(2)}}\rho^t, \quad \forall t \geq 1.$$

In the following lemma we show that deviations from this average are small. In fact, there exists a constant C such that for each $t \geq 0$, S_t is asymptotically stochastically dominated by an Exponential random variable with mean $C\rho^t$. An important ingredient in the proof below is Hoeffding's inequality, which we use to derive a concentration result for the parameter of the Poisson variable in (6.28).

Heuristically, $\sum_{l=1}^{S_t} X_t^{(l)} = \rho S_t + \mathcal{O}(\sqrt{S_t})$, so that in conjunction with large deviation bounds for Poisson random variables (for large $\lambda > 0$, with large probability, $\text{Poi}(\lambda) \leq \lambda + \mathcal{O}(\sqrt{\lambda})$), we expect $\text{Poi}\left(\sum_{l=1}^{S_t} X_t^{(l)}\right) \leq \rho S_t(1 + o(1))$.

Lemma 6.6.4 (Degree-Corrected Extension of Lemma 23 in [17]). *Assume $S_0 = 1$. There exist $c, c' > 0$ such that for all $s \geq 0$,*

$$\mathbb{P}\left(\forall k \geq 1, S_k \leq s\rho^k\right) \geq 1 - c'e^{-cs}.$$

From Theorem 6.6.1 and Corollary 6.6.2, we know that the different components (expressed in the basis of eigenvectors of M) grow exponentially with rate ρ , respectively μ_2 . We now quantify the error. Recall Ψ_t from (6.27).

In case $\mu_2^2 > \rho$, we have

Theorem 6.6.5 (Degree-Corrected Extension of Theorem 24 in [17]). *Assume that $\mu_2^2 > \rho$. Let $\beta > 0$, $Z_0 = \delta_x$ and $\phi_o = \psi_o$ be fixed. There exists $C = C(x, \beta) > 0$ such that with probability at least $1 - n^{-\beta}$, for all $k \in \{1, 2\}$, all $0 \leq s < t \leq C_{\min} \log(n)$, with $0 \leq s < t$,*

$$|\langle g_k, Z_s \rangle - \mu_k^{s-t} \langle g_k, Z_t \rangle| \leq C(s+1)\rho^{s/2}(\log n)^{3/2},$$

and,

$$|\langle g_k, \Psi_s \rangle - \mu_k^{s-t} \langle g_k, \Psi_t \rangle| \leq C\rho^{s/2}(\log n)^{5/2}.$$

In case $\mu_2^2 \leq \rho$, we have

Theorem 6.6.6. *Assume that $\mu_2^2 \leq \rho$. Let $\beta > 0$, $Z_0 = \delta_x$ and $\phi_o = \psi_o$ be fixed. There exists $C = C(x, \beta) > 0$ such that with probability at least $1 - n^{-\beta}$, for all $t \geq 1$,*

$$|\langle g_2, \Psi_t \rangle| \leq Ct^2\rho^{t/2}(\log n)^2,$$

and,

$$\mathbb{E}\left[|\langle g_2, \Psi_t \rangle|^2\right] \leq Ct^3\rho^t.$$

6.6.3 $B^\ell B^{*\ell} \tilde{\chi}_k$ on trees: a cross generation functional

Recall our claim that $B^\ell B^{*\ell} \tilde{\chi}_k$ are asymptotically aligned with the eigenvectors of B . In the DC-SBM, the local-neighbourhood of a vertex has with high probability a tree-like structure described by the branching process above. In this section we analyse $B^\ell B^{*\ell} \tilde{\chi}_k$ on trees.

To this end we define a cross-generational functional slightly different from its analogue in [17] due to the presence of weights:

$$Q_{k,\ell} = \sum_{(u_0, \dots, u_{2\ell+1}) \in \mathcal{P}_{2\ell+1}} g_k(\sigma(u_{2\ell+1})) \phi_{u_{2\ell+1}}, \quad (6.29)$$

where $\mathcal{P}_{2\ell+1}$ is the set of paths $(u_0, \dots, u_{2\ell+1})$ (of length $2\ell + 1$) in the tree starting from $u_0 = o$ with both (u_0, \dots, u_ℓ) and $(u_\ell, \dots, u_{2\ell+1})$ non-backtracking and $u_{\ell-1} = u_{\ell+1}$. Note that these paths thus make a back-track exactly once at step $\ell + 1$.

Explicitly, we have

$$Q_{1,\ell} = \sum_{(u_0, \dots, u_{2\ell+1}) \in \mathcal{P}_{2\ell+1}} \frac{1}{\sqrt{2}} \phi_{u_{2\ell+1}}, \quad (6.30)$$

and,

$$Q_{2,\ell} = \sum_{(u_0, \dots, u_{2\ell+1}) \in \mathcal{P}_{2\ell+1}} \frac{1}{\sqrt{2}} \sigma(u_{2\ell+1}) \phi_{u_{2\ell+1}}. \quad (6.31)$$

Consider a tree \mathcal{T}' and a leaf e_1 on it that has unique neighbour, say, o . Then, if e is the oriented edges from e_1 to o and if $B_{\mathcal{T}'}$ denotes the non-backtracking matrix defined on \mathcal{T}' ,

$$\left(B_{\mathcal{T}'}^\ell, B_{\mathcal{T}'}^{*\ell}, \check{\chi}_k \right) (e) = Q_{k,\ell} + g_k(\sigma(e_1)) \phi_{e_1} \|Z_\ell\|_1, \quad (6.32)$$

where $Q_{k,\ell}$ and Z_ℓ are defined on the tree \mathcal{T} with root o obtained after removing vertex e_1 from \mathcal{T}' .

In the sequel we analyse $Q_{k,\ell}$ on the branching process defined above, starting with a single particle, the root o . Let V indicate the particles of the random tree. Denote the spin of a particle $v \in V$ by $\sigma_v \in \{+, -\}$ and its weight by $\phi_v \in S$.

For $t \geq 0$, let Y_t^v denote the set of particles, *including* their spins and weights, of generation t from v in the subtree of particles with common ancestor $v \in V$. Let $Z_t^v = (Z_t^{v,+}, Z_t^{v,-})$ denote the number of \pm vertices in generation t ; i.e., $Z_t^{v,\pm} = \sum_{u \in Y_t^v} 1_{\sigma(u)=\pm}$. Finally, let $\Psi_t^v = (\Psi_t^{v,+}, \Psi_t^{v,-})$, with $\Psi_t^{v,\pm} = \sum_{u \in Y_t^v} 1_{\sigma(u)=\pm} \phi_u$.

We rewrite $Q_{k,\ell}$ into a more manageable form: First observe that every path in $\mathcal{P}_{2\ell+1}$, after reaching $u_{\ell+1}$, climbs back to a depth t from which it then again moves down the tree (that is, in the direction away from the root). Let us call the vertex at level t (to which the path climbs back before descending again) u . Then, (if $t \neq 0$) there are two children of u , say v and w such that w lies on the path between u and $u_{\ell+1}$ and v is in between u and $u_{2\ell+1}$. For such fixed v and w in Y_1^u , only the children $u_{2\ell+1} \in Y_t^v$ determine the contribution of a path to (6.29), regardless of the choice of $u_{\ell+1} \in Y_{\ell-t-1}^w$. Hence, for such fixed u and $v, w \in Y_1^u$ and $u_{2\ell+1}$, there are $|Y_{\ell-t-1}^w| = S_{\ell-t-1}^w$ paths giving the same contribution to (6.29):

$$Q_{k,\ell} = \sum_{t=0}^{\ell-1} \sum_{u \in Y_t^o} L_{k,\ell}^u, \quad (6.33)$$

where, for $|u| = t \geq 0$,

$$L_{k,\ell}^u = \sum_{w \in Y_1^u} S_{\ell-t-1}^w \left(\sum_{v \in Y_1^u \setminus \{w\}} \langle g_k, \Psi_t^v \rangle \right). \quad (6.34)$$

The following theorem is an extension of Theorem 25 in [17]. The important observation is that, again, for $Z_0 = \delta_\tau$ fixed, $(Q_{2,\ell}/\mu_2^{2\ell})_\ell$ converges to a random

variable with mean a constant times τ , that is, the spin of the root. Its proof uses both martingale theorems stated above. We use the second martingale statement, which is not present in the ordinary SBM, to bound the variance of $Q_{k,\ell}$ in case $\mu_2^2 > \rho$:

Theorem 6.6.7 (Degree-Corrected Extension of Theorem 25 in [17]). *Assume that $\mu_2^2 > \rho$. Let $Z_0 = \delta_x$ and $\phi_o = \psi_o$ be fixed. For $k \in \{1, 2\}$, $(Q_{k,\ell}/\mu_k^{2\ell})_\ell$ converges in L^2 as ℓ tends to infinity to a random variable with mean $\frac{\Phi^{(3)}}{\Phi^{(2)}} \frac{\rho}{\mu_k^2 - \rho} \mu_{k,\psi_o} g_k(x)$. Further, the L^2 -convergence takes place uniformly for all ψ_o .*

In case $\mu_2^2 \leq \rho$, we have:

Theorem 6.6.8. *Assume that $\mu_2^2 \leq \rho$. Let $Z_0 = \delta_x$ and $\phi_o = \psi_o$ be fixed. There exists a constant C such that $\mathbb{E} \left[Q_{2,\ell}^2 \right] \leq C \rho^{2\ell} \ell^5$.*

6.6.4 Orthogonality: Decorrelation in branching process

Again, as in [17], $Q_{1,\ell}$ and $Q_{2,\ell}$ are uncorrelated when defined on the branching process above. The proof presented here is simpler than the corresponding one in [17] and uses that for the two communities-case, $Q_{1,\ell}$ and $Q_{2,\ell}$ are *explicitly* known.

The orthogonality of the candidate eigenvectors (i.e., (iii) – (v) in Proposition 6.5.1) follows from this fact, see Proposition 6.8.3 (ii), (iii) and Proposition 6.8.4 (ii) below.

Theorem 6.6.9 (Degree-Corrected Extension of 28 in [17]). *Assume that the spin σ_o of the root is drawn uniformly from $\{+, -\}$. Then for any $\ell \geq 0$,*

$$\mathbb{E} [Q_{1,\ell} Q_{2,\ell} | \mathcal{T}] = 0.$$

6.7 Coupling of local neighbourhood

The proofs of the statements in this section are deferred to Section 6.13.

6.7.1 Coupling

We distinguish between two different concepts of neighbourhood: the classical neighbourhood (see Section 5.7) that is rooted at a vertex and another neighbourhood that starts with an edge. For the latter, we need the following concept of *oriented* distance \vec{d} , which for $e, f \in \vec{E}(V)$ is defined as

$$\vec{d}(e, f) = \min_{\gamma} \ell(\gamma)$$

where the minimum is taken over all self-avoiding paths $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_{\ell+1})$ in G such that $(\gamma_0, \gamma_1) = e$, $(\gamma_\ell, \gamma_{\ell+1}) = f$ and for all $1 \leq k \leq \ell + 1$, $\{\gamma_k, \gamma_{k+1}\} \in \vec{E}$. and where for such a path γ , $\ell(\gamma) = \ell$. Note that $\vec{d}(e, f) = \vec{d}(f^{-1}, e^{-1})$, i.e., \vec{d} is not symmetric.

We introduce the vector $Y_t(e) = (Y_t(e)(i))_{i \in \{+, -\}}$ where, for $i \in \{+, -\}$,

$$Y_t(e)(i) = \left| \left\{ f \in \vec{E} : \vec{d}(e, f) = t, \sigma(f_2) = i \right\} \right|, \quad (6.35)$$

we denote the number of vertices at oriented distance t from e by

$$S_t(e) = \|Y_t(e)\|_1 = \left| \left\{ f \in \vec{E} : \vec{d}(e, f) = t \right\} \right|,$$

and we define vector $\Psi_t(e) = (\Psi_t(e)(i))_{i \in \{+, -\}}$ where, for $i \in \{+, -\}$,

$$\Psi_t(e)(i) = \sum_{f \in \vec{E}: \vec{d}(e, f) = t} 1_{\sigma(f_2) = i} \phi_{f_2}. \quad (6.36)$$

We denote the classical neighbourhood of radius r rooted at vertex v by $(G, v)_r$ and the neighbourhood around oriented edge $e = (e_1, e_2)$ by $(G, e)_r$. With the definitions above, we then have, $(G, e)_r = (G', e_2)_r$, where G' is the graph G with edge $\{e_1, e_2\}$ removed. In particular,

$$S_t(e) = S'_t(e_2),$$

where S'_t is S_t defined on G' .

The two branching processes that describe the neighbourhoods are almost identical, the only difference lies in the weight of the root: In the classical branching processes, the weight is drawn according to distribution ν . In the branching process starting at an edge oriented towards, say, o , the root o has weight governed by ν^* . See Proposition 6.7.1 below, which differs slightly from Theorem 5.7.1, because here the spins are deterministic according to (6.6).

As a corollary we obtain an analogue of Theorem 6.6.5 for local neighbourhoods: the components of $\Psi_t(e)$ grow exponentially, see Corollary 6.7.3.

We bound the growth of S_t in Lemma 6.7.4. We use a coupling argument to show that the weights of the unexplored vertices and selected vertices are stochastically dominated by variables following law ν , respectively ν^* . This argument is not needed in the ordinary SBM.

Following [94], we need to verify that certain problematic structures, namely *tangles*, are excluded with high probability. We say that a graph H is tangle-free if all its ℓ -neighbourhoods contain at most one cycle. If there is at least one ℓ -neighbourhood in H that contains more than one cycle, we call H tangled. Note that in the sequel we shall often suppress the dependence on ℓ and simply call a graph tangle-free or tangled; the ℓ dependence is then tacitly assumed.

Following standard arguments we establish in Lemma 6.7.5 that the graph is with high probability $\log(n)$ -tangle free.

Proposition 6.7.1 (Degree-Corrected Extension of Proposition 31 in [17]). *Let $\ell = C \log_\rho(n)$, with $C < C_{\text{coupling}}$. Let $\rho \in V$ and $e = (e_1, e_2) \in \vec{E}$. Let (T, o) be the branching process with root o defined in Section 6.6, where the root has spin $\sigma(o)$ and weight governed by ν . Similarly, Let (T', o) be that same branching process, when the root has spin $\sigma(e_2)$ and weight governed by ν^* . Then, the total variation distance between the law of $(G, v)_\ell$ and $(T, o)_\ell$ goes to zero as $1 - n^{-(\frac{\gamma}{2} \wedge \frac{1}{40})}$. The same is true for the difference between the law of $(G, e)_\ell$ and (T', o) .*

Remark 6.7.2. *Note that with the event $(G, v)_\ell = (T, o)_\ell$, we mean that the graph and tree are equal, **including their spins and weights**. See Theorem 5.7.1 for more details.*

Corollary 6.7.3 (Degree-Corrected Extension of Corollary 32 in [17]). *Assume $\mu_2^2 > \rho$. Let $\ell = C \log_\rho n$ with $0 < C < C_{\text{coupling}}$. For $e \in \vec{E}(V)$, we define the event $\mathcal{E}(e)$ that for all $0 \leq t < \ell$ and $k \in \{1, 2\}$: $|\langle g_k, \Psi_t(e) \rangle - \mu_k^{t-\ell} \langle g_k, \Psi_\ell(e) \rangle| \leq (\log n)^3 \rho^{t/2}$. Then, with high probability, the number of edges $e \in \vec{E}$ such that $\mathcal{E}(e)$ does not hold is at most $\log(n) n^{1 - (\frac{\gamma}{2} \wedge \frac{1}{40})}$.*

Lemma 6.7.4 (Degree-Corrected Extension of Lemma 29 in [17]). *There exist $c, c' > 0$ such that for all $s \geq 0$ and for any $w \in [n] \cup \vec{E}(V)$,*

$$\mathbb{P}(\forall t \geq 0 : S_t(w) \leq s \rho_n^t) \geq 1 - ce^{-c's}.$$

Consequently, for any $p \geq 1$, there exists $c'' > 0$ such that

$$\mathbb{E} \left[\max_{v \in [n], t \geq 0} \left(\frac{S_t(v)}{\bar{\rho}_n^t} \right)^p \right] \leq c'' (\log n)^p.$$

Lemma 6.7.5 (Degree-Corrected Extension of Lemma 30 in [17]). *Let $\ell = C \log_\rho(n)$, with $0 < C < C_{\text{coupling}}$. Then, w.h.p., at most $\rho^{2\ell} \log(n)$ vertices have a cycle in their ℓ -neighbourhood. Further, w.h.p., the graph is ℓ -tangle-free.*

6.7.2 Geometric growth

Here we show that for $k \in \{1, 2\}$, $\langle B^\ell \chi_k, \delta_e \rangle$ grows nearly geometrically in t with rate μ_k . Corollary 6.7.7 then establishes a bound for $r \leq \ell$ on $\sup_{\langle B^r \chi_k, x \rangle = 0, \|x\|=1} \|\langle B^r \chi_k, x \rangle\|$ crucial for the norm bounds in Section 6.10.

Proposition 6.7.6 (Degree-Corrected Extension of Proposition 33 in [17]). *Assume $\mu_2^2 > \rho$. Let $\ell = C \log_\rho(n)$, with $0 < C < C_{\text{coupling}} \wedge \left(\frac{1}{2} - \left(\frac{\gamma}{4} \wedge \frac{1}{80}\right)\right) = C_{\text{coupling}}$. For $e \in \vec{E}(V)$, let \vec{E}_ℓ be the set of oriented edges such that either $(G, e_2)_\ell$ is not a tree or the event $\mathcal{E}(e)$ (defined in Corollary 6.7.3) does not hold. Then, w.h.p. for $k \in \{1, 2\}$:*

$$(i) |\vec{E}_\ell| \ll (\log n)^2 n^{1 - \frac{\gamma}{2} \wedge \frac{1}{40}},$$

$$(ii) \text{ for all } e \in \vec{E} \setminus \vec{E}_\ell, 0 \leq r \leq \ell,$$

$$|\langle B^r \chi_k, \delta_e \rangle - \mu_k^{r-\ell} \langle B^\ell \chi_k, \delta_e \rangle| \leq (\log n)^4 \rho^{r/2},$$

$$(iii) \text{ for all } e \in \vec{E}_\ell, 0 \leq r \leq \ell,$$

$$|\langle B^r \chi_k, \delta_e \rangle| \leq (\log n)^2 \rho^r.$$

Corollary 6.7.7 (Degree-Corrected Extension of Corollary 34 in [17]). *Let $\ell = C \log_\rho(n)$, with $0 < C < C_{\text{coupling}} \wedge \left(1 - \frac{\gamma}{2} \wedge \frac{1}{40}\right) \wedge \left(\frac{\gamma}{4} \wedge \frac{1}{80}\right) = C_{\text{coupling}}$. W.h.p. for any $0 \leq r \leq \ell - 1$ and $k \in \{1, 2\}$:*

$$\sup_{\langle B^r \chi_k, x \rangle = 0, \|x\|=1} \|\langle B^r \chi_k, x \rangle\| \leq (\log n)^5 n^{1/2} \rho^{r/2}.$$

6.8 A weak law of large numbers for local functionals on the DC-SBM

The proofs of the statements in this section are deferred to Section 6.14.

Here we show that a weak law of large numbers applies for local functionals defined on *weighted coloured* random graphs generated according to the DC-SBM.

By a *weighted coloured* graph we mean a graph $G = (V, E)$ together with maps $\sigma : V \rightarrow \{+, -\}$ and $\phi : V \rightarrow [\phi_{\min}, \phi_{\max}]$. For $v \in V$, we identify $\sigma(v)$ as the spin of v and $\phi(v)$ as its weight. We denote by \mathcal{G}^* the set of *rooted weighted coloured* graphs. We denote an element of \mathcal{G}^* by (G, o) : $G = (V, E)$ is then a weighted coloured graph and $o \in V$ is some distinguished vertex. A function $\tau : \mathcal{G}^* \rightarrow \mathbb{R}$ is said to be ℓ -local if $\tau(G, o)$ depends only on $(G, o)_\ell$.

To derive the claimed weak law when G is drawn according to the DC-SBM, we prepare with a variance bound for $\sum_{v=1}^n \tau(G, v)$, see Proposition 6.8.1. The bound

follows from the law of total variance,

$$\begin{aligned} \text{Var} \left(\sum_{v=1}^n \tau(G, v) \right) &= \mathbb{E} \left[\text{Var} \left(\sum_{v=1}^n \tau(G, v) \middle| \phi_1, \dots, \phi_n \right) \right] \\ &\quad + \text{Var} \left(\mathbb{E} \left[\sum_{v=1}^n \tau(G, v) \middle| \phi_1, \dots, \phi_n \right] \right), \end{aligned}$$

together with an application of Efron-Stein's inequality to both terms on the right. For instance, Efron-Stein's inequality says that the variance of $\mathbb{E} [\sum_{v=1}^n \tau(G, v) | \phi_1, \dots, \phi_n]$ can be bounded by considering the variance caused by one weight, while keeping all other weights fixed. If τ is ℓ -local, and all weights except at vertex u are fixed, then the only random part of $\sum_{v=1}^n \tau(G, v)$ is $\sum_{v \in \mathcal{N}(u)} \tau(G, v) \leq |\mathcal{N}(u)| \max_v \tau(G, v)$, where $\mathcal{N}(u)$ is the ℓ neighbourhood of u . For bounded functions τ we thus expect the variance to be $\mathcal{O}(n^\beta)$, with $\beta < 2$.

The sample average $\frac{1}{n} \sum_{v=1}^n \tau(G, v)$ concentrates then around $\mathbb{E} [\tau(T, o)]$, where (T, o) is the branching process from Section 6.6, with root o having spin drawn uniformly from $\{+, -\}$ and weight governed by ν , see Proposition 6.8.2. The coupling, and in particular the matching of the weights, plays an important role in its proof.

In the next section we apply the latter proposition to some specific functionals.

Proposition 6.8.1 (Degree-Corrected Extension of Proposition 35 in [17]). *Let G be drawn according to the DC-SBM. There exists $c > 0$ such that if $\tau, \varphi : \mathcal{G}^* \rightarrow \mathbb{R}$ are ℓ -local, $|\tau(G, o)| \leq \varphi(G, o)$ and φ is non-decreasing by the addition of edges, then*

$$\text{Var} \left(\sum_{v=1}^n \tau(G, v) \right) \leq cn\rho^{2\ell} \left(\mathbb{E} \left[\max_{v \in [n]} \varphi^4(G, v) \right] \right)^{1/2}.$$

Proposition 6.8.2 (Degree-Corrected Extension of Proposition 36 in [17]). *Let G be drawn according to the DC-SBM. Let (T, o) be the branching process from Section 6.6, with root o having spin drawn uniformly from $\{+, -\}$ and weight governed by ν . Let $\ell = C \log_\rho(n)$, with $C < C_{\text{coupling}}$. There exists $c > 0$ such that if $\tau, \varphi : \mathcal{G}^* \rightarrow \mathbb{R}$ are ℓ -local, $|\tau(G, o)| \leq \varphi(G, o)$ and φ is non-decreasing by the addition of edges, then*

$$\begin{aligned} &\mathbb{E} \left[\left| \frac{1}{n} \sum_{v=1}^n \tau(G, v) - \mathbb{E} [\tau(T, o)] \right| \right] \\ &\leq c_2 n^{-(\frac{\gamma}{2} \wedge \frac{1}{40})} \left(\mathbb{E} \left[\max_{v \in [n]} \varphi^4(G, v) \right]^{1/4} \vee \mathbb{E} [\varphi^2(T, o)]^{1/2} \right) + \mathcal{O}(n^{-\gamma}) \end{aligned} \tag{6.37}$$

6.8.1 Application with some specific local functionals

Here we consider $\langle B^\ell \chi_1, B^\ell \chi_2 \rangle$, $\langle B^{2\ell} \chi_k, B^\ell \chi_j \rangle$, and $\langle B^\ell B^{*\ell} \chi_1, B^\ell B^{*\ell} \chi_2 \rangle$, quantities occurring in Proposition 6.5.1.

Explicitly, $B^\ell \chi_k(e) = \sum_f B_{ef}^\ell g_k(\sigma(f_2)) \phi_{f_2}$, where we recall that B_{ef}^ℓ is the number of non-backtracking walks from e to f . Now, if the oriented ℓ -neighbourhood of e is a *tree*, then $B^\ell \chi_k(e) = \langle g_k, \Psi_\ell(e) \rangle$. With this intuition in mind, we analyse likewise expressions in Proposition 6.8.3 below.

Inspired by (6.32), which expresses $B^\ell B^{*\ell} \chi_k$ on *trees* in terms of the operator $Q_{k,\ell}$, we extend the latter to an operator defined on general graphs. First, for $e \in \vec{E}(V)$ and $t \geq 0$, set $\mathcal{Y}_t(e) = \{f \in \vec{E} : \vec{d}(e, f) = t\}$. Then, for $k \in \{1, 2\}$, we set

$$P_{k,\ell}(e) = \sum_{t=0}^{\ell-1} \sum_{f \in \mathcal{Y}_t(e)} L_k(f), \tag{6.38}$$

with

$$L_k(f) = \sum_{(g,h) \in \mathcal{Y}_1(f) \setminus \mathcal{Y}_t(e); g \neq h} \langle g_k, \tilde{\Psi}_t(g) \rangle \tilde{S}_{\ell-t-1}(h),$$

where $\tilde{\Psi}_t(g)$, $\tilde{S}_{\ell-t-1}(h) = \|\tilde{Y}_{\ell-t-1}(h)\|_1$ are the variables $\Psi_t(g)$, respectively $S_{\ell-t-1}(h)$, defined on the graph G where all edges in $(G, e_2)_t$ have been removed. Note that, if $(G, e)_{2\ell}$ is a tree, then $\tilde{\Psi}_s(g) = \Psi_s(g)$ for $s \leq 2\ell - t$. Compare $P_{k,\ell}$ to $Q_{k,\ell}$ in (6.29) and $L_k(f)$ to $L_{k,\ell}$ in (6.34).

Finally, define

$$S_{k,\ell}(e) = S_\ell(e) g_k(\sigma(e_1)) \phi_{e_1}. \quad (6.39)$$

We then have an extension of (6.32), when $(G, e_2)_{2\ell}$ is a tree:

$$B^\ell B^{*\ell} \tilde{\chi}_k(e) = P_{k,\ell}(e) + S_{k,\ell}(e). \quad (6.40)$$

We analyse (6.40) in Proposition 6.8.4 below.

We start with the case $\mu_2^2 > \rho$:

Proposition 6.8.3 (Degree-Corrected Extension of Proposition 37 in [17]). *Assume that $\mu_2^2 > \rho$. Let $\ell = C \log_\rho n$ with $0 < C < C_{\text{coupling}}$.*

(i) *For any $k \in \{1, 2\}$, there exists $c'_k > 0$ such that, in probability,*

$$\frac{1}{n} \sum_{e \in \tilde{E}} \frac{\langle g_k, \Psi_\ell(e) \rangle^2}{\mu_k^{2\ell}} \rightarrow c'_k.$$

(ii) *For any $k \in \{1, 2\}$, there exists $c''_k > 0$ such that, in probability,*

$$\frac{1}{n} \sum_{e \in \tilde{E}} \frac{\langle g_k, Y_\ell(e) \rangle^2}{\mu_k^{2\ell}} \rightarrow c''_k.$$

(iii)

$$\mathbb{E} \left[\left| \frac{1}{n} \sum_{e \in \tilde{E}} \langle g_1, \Psi_\ell(e) \rangle \langle g_2, \Psi_\ell(e) \rangle \right| \right] \leq (\log n)^3 n^{2C - (\frac{\gamma}{2} \wedge \frac{1}{40})} + n^{-\gamma}.$$

(iv) *For any $k \neq j \in \{1, 2\}$,*

$$\mathbb{E} \left[\left| \frac{1}{n} \sum_{e \in \tilde{E}} \langle g_k, \Psi_{2\ell}(e) \rangle \langle g_j, \Psi_\ell(e) \rangle \right| \right] \leq (\log n)^3 n^{3C - (\frac{\gamma}{2} \wedge \frac{1}{40})} + n^{-\gamma}.$$

(v) *For any $k \in \{1, 2\}$, in probability*

$$\frac{1}{n} \sum_{e \in \tilde{E}} \frac{\langle g_k, \Psi_{2\ell}(e) \rangle \langle g_k, \Psi_\ell(e) \rangle}{\mu_k^{3\ell}} \rightarrow c'''_k.$$

Proposition 6.8.4 (Degree-Corrected Extension of Proposition 38 in [17]). *Assume that $\mu_2^2 > \rho$. Let $\ell = C \log_\rho n$ with $C < C_{\text{coupling}}$.*

(i) *For any $k \in \{1, 2\}$, there exists $c''''_k > 0$ such that in probability*

$$\frac{1}{n} \sum_{e \in \tilde{E}} \frac{P_{k,\ell}^2(e)}{\mu_k^{4\ell}} \rightarrow c''''_k.$$

(ii)

$$\mathbb{E} \left[\left| \frac{1}{n} \sum_{e \in \vec{E}} (P_{1,\ell}(e) + S_{1,\ell}(e))(P_{2,\ell}(e) + S_{2,\ell}(e)) \right| \right] \leq (\log n)^8 n^{4C - (\frac{7}{2} \wedge \frac{1}{40})}$$

In case $\mu_2^2 \leq \rho$, most of the above claims continue to hold. We treat the exceptions here:

Proposition 6.8.5. *Assume that $\mu_2^2 \leq \rho$. Let $\ell = C \log_\rho n$ with $0 < C < C_{\text{coupling}}$. There exists some $c > 0$, such that w.h.p.,*

$$\frac{1}{n} \sum_{e \in \vec{E}} \frac{\langle g_2, \Psi_\ell(e) \rangle^2}{\rho^\ell} \geq c.$$

Proposition 6.8.6. *Assume that $\mu_2^2 \leq \rho$. Let $\ell = C \log_\rho n$ with $C < C_{\text{coupling}}$. There exists $c > 0$ such that w.h.p.,*

$$\frac{1}{n} \sum_{e \in \vec{E}} \frac{P_{2,\ell}^2(e)}{\rho^{2\ell} \log^5(n)} \leq c.$$

6.9 Proof of Propositions 6.5.1 and 6.5.3

We introduce for $k \in \{1, 2\}$ the vector $N_{k,\ell}$, defined on $e \in \vec{E}$ as

$$N_{k,\ell}(e) = \langle g_k, \Psi_\ell(e) \rangle.$$

If $(G, e_2)_\ell$ is a tree, then

$$N_{k,\ell}(e) = \langle B^\ell \chi_k, \delta_e \rangle,$$

and we have a similar expression for $B^\ell B^{*\ell} \check{\chi}_k$ in (6.40). Now, at most $\rho^{2\ell} \log(n)$ vertices have a cycle in their ℓ -neighbourhood (see Lemma 6.7.5). Therefore:

Lemma 6.9.1 (Degree-Corrected Extension of Lemma 39 in [17]). *Let $\ell = C \log_\rho n$ with $0 < C < C_{\text{min}}$. Then, w.h.p. $\|B^\ell \chi_k - N_{k,\ell}\| = O((\log n)^{5/2} \rho^{2\ell}) = o(\rho^{\ell/2} \sqrt{n})$, $\|B^\ell B^{*\ell} \check{\chi}_k - P_{k,\ell} - S_{k,\ell}\| = O((\log n)^4 \rho^{4\ell})$ and $\|B^\ell B^{*\ell} \check{\chi}_k - P_{k,\ell}\| = O(\rho^\ell \sqrt{n})$.*

Proof. The proof of Lemma 39 in [17] can be easily adapted to the current setting. The key idea is pointed out above. It thus remains to bound $|(B^\ell \chi_k - N_{k,\ell})(e)|$ and $|(B^\ell B^{*\ell} \check{\chi}_k - P_{k,\ell})(e)|$ on edges e for which $(G, e_2)_\ell$ is not a tree. For this, use that with high probability the graph is 2ℓ -tangle free so that there are at most two non-backtracking paths between e and any edge at distance ℓ . \square

We can thus in our calculations replace $B^\ell \chi_k$ by $N_{k,\ell}$ and $B^\ell B^{*\ell} \check{\chi}_k$ by $P_{k,\ell}$. From Propositions 6.8.3 and 6.8.4, Proposition 6.5.1 then follows:

Proof of Proposition 6.5.1. This proof follows the corresponding proof in [17]. We give the key observations: (i) From Proposition 6.8.3 (i), $\|N_{k,\ell}\| \sim \sqrt{n} \mu_k^\ell$ and from Proposition 6.8.4 (i), $\|P_{k,\ell}\| \sim \sqrt{n} \mu_k^{2\ell}$.

(ii) From Proposition 6.8.3 (v), $|\langle N_{k,\ell}, N_{k,2\ell} \rangle| \sim n \mu_k^{3\ell}$.

(iii) From Proposition 6.8.3 (iii), $|\langle N_{1,\ell}, N_{2,\ell} \rangle| \sim (\log n)^3 n^{3C - (\frac{7}{2} \wedge \frac{1}{40})}$.

(iv) From Proposition 6.8.3 (iv), $|\langle N_{k,2\ell}, N_{j,\ell} \rangle| \sim (\log n)^3 n^{4C - (\frac{7}{2} \wedge \frac{1}{40})}$.

(v) From Proposition 6.8.4 (ii), $|\langle P_{1,\ell} + S_{1,\ell}, P_{2,\ell} + S_{2,\ell} \rangle| \sim (\log n)^8 n^{5C - (\frac{7}{2} \wedge \frac{1}{40})}$. \square

Proposition 6.5.3 follows similarly from the case $\mu_2^2 \leq \rho$ treated in Section 6.8.1:

Proof of Proposition 6.5.3. This follows from Propositions 6.8.5 and 6.8.6 in conjunction with Lemma 6.9.1. \square

6.10 Norm of non-backtracking matrices

The proofs of the statements in this section are deferred to Section 6.15.

In this section the product over an empty set is defined to be one.

It is convenient to extend matrix B and vector χ_k to the set of directed edges on the *complete* graph, $\vec{E}_K(V) = \{(u, v) : u \neq v \in V\}$: For $e, f \in \vec{E}_K(V)$, B_{ef} is then extended to

$$B_{ef} = A_e A_f 1_{e_2=f_1} 1_{e_1 \neq f_2}, \quad (6.41)$$

where A is the adjacency matrix. For each $e \in \vec{E}_K(V)$ we set $\chi_k(e) = g_k(\sigma(e_2))\phi_{e_2}$.

For integer $k \geq 1$, $e, f \in \vec{E}_K(V)$, we let Γ_{ef}^k be the set of non-backtracking walks $\gamma = (\gamma_0, \dots, \gamma_k)$ of length k from $(\gamma_0, \gamma_1) = e$ to $(\gamma_{k-1}, \gamma_k) = f$ on the *complete* graph with vertex set V .

By induction it follows that

$$(B^k)_{ef} = \sum_{\gamma \in \Gamma_{ef}^{k+1}} \prod_{s=0}^k A_{\gamma_s \gamma_{s+1}}. \quad (6.42)$$

Indeed, note that $\prod_{s=0}^k A_{\gamma_s \gamma_{s+1}}$ is one when γ is a path in G and zero otherwise.

To each walk $\gamma = (\gamma_0, \dots, \gamma_k)$, we associate the graph $G(\gamma) = (V(\gamma), E(\gamma))$, with the set of vertices $V(\gamma) = \{\gamma_i, 0 \leq i \leq k\}$ and the set of edges $E(\gamma) = \{\{\gamma_i, \gamma_{i+1}\}, 0 \leq i \leq k-1\}$.

From Lemma 6.7.5, the graphs following the DC-SBM are tangle-free with high probability. Hence, it makes sense to consider the subset $F_{ef}^{k+1} \subset \Gamma_{ef}^{k+1}$ of tangle-free non-backtracking walks on the *complete* graph. Indeed, if G is tangle-free, we need only consider the tangle-free paths in the summation (6.42):

$$(B^k)_{ef} = \sum_{\gamma \in F_{ef}^{k+1}} \prod_{s=0}^k A_{\gamma_s \gamma_{s+1}}, \quad (6.43)$$

and $B^k = B^{(k)}$ for $1 \leq k \leq \ell$.

Define for $u \neq v$ the *centred* random variable

$$\underline{A}_{uv} = A_{uv} - \frac{\phi_u \phi_v}{n} W_{\sigma_u \sigma_v}, \quad (6.44)$$

where

$$W = \begin{pmatrix} a & b \\ b & a \end{pmatrix}.$$

Compare this to the SBM *without* degree-corrections in Section 10.1 of [17]: $\phi_u = 1$ for all u in the latter model.

Using \underline{A} we shall attempt to center B^k when the underlying graph G is tangle-free through considering

$$\Delta_{ef}^{(k)} = \sum_{\gamma \in F_{ef}^{k+1}} \prod_{s=0}^k \underline{A}_{\gamma_s \gamma_{s+1}}. \quad (6.45)$$

Further, we set

$$\Delta_{ef}^{(0)} = 1_{e=f} \underline{A}_e \quad \text{and} \quad B_{ef}^{(0)} = 1_{e=f} A_e. \quad (6.46)$$

To decompose (6.43), following a decomposition that appeared first in [85], we use

$$\prod_{s=0}^{\ell} x_s = \prod_{s=0}^{\ell} y_s + \sum_{t=0}^{\ell} \prod_{s=0}^{t-1} y_s (x_t - y_t) \prod_{s=t+1}^{\ell} x_s,$$

with $x_s = A_{\gamma_s \gamma_{s+1}}$ and $y_s = \underline{A}_{\gamma_s \gamma_{s+1}}$ on a path $\gamma \in F_{ef}^{k+1}$:

$$\prod_{s=0}^{\ell} A_{\gamma_s \gamma_{s+1}} = \prod_{s=0}^{\ell} \underline{A}_{\gamma_s \gamma_{s+1}} + \sum_{t=0}^{\ell} \prod_{s=0}^{t-1} \underline{A}_{\gamma_s \gamma_{s+1}} \left(\frac{\phi_{\gamma_t} \phi_{\gamma_{t+1}}}{n} W_{\sigma_{\gamma_t} \sigma_{\gamma_{t+1}}} \right) \prod_{s=t+1}^{\ell} A_{\gamma_s \gamma_{s+1}}.$$

Summing over all $\gamma \in F_{ef}^{\ell+1}$ then gives

$$\begin{aligned} B_{ef}^{(\ell)} &= \sum_{\gamma \in F_{ef}^{\ell+1}} \prod_{s=0}^{\ell} \underline{A}_{\gamma_s \gamma_{s+1}} \\ &\quad + \sum_{t=0}^{\ell} \sum_{\gamma \in F_{ef}^{\ell+1}} \prod_{s=0}^{t-1} \underline{A}_{\gamma_s \gamma_{s+1}} \left(\frac{\phi_{\gamma_t} \phi_{\gamma_{t+1}}}{n} W_{\sigma_{\gamma_t} \sigma_{\gamma_{t+1}}} \right) \prod_{s=t+1}^{\ell} A_{\gamma_s \gamma_{s+1}} \\ &= \Delta_{ef}^{(\ell)} + \sum_{t=0}^{\ell} \sum_{\gamma \in F_{ef}^{\ell+1}} \prod_{s=0}^{t-1} \underline{A}_{\gamma_s \gamma_{s+1}} \left(\frac{\phi_{\gamma_t} \phi_{\gamma_{t+1}}}{n} W_{\sigma_{\gamma_t} \sigma_{\gamma_{t+1}}} \right) \prod_{s=t+1}^{\ell} A_{\gamma_s \gamma_{s+1}}. \end{aligned} \quad (6.47)$$

Consider the two products in the summation over $F_{ef}^{\ell+1}$ on the right of (6.47): We can, for $1 \leq t \leq \ell - 1$, replace the summation over $F_{ef}^{\ell+1}$ by summing over all pairs $\gamma' = (\gamma_0, \dots, \gamma_t) \in F_{eg}^t$ and $\gamma'' = (\gamma_{t+1}, \dots, \gamma_{\ell+1}) \in F_{g'f}^{\ell-t}$ for some $g, g' \in \vec{E}(V)$ such that there exists a non-backtracking path with one intermediate edge, on the *complete* graph, between oriented edges g and g' (we denote this property by $g \xrightarrow{2} g'$). However caution is needed, as this summation also includes *tangled* paths, namely those in the sets $\{F_{t,ef}^{\ell+1}\}_{t=0}^{\ell}$. Where, for $1 \leq t \leq \ell - 1$, $F_{t,ef}^{\ell+1}$ is defined as the collection of all *tangled* paths $\gamma = (\gamma_0, \dots, \gamma_{\ell+1}) = (\gamma', \gamma'') \in \Gamma_{ef}^{\ell+1}$ with γ' and γ'' as above. For $t = 0$, $F_{0,ef}^{\ell+1}$ consists of all non-backtracking *tangled* paths (γ', γ'') with $\gamma' = (e_1)$ and $\gamma'' \in F_{g'f}^{\ell}$ for any g' such that $g'_1 = e_2$. For $t = \ell$, $F_{\ell,ef}^{\ell+1}$ is the set of non-backtracking *tangled* paths (γ', γ'') such that $\gamma'' = (f_2)$ and $\gamma' \in F_{eg}^{\ell}$ for some $g \in \vec{E}(V)$ with $g_2 = f_1$. We rewrite (6.47) as

$$B^{(\ell)} = \Delta^{(\ell)} + \frac{1}{n} K B^{(\ell-1)} + \frac{1}{n} \sum_{t=1}^{\ell-1} \Delta^{(t-1)} K^{(2)} B^{(\ell-t-1)} + \frac{1}{n} \Delta^{(\ell-1)} \widehat{K} - \frac{1}{n} \sum_{t=0}^{\ell} R_t^{(\ell)}, \quad (6.48)$$

where for $e, f \in E_K$,

$$K_{ef} = 1_{e \rightarrow f} \phi_{e_1} \phi_{e_2} W_{\sigma_{(e_1)} \sigma_{(e_2)}}, \quad (6.49)$$

the *weighted* non-backtracking matrix on the *complete* graph (recall that $e \rightarrow f$ represents the non-backtracking property),

$$\widehat{K}_{ef} = 1_{e \rightarrow f} \phi_{f_1} \phi_{f_2} W_{\sigma_{(f_1)} \sigma_{(f_2)}}, \quad (6.50)$$

$$K_{ef}^{(2)} = 1_{e \xrightarrow{2} f} \phi_{e_2} \phi_{f_1} W_{\sigma(e_2)\sigma(f_1)}, \quad (6.51)$$

and where

$$(R_t^{(\ell)})_{ef} = \sum_{\gamma \in F_{t,ef}^{\ell+1}} \prod_{s=0}^{t-1} A_{\gamma_s \gamma_{s+1}} \phi_{\gamma_t} \phi_{\gamma_{t+1}} W_{\sigma(\gamma_t)\sigma(\gamma_{t+1})} \prod_{s=t+1}^{\ell} A_{\gamma_s \gamma_{s+1}}. \quad (6.52)$$

Indeed,

$$\begin{aligned} \left(\sum_{t=1}^{\ell-1} \Delta^{(t-1)} K^{(2)} B^{(\ell-t-1)} \right)_{ef} &= \sum_{t=1}^{\ell-1} \sum_{g, g'} \Delta_{eg}^{(t-1)} K_{gg'}^{(2)} B_{g'f}^{(\ell-t-1)} \\ &= \sum_{t=1}^{\ell-1} \sum_{g, g'} \sum_{\gamma' \in F_{eg}^t} \sum_{\gamma'' \in F_{g'f}^{\ell-t}} \prod_{s=0}^{t-1} A_{\gamma'_s \gamma'_{s+1}} 1_{g \xrightarrow{2} g'} \phi_{\gamma'_t} \phi_{\gamma''_0} \\ &\quad \cdot W_{\sigma(\gamma'_t)\sigma(\gamma''_0)} \prod_{s=0}^{\ell-t-1} A_{\gamma''_s \gamma''_{s+1}}, \end{aligned} \quad (6.53)$$

$$\left(KB^{(\ell-1)} \right)_{ef} = \sum_g \sum_{\gamma'' \in F_{gf}^\ell} 1_{e \rightarrow g} \phi_{e_1} \phi_{e_2} W_{\sigma(e_1)\sigma(e_2)} A_{e_2, g_2} \prod_{s=1}^{\ell-2} A_{\gamma''_s \gamma''_{s+1}} A_{f_1, f_2}, \quad (6.54)$$

and,

$$\left(\Delta^{(\ell-1)} \widehat{K} \right)_{ef} = \sum_g \sum_{\gamma' \in F_{eg}^\ell} A_{e_1, e_2} \prod_{s=1}^{\ell-2} A_{\gamma'_s \gamma'_{s+1}} A_{g_1, f_1} 1_{g \rightarrow f} \phi_{f_1} \phi_{f_2} W_{\sigma(f_1)\sigma(f_2)} \quad (6.55)$$

that is exactly the splitting described just below (6.47), where we also pointed out the need to compensate for *tangled* paths occurring in (6.53), which is precisely the role of $R_t^{(\ell)}$ in (6.48).

To bound (6.48), we introduce

$$\overline{W} = \frac{2}{\Phi^{(2)}} (\rho \chi_1 \check{\chi}_1^* + \mu_2 \chi_2 \check{\chi}_2^*) = (\phi_{e_2} \phi_{f_1} W_{\sigma(e_2)\sigma(f_1)})_{ef}, \quad (6.56)$$

and,

$$L = K^{(2)} - \overline{W}. \quad (6.57)$$

Note the presence of weights in (6.56), hence our choice for the candidate eigenvectors.

Further, we set for $1 \leq t \leq \ell - 1$,

$$S_t^{(\ell)} = \Delta^{(t-1)} L B^{(\ell-t-1)}. \quad (6.58)$$

We then have:

Proposition 6.10.1 (Degree-Corrected Extension of Proposition 13 in [17]). *If G is tangle-free and $x \in \mathbb{C}^{\vec{E}(V)}$ with norm smaller than one, we have*

$$\begin{aligned} \|B^\ell x\| &\leq \|\Delta^{(\ell)}\| + \frac{1}{n} \|KB^{(\ell-1)}\| + \frac{1}{n} \sum_{j=1,2} \frac{2\mu_j}{\Phi^{(2)}} \sum_{t=1}^{\ell-1} \|\Delta^{(t-1)} \chi_j\| \|\langle \check{\chi}_j, B^{\ell-t-1} x \rangle\| \\ &\quad + \frac{1}{n} \sum_{t=1}^{\ell-1} \|S_t^{(\ell)}\| + \phi_{\max}^2(a \vee b) \|\Delta^{(\ell-1)}\| + \frac{1}{n} \sum_{t=0}^{\ell} \|R_t^{(\ell)}\|. \end{aligned}$$

Proof. Due to the tangle-freeness, $B^\ell = B^{(\ell)}$. Further $K^{(2)} = L + \overline{W}$ and $\|K\| \leq \phi_{\max}^2(a \vee b)n$. \square

In Section 6.15 we prove the following bounds on the matrices in Proposition 6.10.1:

Proposition 6.10.2 (Degree-Corrected Extension of Proposition 14 in [17]). *Let $\ell = C \log_\rho n$ with $C < 1$. With high probability, the following norm bounds hold for all k , $0 \leq k \leq \ell$, and $i = 1, 2$:*

$$\|\Delta^{(k)}\| \leq (\log n)^{10} \rho^{k/2}, \quad (6.59)$$

$$\|\Delta^{(k)} \chi_i\| \leq (\log n)^5 \rho^{k/2} \sqrt{n}, \quad (6.60)$$

$$\|R_k^{(\ell)}\| \leq (\log n)^{25} \rho^{\ell-k/2}, \quad (6.61)$$

$$\|KB^{(k)}\| \leq \sqrt{n} (\log n)^{10} \rho^k, \quad (6.62)$$

and the following bound holds for all k , $1 \leq k \leq \ell - 1$:

$$\|S_k^{(\ell)}\| \leq \sqrt{n} (\log n)^{20} \rho^{\ell-k/2}. \quad (6.63)$$

6.10.1 Proof of Proposition 6.5.2

From Propositions 6.10.1 and 6.10.2, the geometric growth in Corollary 6.7.7 together with the tangle-freeness due to Lemma 6.7.5, the proof of Proposition 6.5.2 follows:

Let $j \in \{1, 2\}$. If, for some vector x , $\langle \check{\varphi}_j, x \rangle = 0$, then $\langle B^\ell \chi_j, \check{x} \rangle = 0$. Therefore, using Corollary 6.7.7,

$$\begin{aligned} \sup_{\|x\|=1, \langle \check{\varphi}_j, x \rangle = 0} \langle \check{\chi}_j, B^{\ell-t-1} x \rangle &= \sup_{\|x\|=1, \langle B^\ell \chi_j, \check{x} \rangle = 0} \langle B^{\ell-t-1} \chi_j, \check{x} \rangle \\ &= \sup_{\|\check{x}\|=1, \langle B^\ell \chi_j, \check{x} \rangle = 0} \langle B^{\ell-t-1} \chi_j, \check{x} \rangle \\ &\leq \log^2(n) n^{1/2} \rho^{\frac{\ell-t-1}{2}}. \end{aligned} \quad (6.64)$$

With high probability, the graph is ℓ -tangle free (Lemma 6.7.5). Thus, invoking Propositions 6.10.1 and 6.10.2, with high probability,

$$\begin{aligned} \sup_{x \in H^\perp, \|x\|=1} \|B^\ell x\| &\leq \log^{10}(n) \rho^{\frac{\ell}{2}} + n^{-1/2} \log^{10}(n) \rho^{\ell-1} \\ &\quad + c_1 \log^8(n) \rho^{\frac{\ell}{2}} + n^{-1/2} \log^{21}(n) \rho^\ell \\ &\quad + c_2 \log^{10}(n) \rho^{\frac{\ell}{2}} + n^{-1} \log^{26}(n) \rho^\ell \\ &\leq \log^c(n) \rho^{\frac{\ell}{2}}, \end{aligned} \quad (6.65)$$

since $C < 1$.

6.10.2 Comparison with the Stochastic Block Model in [17]

Putting $\phi_u = 1$ for all u , we retrieve exactly the same bounds as in the **Stochastic Block Model**, that is equations (30) – (34) in [17].

Below we use the trace method and therefore path counting combinatorial arguments to establish Proposition 6.10.2. In particular, we bound the expectation of expressions of the form

$$\mathbb{E} \left[\prod_{i=1}^{2m} \prod_{s=1}^k A_{\gamma_{i,s-1} \gamma_{i,s}} \right], \quad (6.66)$$

for certain paths $\gamma = (\gamma_1, \dots, \gamma_{2m})$ with $\gamma_i = (\gamma_{i,0}, \dots, \gamma_{i,k}) \in V^{k+1}$, where \underline{A} is defined in (6.44).

In bounding (6.66) the following term occurs:

$$\prod_{u \in V(\gamma)} \Phi^{(d_u)},$$

where $(d_u)_u$ are the degrees of the vertices in a specific tree (or forest) spanning the path γ . See, for instance, (6.130) and (6.143) below. Here lies a major complication with respect to the Stochastic Block Model: those terms are not present in the latter model. In (6.134) and (6.145) we find

$$\prod_{u=1}^{|V(\gamma)|} \Phi^{(d_u)} \leq C_2^{\sum_{u: d_u > 2} (d_u - 2)} \left(\Phi^{(2)} \right)^{|V(\gamma)| - n_c},$$

where $C_2 > 1$ is some constant and where $n_c \geq 1$ is the number of components on the path γ . To compare this term with powers of $\Phi^{(2)}$ (which are present in powers of $\rho = \frac{a+b}{2} \Phi^{(2)}$), we bound $\sum_{u: d_u > 2} (d_u - 2)$, see in particular Lemma 6.15.2 and Lemma 6.15.5 below.

6.11 Detection: Proof of Theorem 6.2.2

The proofs of the statements in this section are deferred to Section 6.16.

We need the following special case of a lemma in [17]:

Lemma 6.11.1 (Special case of Lemma 40 in [17]). *Assume that there exists a function $F : V \rightarrow \{0, 1\}$ such that in probability, for any $i \in \{+, -\}$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{v=1}^n 1_{\sigma(v)=i} F(v) = \frac{f(i)}{2},$$

where $f : \{+, -\} \rightarrow [0, 1]$ is such that $f(+)>f(-)$. Then, assigning to each vertex a label $\hat{\sigma}(v) = +$ if $F(v) = 1$ and $\hat{\sigma}(v) = -$ if $F(v) = 0$, yields asymptotically positive overlap with the true spins.

Recall the eigenvector ξ_2 from Theorem 6.2.1. Below we use the function $F : v \mapsto 1_{\sum_{e: e_2=v} \xi_2(e) > \frac{\tau}{\sqrt{n}}}$ or $F : v \mapsto 1_{\sum_{e: e_2=v} \xi_2(e) \leq \frac{\tau}{\sqrt{n}}}$ for some fixed parameter τ . We verify also that ξ_2 is aligned with $P_{2,\ell}$. It is therefore useful to introduce the vector I_ℓ , defined element-wise by

$$I_\ell(v) = \sum_{e \in \vec{E}: e_2=v} P_{2,\ell}(e), \quad (6.67)$$

for $v \in V$.

Further, put

$$\hat{c} = \frac{a+b}{2} \frac{(\Phi^{(1)})^2 \Phi^{(3)}}{\Phi^{(2)}} \frac{\rho}{\mu_2^2 - \rho} \mu_2$$

The following lemma shows that I_ℓ is correlated with the spins:

Lemma 6.11.2 (Degree-Corrected Extension of Lemma 41 in [17]). *Let $\ell = C \log_\rho n$ with $C < C_{\text{coupling}}$ and $i \in \{+, -\}$. There exists a random variable Y_i such that $\mathbb{E}[Y_i] = 0$, $\mathbb{E}[|Y_i|] < \infty$ and for any continuity point t of the distribution of Y_i , in L^2 ,*

$$\frac{1}{n} \sum_{v=1}^n 1_{\sigma(v)=i} 1_{I_\ell(v) \mu_2^{-2\ell} - \hat{c} g_2(i) \geq t} \rightarrow \frac{1}{2} \mathbb{P}(Y_i \geq t).$$

Recall from Theorem 6.2.1 that the eigenvector ξ_2 is asymptotically *aligned* with

$$\frac{B^\ell B^{*\ell} \check{\chi}_2}{\|B^\ell B^{*\ell} \check{\chi}_2\|}, \quad (6.68)$$

where $\ell \sim \log_\rho(n)$. Hence, for some unknown sign ω , the vector $\xi'_2 = \omega \xi_2$ is asymptotically *close* to (6.68). From Lemma 6.9.1 we know that $B^\ell B^{*\ell} \check{\chi}_2$ and $P_{2,\ell}$ are asymptotically close. Consequently, properly renormalizing ξ'_2 will make it asymptotically close to $P_{2,\ell}$, so that we can replace $P_{2,\ell}$ in (6.67) by ξ'_2 . That is, we set for $v \in V$,

$$I(v) = \sum_{e:e_2=v} s \sqrt{n} \xi'_2(e),$$

with $s = \sqrt{c_2''''}$ the limit in Proposition 6.8.4. Then, I and $I_\ell / \mu_2^{2\ell}$ are close, which leads to the following lemma:

Lemma 6.11.3 (Degree-Corrected Extension of Lemma 42 in [17]). *Let $i \in \{+, -\}$ and \widehat{Y}_i be as in Lemma 6.11.2. For any continuity point t of the distribution of \widehat{Y}_i , in L^2 ,*

$$\frac{1}{n} \sum_{v=1}^n 1_{\sigma(v)=i} 1_{I(v) - \widehat{c}g_2(i) \geq t} \rightarrow \frac{1}{2} \mathbb{P}(\widehat{Y}_i \geq t).$$

Put for $i \in \{+, -\}$, $X_i = \widehat{Y}_i + \widehat{c}g_2(i) = \widehat{Y}_i + \frac{1}{\sqrt{2}} \widehat{c}i$. Then, for all $t \in \mathbb{R}$ that are continuity points of the distribution of X_i , the following convergence holds in probability

$$\frac{1}{n} \sum_{v=1}^n 1_{\sigma(v)=i} 1_{I(v) > t} \rightarrow \frac{1}{2} \mathbb{P}(X_i > t).$$

Since $\mathbb{E}[X_+] > 0$, the argument below (90) in [17] establishes the existence of a continuity point $t_0 \in \mathbb{R}$ such that $\mathbb{P}(X_+ > t_0) > \mathbb{P}(X_- > t_0)$.

Further, we note that X_+ is in distribution equal to $-X_-$, a fact that we use below.

We are now in a position to apply Lemma 6.11.1 and thereby finishing the proof of Theorem 6.2.2:

If $\omega = 1$, then we define F , for $v \in V$, by

$$F(v) = 1_{\sum_{e:e_2=v} \xi_2(e) > \frac{t_0}{s\sqrt{n}}} = 1_{I(v) > t_0}.$$

Then,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{v=1}^n 1_{\sigma(v)=+} F(v) = \frac{1}{2} \mathbb{P}(X_+ > t_0) =: \frac{f(+)}{2},$$

and,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{v=1}^n 1_{\sigma(v)=-} F(v) = \frac{1}{2} \mathbb{P}(X_- > t_0) =: \frac{f(-)}{2},$$

so that $f(+)$ > $f(-)$ and Lemma 6.11.1 applies.

If, however, $\omega = -1$, then we define F , for $v \in V$, by

$$F(v) = 1_{\sum_{e:e_2=v} \xi_2(e) \leq \frac{t_0}{s\sqrt{n}}} = 1_{-I(v) \leq t_0}.$$

Then, this time,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{v=1}^n 1_{\sigma(v)=+} F(v) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{v=1}^n 1_{\sigma(v)=+} 1_{I(v) > -t_0} = \frac{1}{2} \mathbb{P}(X_+ > -t_0) =: \frac{f(+)}{2},$$

since $-t_0$ is a continuity point of X_+ , which follows from the fact that X_+ is in distribution equal to $-X_-$ and t_0 is a continuity point of X_- .

Similarly,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{v=1}^n 1_{\sigma(v)=-} F(v) = \frac{1}{2} \mathbb{P}(X_- > -t_0) =: \frac{f(-)}{2}.$$

Now,

$$f(+) = \mathbb{P}(X_+ > -t_0) = 1 - \mathbb{P}(X_- > t_0) > 1 - \mathbb{P}(X_+ > t_0) = \mathbb{P}(X_- > -t_0) = f(-),$$

exactly the setting of Lemma 6.11.1.

6.12 Proofs of Section 6.6

Proof of Theorem 6.6.1. For $1 \leq q < t$, we have

$$Z_t - M^{t-s} Z_s = \sum_{u=s}^{t-1} M^{t-u-1} (Z_{u+1} - M Z_u),$$

consequently, as $g_k^* M = \mu_k g_k^*$,

$$\frac{\langle g_k, Z_t \rangle}{\mu_k^{t-1}} = \frac{\langle g_k, Z_q \rangle}{\mu_k^{q-1}} + \sum_{u=q}^{t-1} \frac{\langle g_k, Z_{u+1} - M Z_u \rangle}{\mu_k^u}, \quad (6.69)$$

compare to (55) in [17]. Hence, $(X_k(t))_{t \geq 1}$ is an \mathcal{F}_t -martingale with mean 0. We shall invoke Doob's martingale convergence theorem to prove the assertion. That is, we shall show that for some $C > 0$ and all $t \geq 1$,

$$\mathbb{E} [X_k^2(t) | Z_1] \leq C \|Z_1\|_1.$$

Let, for $i, j \in \{+, -\}$, $Z_{s+1}(i, j)$ denote the number of type i individuals in generation $s+1$ which descend from from a type j particle in the s -th generation. Then,

$$\mathbb{E} [\|Z_{s+1} - M Z_s\|_2^2 | Z_s] = \sum_{i, j \in \{+, -\}} \mathbb{E} [(Z_{s+1}(i, j) - M_{ij} Z_s(j))^2 | Z_s(j)]. \quad (6.70)$$

We calculate first, for some integer $z \geq 0$,

$$\begin{aligned} \mathbb{E} [(Z_{s+1}(i, j) - M_{ij} Z_s(j))^2 | Z_s(j) = z] &= \mathbb{E} \left[\left(\sum_{l=1}^z (Y_l(i, j) - M_{ij}) \right)^2 \middle| Z_s(j) = z \right] \\ &= \sum_{l=1}^z \mathbb{E} [(Y_l(i, j) - M_{ij})^2], \end{aligned} \quad (6.71)$$

where $(Y_l(i, j))_{l=1}^z$ are i.i.d. copies of $\text{Poi}\left(\frac{1_{i=j} a + 1_{i \neq j} b}{2} \Phi^{(1)} \phi^*\right)$, where ϕ^* follows the biased law ν^* .

Put $c_1 = \max_{i, j \in \{+, -\}} \mathbb{E} [(Y_l(i, j) - M_{ij})^2] < \infty$. Then, plugging (6.71) into (6.70), we obtain

$$\mathbb{E} [\|Z_{s+1} - M Z_s\|_2^2 | Z_s] \leq 2c_1 \|Z_s\|_1.$$

Consequently,

$$\begin{aligned}\mathbb{E} [\|Z_{s+1} - MZ_s\|_2^2 | Z_1] &= \mathbb{E} [\mathbb{E} [\|Z_{s+1} - MZ_s\|_2^2 | Z_s] | Z_1] \\ &\leq 2c_1 \mathbb{E} [\|Z_s\|_1 | Z_1] \\ &= 2c_1 \rho^{s-1} \|Z_1\|_1.\end{aligned}\tag{6.72}$$

Combining the above with (6.69) for $q = 1$, we obtain

$$\begin{aligned}\mathbb{E} [X_k^2(t) | Z_1] &= \sum_{s=1}^{t-1} \frac{\mathbb{E} [\langle g_k, (Z_{s+1} - MZ_s) \rangle^2 | Z_1]}{\mu_k^{2s}} \\ &\leq \|g_k\|_2^2 \sum_{s=1}^{t-1} \frac{\mathbb{E} [\|Z_{s+1} - MZ_s\|_2^2 | Z_1]}{\mu_k^{2s}} \\ &\leq 2c_1 \|g_k\|_2^2 \sum_{s=1}^{\infty} \frac{\rho^{s-1}}{\mu_k^{2s}} \|Z_1\|_1.\end{aligned}\tag{6.73}$$

The assertion now follows upon noting that

$$C := 2c_1 \max_{k \in \{+, -\}} \|g_k\|_2^2 \sum_{s=1}^{\infty} \frac{\rho^{s-1}}{\mu_k^{2s}} < \infty,$$

since $\rho < \mu_k^2$. \square

Proof of Corollary 6.6.2. From Theorem 6.6.1 we know that there exists a random variable $X_k(\infty)$ such that

$$X_k(t) := \frac{\langle g_k, Z_t \rangle}{\mu_k^{t-1}} - \langle g_k, Z_1 \rangle \xrightarrow{\text{a.s.}} X_k(\infty),$$

as $t \rightarrow \infty$. Now,

$$\langle g_k, Z_1 \rangle = \mu_{k, \psi_o} \langle g_k, Z_0 \rangle + \langle g_k, Z_1 - M_{\psi_o} Z_0 \rangle.$$

We combine this with the definition of $X_k(t)$ to obtain

$$\frac{\langle g_k, Z_t \rangle}{\mu_{k, \psi_o} \mu_k^{t-1}} = \langle g_k, Z_0 \rangle + \frac{\langle g_k, Z_1 - M_{\psi_o} Z_0 \rangle}{\mu_{k, \psi_o}} + \frac{X_k(t)}{\mu_{k, \psi_o}},$$

where the right hand side is seen to converge in both senses to the random variable

$$Y_{k, \psi_o}(\infty) = \langle g_k, Z_0 \rangle + \frac{\langle g_k, Z_1 - M_{\psi_o} Z_0 \rangle}{\mu_{k, \psi_o}} + \frac{X_k(\infty)}{\mu_{k, \psi_o}}.$$

Indeed,

$$\left| \frac{\langle g_k, Z_t \rangle}{\mu_{k, \psi_o} \mu_k^{t-1}} - Y_{k, \psi_o}(\infty) \right| \leq \frac{1}{\mu_{k, \phi_{\min}}} |X_k(t) - X_k(\infty)|,$$

for all ψ_o . The uniform convergence follows, since

$$\begin{aligned}&\mathbb{E} [|X_k(t) - X_k(\infty)|^2 | \phi_0 = \psi_o] \\ &= \sum_{z=0}^{\infty} \mathbb{E} [|X_k(t) - X_k(\infty)|^2 | \|Z_1\| = z] \mathbb{P} (\|Z_1\| = z | \phi_0 = \psi_o) \\ &\leq e^{\frac{a+b}{2} \Phi^{(1)}(\phi_{\max} - \phi_{\min})} \mathbb{E} [|X_k(t) - X_k(\infty)|^2 | \phi_0 = \phi_{\max}]\end{aligned}\tag{6.74}$$

\square

Proof of Theorem 6.6.3. For $1 \leq q < t$, we have again

$$\frac{\langle g_k, \Psi_t \rangle}{\mu_k^{t-1}} = \frac{\langle g_k, \Psi_q \rangle}{\mu_k^{q-1}} + \sum_{u=q}^{t-1} \frac{\langle g_k, \Psi_{u+1} - M\Psi_u \rangle}{\mu_k^u}. \quad (6.75)$$

Since $\mathbb{E}[\Psi_{u+1}|\Psi_u] = M\Psi_u$, $(X_k(t))_{t \geq 1}$ is an \mathcal{G}_t -martingale with mean 0. We show again that for some $C > 0$ and all $t \geq 1$,

$$\mathbb{E}[X_k^2(t)|Z_1] \leq C\|Z_1\|_1.$$

Let, for $i, j \in \{+, -\}$, $\Psi_{s+1}(i, j)$ denote the sum over the weights of type i individuals in generation $s+1$ which descend from a type j particle in the s -th generation. Then,

$$\mathbb{E}[\|\Psi_{s+1} - M\Psi_s\|_2^2 | Z_s] = \sum_{i, j \in \{+, -\}} \mathbb{E}[(\Psi_{s+1}(i, j) - M_{ij}\Psi_s(j))^2 | Z_s(j)]. \quad (6.76)$$

We calculate first, for some integer $z \geq 0$,

$$\mathbb{E}[(\Psi_{s+1}(i, j) - M_{ij}\Psi_s(j))^2 | Z_s(j) = z] = \mathbb{E}\left[\left(\sum_{l=1}^z \left(\sum_{l'=1}^{Y_l(i, j)} \phi_{l'}^i - M_{ij}\phi_l^j\right)\right)^2 \middle| Z_s(j) = z\right] \quad (6.77)$$

where $\phi_{l'}^i$ and ϕ_l^j are all independent and governed by the biased law ν^* , and where $(Y_l(i, j))_{l=1}^z$ are i.i.d. copies of $\text{Poi}\left(\frac{1_{i=j}a + 1_{i \neq j}b}{2}\Phi^{(1)}\phi^*\right)$, with ϕ^* governed by ν^* . Thus the summands indexed by l are independent. We have

$$\mathbb{E}\left[\sum_{l'=1}^{Y_l(i, j)} \phi_{l'}^i - M_{ij}\phi_l^j \middle| Z_s(j)\right] = \frac{1_{i=j}a + 1_{i \neq j}b}{2}\Phi^{(1)}\frac{\Phi^{(2)}}{\Phi^{(1)}}\frac{\Phi^{(2)}}{\Phi^{(1)}} - M_{ij}\frac{\Phi^{(2)}}{\Phi^{(1)}} = 0$$

Therefore,

$$\mathbb{E}[(\Psi_{s+1}(i, j) - M_{ij}\Psi_s(j))^2 | Z_s(j) = z] = \sum_{l=1}^z \mathbb{E}\left[\left(\sum_{l'=1}^{Y_l(i, j)} \phi_{l'}^i - M_{ij}\phi_l^j\right)^2\right]. \quad (6.78)$$

Put $c_1 = \max_{i, j \in \{+, -\}} \mathbb{E}\left[\left(\sum_{l'=1}^{Y_l(i, j)} \phi_{l'}^i - M_{ij}\phi_l^j\right)^2\right] < \infty$. Then, plugging (6.78) into (6.76), we obtain

$$\mathbb{E}[\|\Psi_{s+1} - M\Psi_s\|_2^2 | Z_s] \leq 2c_1\|Z_s\|_1. \quad \square$$

Proof of Lemma 6.6.4. For $k \geq 1$, put

$$\epsilon_k = \rho^{-k/2}\sqrt{k} \quad \text{and} \quad f_k = \prod_{\ell=1}^k (1 + \epsilon_\ell).$$

Due to convergence of $(f_k)_k$, there exist constants $c_0, c_1 > 0$ such that for all $k \geq 1$,

$$c_0 \leq f_k \leq c_1 \quad \text{and} \quad \epsilon_k \leq c_1, \quad (6.79)$$

exactly as (57) in [17].

Recall the law of S_{k+1} from (6.28). We shall firstly derive a concentration result for $\sum_{l=1}^{S_k} X_k^{(l)}$, by using Hoeffding's inequality. Note that by definition $X_k^{(l)} \in \frac{a+b}{2} \Phi^{(1)}[\phi_{\min}, \phi_{\max}]$. Put $\gamma = (\frac{a+b}{2} \Phi^{(1)})^2 (\phi_{\max} - \phi_{\min})^2$, then Hoeffding's equality reads

$$\mathbb{P} \left(\left| \sum_{l=1}^n X_k^{(l)} - n\rho \right| \geq t \right) \leq 2 \exp \left(-\frac{2t^2}{n\gamma} \right).$$

Hence, in particular,

$$\mathbb{P} \left(\left| \sum_{l=1}^{sf_k \rho^k} X_k^{(l)} - sf_k \rho^k \rho \right| \geq sf_k \rho^k \rho \frac{\epsilon_{k+1}}{2} \right) \leq 2 \exp \left(-\frac{f_k \rho (k+1)}{2\gamma} s \right) \leq 2 \exp(-c_2 s), \quad (6.80)$$

for some $c_2 > 0$, due to (6.79). We use the last result to obtain

$$\begin{aligned} & \mathbb{P} \left(S_{k+1} > sf_{k+1} \rho^{k+1} | S_k \leq sf_k \rho^k \right) \\ & \leq \mathbb{P} \left(\text{Poi} \left(\sum_{l=1}^{sf_k \rho^k} X_k^{(l)} \right) > sf_{k+1} \rho^{k+1} \right) \\ & \leq \mathbb{P} \left(\text{Poi} \left(sf_k \rho^{k+1} \left(1 + \frac{\epsilon_{k+1}}{2} \right) \right) > sf_{k+1} \rho^{k+1} \right) (1 - 2e^{-c_2 s}) \\ & \quad + 2e^{-c_2 s}. \end{aligned} \quad (6.81)$$

We bound

$$\begin{aligned} sf_{k+1} \rho^{k+1} &= sf_k \rho^{k+1} \left(1 + \frac{\epsilon_{k+1}}{2} \right) \frac{1 + \epsilon_{k+1}}{1 + \frac{\epsilon_{k+1}}{2}} \\ &\geq sf_k \rho^{k+1} \left(1 + \frac{\epsilon_{k+1}}{2} \right) (1 + c_3 \epsilon_{k+1}), \end{aligned}$$

where $c_3 = \frac{1}{2} \frac{1}{1 + \max_l \epsilon_l / 2} > 0$. Combining the last estimate with (6.81) and the inequality

$$\mathbb{P}(\text{Poi}(\lambda) \geq \lambda s) \leq e^{-\lambda I(s)},$$

where

$$I : x \mapsto \begin{cases} x \log x - x + 1 & \text{if } x > 0; \\ \infty & \text{if } x \leq 0, \end{cases} \quad (6.82)$$

entails that

$$\mathbb{P} \left(S_{k+1} > sf_{k+1} \rho^{k+1} | S_k \leq sf_k \rho^k \right) \leq \exp \left(-sf_k \rho^{k+1} \left(1 + \frac{\epsilon_{k+1}}{2} \right) I(1 + c_3 \epsilon_{k+1}) \right) + 2e^{-c_2 s}.$$

It remains to bound $I(1 + c_3 \epsilon_k)$ from below. But, due to the form of I , there exists a $\theta > 0$ such that for $x \in [0, c_3 \max_k \epsilon_k]$, $I(1 + x) \geq \theta x^2$. Consequently

$$\mathbb{P} \left(S_{k+1} > sf_{k+1} \rho^{k+1} | S_k \leq sf_k \rho^k \right) \leq 3e^{-c_4 s k},$$

for some constant $c_4 > 0$. Hence,

$$\mathbb{P} \left(\exists k : S_k > sc_1 \rho^k \right) \leq \sum_{k=1}^{\infty} 3e^{-c_4 s k} = \frac{3}{1 - e^{-c_4 s}} e^{-c_4 s},$$

from which the statement follows. \square

Proof of Theorem 6.6.5. We claim that there exist constants $c, c' > 0$ such that for any $s \geq 0$

$$\mathbb{P} \left(\|Z_{t+1} - MZ_t\|_2 > s \|Z_t\|_1^{1/2} \mid \mathcal{F}_t \right) \leq c' e^{-c(s \wedge s^2)}. \quad (6.83)$$

To prove (6.83), we shall employ Hoeffding's inequality to establish a concentration result for

$$\lambda^+ = \frac{\Phi^{(1)}}{2} \left(a \sum_{i=1}^{Z_t^+} \Phi_i^+ + b \sum_{i=1}^{Z_t^-} \Phi_i^- \right), \quad (6.84)$$

and,

$$\lambda^- = \frac{\Phi^{(1)}}{2} \left(b \sum_{i=1}^{Z_t^+} \Phi_i^+ + a \sum_{i=1}^{Z_t^-} \Phi_i^- \right) \quad (6.85)$$

around their respective means $y^+ = \mathbb{E}_*[\lambda^+]$ and $y^- = \mathbb{E}_*[\lambda^-]$, where $(\Phi_i^\pm)_i$ are i.i.d. random variables with law ν^* , and where $\mathbb{E}_*[\cdot] = \mathbb{E}[\cdot | Z_t]$. This in conjunction with the classical tail bound for $Y \stackrel{d}{=} \text{Poi}(\lambda)$:

$$\mathbb{P}(|Y - \lambda| > \lambda s) \leq 2e^{-\lambda \delta(s)}, \quad (6.86)$$

where $\delta : x \mapsto I(1-x) \wedge I(1+x)$, with I defined in (6.82), shall allow us to prove concentration of $\begin{pmatrix} Z_{t+1}^+ \\ Z_{t+1}^- \end{pmatrix} = \begin{pmatrix} \text{Poi}(\lambda^+) \\ \text{Poi}(\lambda^-) \end{pmatrix}$ around $\mathbb{E}_* \left[\begin{pmatrix} Z_{t+1}^+ \\ Z_{t+1}^- \end{pmatrix} \right] = \begin{pmatrix} y^+ \\ y^- \end{pmatrix} = MZ_t$.

Let $t^+, t^- > 0$. Then, Hoeffding's inequality gives

$$\mathbb{P}_* \left(\left| \sum_{i=1}^{Z_t^\pm} \Phi_i^\pm - Z_t^\pm \frac{\Phi^{(2)}}{\Phi^{(1)}} \right| \geq t^\pm \right) \leq 2 \exp \left(-\frac{2(t^\pm)^2}{Z_t^\pm \gamma} \right), \quad (6.87)$$

where $\gamma = (\phi_{\min} - \phi_{\max})^2$, and where $\mathbb{P}_*(\cdot) = \mathbb{P}(\cdot | Z_t)$. Hence,

$$\begin{aligned} & \mathbb{P}_* \left(|\lambda^+ - y^+| \leq \frac{\Phi^{(1)}}{2} (at^+ + bt^-) \right) \\ & \geq \mathbb{P}_* \left(\left| \sum_{i=1}^{Z_t^+} \Phi_i^+ - Z_t^+ \frac{\Phi^{(2)}}{\Phi^{(1)}} \right| \leq t^+, \left| \sum_{i=1}^{Z_t^-} \Phi_i^- - Z_t^- \frac{\Phi^{(2)}}{\Phi^{(1)}} \right| \leq t^- \right) \\ & \geq \left(1 - 2 \exp \left(-\frac{2(t^+)^2}{Z_t^+ \gamma} \right) \right) \left(1 - 2 \exp \left(-\frac{2(t^-)^2}{Z_t^- \gamma} \right) \right). \end{aligned} \quad (6.88)$$

Plugging $t^+ = \frac{s\sqrt{y^+}}{\sqrt{3}\Phi^{(1)}_a}$ and $t^- = \frac{s\sqrt{y^-}}{\sqrt{3}\Phi^{(1)}_b}$ into the last equation leads to

$$\begin{aligned} & \mathbb{P}_* \left(|\lambda^+ - y^+| \leq \frac{s}{2} \|y\|_1^{1/2} \right) \\ & \geq \left(1 - 2 \exp \left(-\frac{4/3}{(\Phi^{(1)})^2 a^2 \gamma} \frac{y^+}{Z_t^+} s^2 \right) \right) \left(1 - 2 \exp \left(-\frac{4/3}{(\Phi^{(1)})^2 a^2 \gamma} \frac{y^-}{Z_t^-} s^2 \right) \right) \\ & \geq \left(1 - 2e^{-c_0 s^2} \right)^2 \\ & \geq 1 - 4e^{-c_0 s^2}, \end{aligned} \quad (6.89)$$

for some constant $c_0 > 0$, since $\frac{y^\pm}{Z_t^\pm}$ is bounded away from zero by some constant.

We use the last inequality to obtain

$$\begin{aligned} \mathbb{P}_* \left(Z_{t+1}^+ - y^+ > s \|y\|_1^{1/2} \right) &\leq \mathbb{P}_* \left(\text{Poi} \left(y^+ + \frac{s}{2} \|y\|_1^{1/2} \right) - \left(y^+ + \frac{s}{2} \|y\|_1^{1/2} \right) > \frac{s}{2} \|y\|_1^{1/2} \right) \\ &\quad + 4e^{-c_0 s^2}. \end{aligned} \quad (6.90)$$

We continue by invoking (6.86),

$$\begin{aligned} &\mathbb{P}_* \left(\text{Poi} \left(y^+ + \frac{s}{2} \|y\|_1^{1/2} \right) - \left(y^+ + \frac{s}{2} \|y\|_1^{1/2} \right) > \frac{s}{2} \|y\|_1^{1/2} \right) \\ &\leq 2 \exp \left(- \left(y^+ + \frac{s}{2} \|y\|_1^{1/2} \right) \delta \left(\frac{\frac{s}{2} \|y\|_1^{1/2}}{y^+ + \frac{s}{2} \|y\|_1^{1/2}} \right) \right). \end{aligned}$$

We note the existence of a $\theta > 0$ such that for all $x \in [0, 1]$, $\delta(x) \geq \theta x^2$, so that

$$\left(y^+ + \frac{s}{2} \|y\|_1^{1/2} \right) \delta \left(\frac{\frac{s}{2} \|y\|_1^{1/2}}{y^+ + \frac{s}{2} \|y\|_1^{1/2}} \right) \geq \frac{\theta \frac{s^2}{4} \|y\|_1}{y^+ + \frac{s}{2} \|y\|_1^{1/2}} \geq c_2 (s^2 \wedge s),$$

for some constant $c_2 > 0$, because $y^+ + \frac{s}{2} \|y\|_1^{1/2} \leq \max\{2y^+, s \|y\|_1^{1/2}\}$.

Similarly, to bound $\mathbb{P}_* \left(Z_{t+1}^+ - y^+ \leq -s \|y\|_1^{1/2} \right)$ from above, we need to estimate

$$\begin{aligned} &\mathbb{P}_* \left(\text{Poi} \left(y^+ - \frac{s}{2} \|y\|_1^{1/2} \right) - \left(y^+ - \frac{s}{2} \|y\|_1^{1/2} \right) \leq -\frac{s}{2} \|y\|_1^{1/2} \right) \\ &\leq 2 \exp \left(- \left(y^+ - \frac{s}{2} \|y\|_1^{1/2} \right) \delta \left(\frac{\frac{s}{2} \|y\|_1^{1/2}}{y^+ - \frac{s}{2} \|y\|_1^{1/2}} \right) \right), \end{aligned} \quad (6.91)$$

when $y^+ > \frac{s}{2} \|y\|_1^{1/2}$ (if $y^+ < \frac{s}{2} \|y\|_1^{1/2}$, then $Z_{t+1}^+ - y^+ > -\frac{s}{2} \|y\|_1^{1/2}$, so that $\mathbb{P}_* \left(Z_{t+1}^+ - y^+ \leq -s \|y\|_1^{1/2} \right) = 0$).

We distinguish between two cases: Firstly, when $y^+ - \frac{s}{2} \|y\|_1^{1/2} > \frac{s}{2} \|y\|_1^{1/2}$, we have

$$\left(y^+ - \frac{s}{2} \|y\|_1^{1/2} \right) \delta \left(\frac{\frac{s}{2} \|y\|_1^{1/2}}{y^+ - \frac{s}{2} \|y\|_1^{1/2}} \right) \geq \frac{\theta \frac{s^2}{4} \|y\|_1}{y^+ - \frac{s}{2} \|y\|_1^{1/2}} \geq \theta \frac{\|y\|_1 s^2}{y^+ 4} \geq c_3 s^2, \quad (6.92)$$

for some constant c_3 , due to our observation above.

Secondly, in case $y^+ - \frac{s}{2} \|y\|_1^{1/2} < \frac{s}{2} \|y\|_1^{1/2}$, we use the existence of a $\theta' > 0$ such that for all $x \geq 1$, $\delta(x) \geq \theta' x$:

$$\left(y^+ - \frac{s}{2} \|y\|_1^{1/2} \right) \delta \left(\frac{\frac{s}{2} \|y\|_1^{1/2}}{y^+ - \frac{s}{2} \|y\|_1^{1/2}} \right) \geq \theta' \frac{\|y\|_1^{1/2}}{2} s \geq c_4 s, \quad (6.93)$$

for some constant $c_4 > 0$.

Combining (6.90) - (6.93), leads to

$$\begin{aligned} \mathbb{P} \left(|Z_{t+1}^+ - y^+| > s \|y\|_1^{1/2} \right) &\leq 2 \left(e^{-c_2 (s^2 \wedge s)} + e^{-c_4 s} + e^{-c_3 s^2} \right) + 8e^{-c_0 s^2} \\ &\leq c_5 e^{-c_6 (s^2 \wedge s)}. \end{aligned} \quad (6.94)$$

An identical bound holds (after possibly redefining the values of c_5 and c_6) for $|Z_{t+1}^- - y^-|$.

Finally, noting that $\|y\|_1 = \rho\|Z_t\|_2$, we have

$$\begin{aligned} \mathbb{P}\left(\|Z_{t+1} - MZ_t\|_2 > s\|Z_t\|_1^{1/2} \mid \mathcal{F}_t\right) &\leq \mathbb{P}\left(|Z_{t+1}^+ - y^+| \geq \frac{s}{\sqrt{2}}\|Z_t\|_1^{1/2} \mid \mathcal{F}_t\right) \\ &\quad + \mathbb{P}\left(|Z_{t+1}^- - y^-| \geq \frac{s}{\sqrt{2}}\|Z_t\|_1^{1/2} \mid \mathcal{F}_t\right) \\ &\leq c'e^{-c(s^2 \wedge s)}, \end{aligned} \tag{6.95}$$

that is exactly claim (6.83).

We are now in a position to derive a similar bound as (59) in [17]:

$$\mathbb{P}\left(\forall t \geq 1 : \|Z_{t+1} - MZ_t\|_2 \leq u(t+1)\log n\|Z_t\|_1^{1/2}\right) \geq 1 - c' \sum_{t \geq 1} e^{-cut \log n} \geq 1 - C'n^{-Cu}. \tag{6.96}$$

Recalling (6.69), we have, for $s \geq 1$,

$$|\langle g_k, Z_s \rangle - \mu_k^{s-t} \langle g_k, Z_t \rangle| \leq \mu_k^{s-1} \|g_k\|_2 \sum_{u=s}^{t-1} \frac{\|Z_{u+1} - MZ_u\|_2}{\mu_k^u}.$$

From Equation (6.96) we know that, for all $u \geq 1$,

$$\|Z_{u+1} - MZ_u\|_2 \leq c_9(\log n)(u+1)\|Z_u\|_1^{1/2}, \tag{6.97}$$

where c_9 is so large that 6.97 holds with probability $1 - n^{-\beta}$. Further, $\|Z_h\|_1$ itself is bounded by Lemma 6.6.4:

$$\|Z_h\|_1 \leq c_{10}(\log n)\rho^h, \tag{6.98}$$

also with probability at least $1 - n^{-\beta}$.

With the same probability, for $k \in \{1, 2\}$,

$$\begin{aligned} |\langle g_k, Z_s \rangle - \mu_k^{s-t} \langle g_k, Z_t \rangle| &\leq c_{11}(\log n)^{3/2} \mu_k^{s-1} \sum_{u=s}^{t-1} (u+1) \frac{\sqrt{\rho}^u}{\mu_k} \\ &\leq c_{12}(\log n)^{3/2} (s+1) \rho^{s/2}. \end{aligned} \tag{6.99}$$

The proof the last claim, write

$$\langle g_k, \Psi_s \rangle - \mu_k^{s-t} \langle g_k, \Psi_t \rangle = \frac{\Phi^{(2)}}{\Phi^{(1)}} (\langle g_k, Z_s \rangle - \mu_k^{s-t} \langle g_k, Z_t \rangle) + \epsilon_s - \mu_k^{s-t} \epsilon_t, \tag{6.100}$$

where, for $s \geq 1$,

$$\epsilon_s = g_k(+)\left(\Psi_s(+)-Z_s^+ \frac{\Phi^{(2)}}{\Phi^{(1)}}\right) + g_k(-)\left(\Psi_s(-)-Z_s^- \frac{\Phi^{(2)}}{\Phi^{(1)}}\right).$$

We bound ϵ_t using (6.87),

$$\mathbb{P}\left(\forall t \geq 1 : \epsilon_t \leq t \log n \|Z_t\|_1^{1/2}\right) \geq 1 - c_{13} \sum_{t \geq 1} e^{-c_{14} t^2 \log^2 n} \geq 1 - C'n^{-Cu}. \tag{6.101}$$

So that, with probability $1 - n^{-\beta}$,

$$|\epsilon_s - \mu_k^{s-t} \epsilon_t| \leq c_{15} \log^{5/2}(n) \left(\rho^{s/2} + |\mu_k|^{s-t} \rho^{t/2}\right) \leq c_{16} \log^{5/2}(n) \rho^{s/2},$$

since $|\mu_k| > \rho^{1/2}$. \square

Proof of Theorem 6.6.6. We have,

$$\|\Psi_{u+1} - M\Psi_u\|_2 \leq \frac{\Phi^{(2)}}{\Phi^{(1)}} \|Z_{u+1} - MZ_u\|_2 + \|\Psi_{u+1} - \frac{\Phi^{(2)}}{\Phi^{(1)}} Z_{u+1}\|_2 + \|M \left(\Psi_u - \frac{\Phi^{(2)}}{\Phi^{(1)}} Z_u \right)\|_2. \quad (6.102)$$

We use (6.87), to obtain that for any $\beta > 0$ (similar to (6.96))

$$\mathbb{P} \left(\forall t \geq 1 : \|\Psi_t - \frac{\Phi^{(2)}}{\Phi^{(1)}} Z_t\|_2 \leq t \log n \|Z_t\|_1^{1/2} \right) \geq 1 - n^{-\beta}. \quad (6.103)$$

Combing (6.102), (6.103) and (6.96), gives that with probability $1 - n^{-\beta}$, for all $u \geq 1$,

$$\|\Psi_{u+1} - M\Psi_u\|_2 \leq c_2 u \log n \|Z_u\|_1^{1/2}. \quad (6.104)$$

We can now apply the argument at the end of Theorem 24 in [17]. The second claim follows by using the last part of the proof of Theorem 24 in [17], where the variable U needs to be replaced by

$$U = \sup_{t \geq 1} \frac{\|\Psi_{t+1} - M\Psi_t\|_2}{t \|Z_t\|_1^{1/2}}.$$

It is important here that $\mathbb{E}[U^4] = \mathcal{O}(1)$, which is ensured by (6.104). \square

Proof of Theorem 6.6.7. We start by calculating the expectation and variance of $\sum_{u \in Y_t^o} L_{k,\ell}^u$ conditional on \mathcal{F}_t (defined in Theorem 6.6.1). We use this to show that, as $\ell \rightarrow \infty$, uniformly for all ψ_o ,

$$\frac{\bar{Q}_{k,\ell}}{\mu_k^{2\ell}} \rightarrow \frac{\Phi^{(3)}}{\Phi^{(2)}} \frac{\rho}{\mu_k^2 - \rho} \mu_{k,\psi_o} Y_{k,\psi_o}(\infty), \quad (6.105)$$

almost surely and in L^2 , where $Y_{k,\psi_o}(\infty)$ is defined in Corollary 6.6.2, and where

$$\bar{Q}_{k,\ell} = \sum_{t=0}^{\ell-1} \mathbb{E}_{\mathcal{F}_t} \sum_{u \in Y_t^o} L_{k,\ell}^u.$$

The latter is reminiscent of

$$Q_{k,\ell} = \sum_{t=0}^{\ell-1} \sum_{u \in Y_t^o} L_{k,\ell}^u,$$

and we show that $\bar{Q}_{k,\ell}$ and $Q_{k,\ell}$ are in fact close in L^2 -distance:

$$\|\bar{Q}_{k,\ell} - Q_{k,\ell}\| = o(\mu_k^{2\ell}).$$

Consider for $t \geq 0$ and $\ell \geq t + 2$,

$$\begin{aligned} \mathbb{E}_{\mathcal{F}_t, Y_t^o, Y_1^u} L_{k,\ell}^u &= \sum_{(v,w) \in Y_1^u, v \neq w} \mathbb{E}_{\mathcal{F}_t, Y_t^o, Y_1^u} S_{\ell-t-1}^{wv} \mathbb{E}_{\mathcal{F}_t, Y_t^o, Y_1^u} \langle g_k, \Psi_t^v \rangle \\ &= \sum_{(v,w) \in Y_1^u, v \neq w} \rho_w \rho^{\ell-t-2} \phi_v \mu_k^t \langle g_k, Z_0^v \rangle \end{aligned} \quad (6.106)$$

where $\rho_w = \frac{a+b}{2} \Phi^{(1)} \phi_w$, with ϕ_w a random variable that follows law ν^* . The second equality in (6.106) follows after calculating

$$\mathbb{E}[\Psi_t^v | Y_0^v] = \frac{\Phi^{(2)}}{\Phi^{(1)}} \mathbb{E}[Z_t^v | Y_0^v] = \frac{\Phi^{(2)}}{\Phi^{(1)}} \frac{\Phi^{(1)} \phi_v}{\Phi^{(2)}} M^t Z_0^v = \phi_v M^t Z_0^v,$$

where the factor $\frac{\Phi^{(1)}\phi_v}{\Phi^{(2)}}$ accounts for the fact that the "parental" vertex v has *deterministic* type ϕ_v (and transitions are thus given by $M_{\phi_v} = \frac{\Phi^{(1)}\phi_v}{\Phi^{(2)}}M$), whereas vertices in the later generations have i.i.d. weights (for which M is the transition matrix). Now,

$$\begin{aligned}\mathbb{E}_{\mathcal{F}_t, Y_t^o} L_{k,\ell}^u &= \mathbb{E}_{\mathcal{F}_t, Y_t^o} \mathbb{E}_{\mathcal{F}_t, Y_t^o, Y_1^u} L_{k,\ell}^u \\ &= \mathbb{E}_{\mathcal{F}_t, Y_t^o} \sum_{(v,w) \in Y_1^u, v \neq w} \rho_w \rho^{\ell-t-2} \phi_v \mu_k^t \langle g_k, Z_0^v \rangle \\ &= \rho^{\ell-t-2} \mu_k^t \mathbb{E}_{\mathcal{F}_t, Y_t^o} |Y_1^u| (|Y_1^u| - 1) \mathbb{E}_{\mathcal{F}_t, Y_t^o} \rho^* \mathbb{E}_{\mathcal{F}_t, Y_t^o} \phi^* \langle g_k, \begin{pmatrix} 1_{\sigma^*=+} \\ 1_{\sigma^*=-} \end{pmatrix} \rangle,\end{aligned}\tag{6.107}$$

where ϕ^* has law ν^* , ρ^* is an i.i.d. copy of $\frac{a+b}{2}\Phi^{(1)}\phi^*$ and $\sigma^* = \sigma_u$ with probability $\frac{a}{a+b}$, and $\sigma^* = -\sigma_u$ with probability $\frac{b}{a+b}$ (further, ρ^* , ϕ^* and σ^* are independent).

We thus have

$$\mathbb{E}_{\mathcal{F}_t, Y_t^o} L_{k,\ell}^u = \rho^{\ell-t-2} \mu_k^t \cdot \rho_u^2 \cdot \rho \cdot \frac{\Phi^{(2)}}{\Phi^{(1)}} \cdot (g_k(1)c(\sigma_u, +) + g_k(2)c(\sigma_u, -)),\tag{6.108}$$

where $\rho_u = \frac{a+b}{2}\Phi^{(1)}\phi_u$ (with ϕ_u the weight of u) and for $(x, y) \in \{+, -\} \times \{+, -\}$, $c(x, y) = \frac{a}{a+b}$ if $x = y$ and $c(x, y) = \frac{b}{a+b}$ otherwise.

Now, as g_k is an eigenvector of M with eigenvalue μ_k , we have

$$(g_k(1)c(\sigma_u, +) + g_k(2)c(\sigma_u, -)) = \frac{2}{a+b} \frac{\mu_k}{\Phi^{(2)}} \langle g_k, Z_0^u \rangle = \frac{\mu_k}{\rho} \langle g_k, Z_0^u \rangle.$$

Together with (6.107) this gives

$$\mathbb{E}_{\mathcal{F}_t, Y_t^o} L_{k,\ell}^u = \rho^{\ell-t-2} \mu_k^{t+1} \rho_u^2 \frac{\Phi^{(2)}}{\Phi^{(1)}} \langle g_k, Z_0^u \rangle.\tag{6.109}$$

Summing over $u \in Y_t^o$ using the last equation yields

$$\begin{aligned}\mathbb{E}_{\mathcal{F}_t} \sum_{u \in Y_t^o} L_{k,\ell}^u &= \mathbb{E}_{\mathcal{F}_t} \sum_{u \in Y_t^o} \mathbb{E}_{\mathcal{F}_t, Y_t^o} L_{k,\ell}^u \\ &= \rho^{\ell-t-2} \mu_k^{t+1} \frac{\Phi^{(2)}}{\Phi^{(1)}} \mathbb{E}_{\mathcal{F}_t} \sum_{u \in Y_t^o} \rho_u^2 \langle g_k, Z_0^u \rangle \\ &= \rho^{\ell-t-2} \mu_k^{t+1} \langle g_k, Z_t \rangle \left(\frac{a+b}{2} \right)^2 \Phi^{(2)} \cdot \begin{cases} \psi_o \Phi^{(2)} & \text{if } t = 0; \\ \Phi^{(3)} & \text{if } t > 0. \end{cases}\end{aligned}\tag{6.110}$$

We leave it to the reader to verify that the same inequality holds for $l = t + 1$.

We continue by bounding the variance of $L_{k,\ell}^u$:

$$\begin{aligned}\text{Var}_{\mathcal{F}_t} L_{k,\ell}^u &\leq \mathbb{E}_{\mathcal{F}_t} (L_{k,\ell}^u)^2 \\ &= \mathbb{E}_{\mathcal{F}_t} \sum_{(v,w) \in Y_1^u, v \neq w} \sum_{(v',w') \in Y_1^u, v' \neq w'} S_{\ell-t-1}^{w'} S_{\ell-t-1}^w \langle g_k, \Psi_t^v \rangle \langle g_k, \Psi_t^{v'} \rangle \\ &\leq \mathbb{E}_{\mathcal{F}_t} |Y_1^u|^2 \mathbb{E}_{\infty} S_{\ell-t-1}^2 \mathbb{E}_{\infty} \langle g_k, \Psi_t^v \rangle^2,\end{aligned}\tag{6.111}$$

where $\mathbb{E}_{\infty}[\cdot] = \max_{\tau' \in \{+, -\}} \mathbb{E}[\cdot | \phi_o = \phi_{\max}, \sigma_o = \tau']$. Now, $\mathbb{E}_{\mathcal{F}_t} |Y_1^u|^2 \leq c_0$. From Lemma 6.6.4, we know that $S_k \stackrel{d}{\leq} \text{Exp}(c_1 \rho^k)$, hence $\mathbb{E}_{\infty} S_{\ell-t-1}^2 \leq 2c_1^2 (\rho^{\ell-t-1})^2$. To

bound $\mathbb{E}_\infty \langle g_k, \Psi_t^v \rangle^2$, recall from Theorem 6.6.3 that

$$\mathbb{E} \left[\left(\frac{\langle \phi_k, \Psi_t \rangle}{\mu_k^{t-1}} - \langle g_k, \Psi_1 \rangle \right)^2 \middle| Z_1 \right] \leq C_2 \|Z_1\|_1.$$

Consequently, as $\mathbb{E} [\|Z_1\|_1]$ is bounded,

$$\mathbb{E}_\infty \langle g_k, \Psi_t^v \rangle^2 \leq c_3 \mu_k^{2t}.$$

Returning to (6.111), we have

$$\text{Var}_{\mathcal{F}_t} \sum_{u \in Y_t^o} L_{k,\ell}^u \leq c_4 \mu_k^{2t} \rho^{2(\ell-t)} S_t. \quad (6.112)$$

We have

$$\begin{aligned} \bar{Q}_{k,\ell} &= \sum_{t=0}^{\ell-1} \mathbb{E}_{\mathcal{F}_t} \sum_{u \in Y_t^o} L_{k,\ell}^u \\ &= \rho^\ell \mu_k \langle g_k, Z_0 \rangle \psi_o + \sum_{t=1}^{\ell-1} \rho^{\ell-t} \mu_k^{t+1} \langle g_k, Z_t \rangle \frac{\Phi^{(3)}}{\Phi^{(2)}} \\ &= \rho^\ell \mu_k \langle g_k, Z_0 \rangle \psi_o + \frac{\Phi^{(3)}}{\Phi^{(2)}} \sum_{t=1}^{\ell-1} \rho^{\ell-t} \mu_k^{2t} \mu_{k,\psi_o} Y_{k,\psi_o}(t), \end{aligned} \quad (6.113)$$

where $Y_{k,\psi_o}(t)$ is defined in Corollary 6.6.2.

We consider

$$\frac{\bar{Q}_{k,\ell}}{\mu_k^{2\ell}} = o(1) + \frac{\Phi^{(3)}}{\Phi^{(2)}} \sum_{t=1}^{\ell-1} \left(\frac{\mu_k^2}{\rho} \right)^{t-\ell} \mu_{k,\psi_o} Y_{k,\psi_o}(t), \quad (6.114)$$

and verify our claim (6.105). To do so, split for arbitrary *fixed* $\epsilon > 0$,

$$\sum_{t=1}^{\ell-1} r^{t-\ell} Y_k(t) = \sum_{t=1}^{T_\epsilon-1} r^{t-\ell} Y_k(t) + \sum_{t=T_\epsilon}^{\ell-1} r^{t-\ell} Y_k(t),$$

where $r = \frac{\mu_k^2}{\rho}$, Y_k is shorthand notation for Y_{k,ψ_o} , and where

$$T_\epsilon = \min\{t : \forall s \geq t, |Y_k(\infty) - Y_k(s)| \leq \epsilon\}.$$

Then,

$$\sum_{t=1}^{T_\epsilon-1} r^{t-\ell} Y_k(t) \leq |\sup_t Y_k(t)| r^{-\ell} r^{T_\epsilon} T_\epsilon \xrightarrow{\text{a.s.}} 0,$$

as $\ell \rightarrow \infty$, since $(Y_k(t))_t$ is convergent (uniformly in ψ_o) and hence bounded. Further,

$$\begin{aligned} \sum_{t=T_\epsilon}^{\ell-1} r^{t-\ell} Y_k(t) &= \sum_{u=1}^{\ell-T_\epsilon} (Y_k(\infty) + \mathcal{O}(\epsilon)) \\ &\xrightarrow{\text{a.s.}} \sum_{u=1}^{\infty} r^{-u} (Y_k(\infty) + \mathcal{O}(\epsilon)) \\ &= \frac{1}{r-1} (Y_k(\infty) + \mathcal{O}(\epsilon)), \end{aligned} \quad (6.115)$$

where the limit is taken for $\ell \rightarrow \infty$. Since $\epsilon > 0$ was arbitrary, (6.105) follows.

L^2 -convergence follows from [17] (this convergence takes place uniformly for all ψ_o due to Theorem 6.6.1).

Further, that $\|\bar{Q}_{k,\ell} - Q_{k,\ell}\| = o(\mu_k^{2\ell})$ can be established by following the proof in [17]. Indeed, from the latter proof we know that, for some constant c_6 independent of ψ_o ,

$$\begin{aligned} \|Q_{k,\ell} - \bar{Q}_{k,\ell}\|_2 &\leq \sum_{t=0}^{\ell-1} \left\| \left(\text{Var}_{\mathcal{F}_t} \left(\sum_{u \in Y_t^o} L_{k,\ell}^u \right) \right)^{1/2} \right\|_2 \\ &\leq c_5 \sum_{t=0}^{\ell-1} \mu_k^t \rho^{\ell-t} \|\sqrt{S_t}\|_2 \\ &\leq c_6 \mu_k^\ell \rho^{\ell/2}, \end{aligned} \tag{6.116}$$

due to the variance bound in (6.112) and Lemma 6.6.4.

Finally, combining the uniform bounds (6.105) and (6.116), entails that

$$\left\| \frac{Q_{k,\ell}}{\mu_k^{2\ell}} - \frac{\Phi^{(3)}}{\Phi^{(2)}} \frac{\rho}{\mu_k^2 - \rho} \mu_{k,\psi_o} Y_{k,\psi_o}(\infty) \right\|_2 \rightarrow 0,$$

uniformly for all ψ_o . □

Proof of Theorem 6.6.8. Using (6.111) and Theorem 6.6.6, we have

$$\text{Var}_{\mathcal{F}_t} L_{k,\ell}^u \leq c_1 \rho^{2(\ell-t)} t^3 \rho^t.$$

Plugging this bound, together with (6.110) here, into (66) in [17] establishes the claim. □

Proof of Theorem 6.6.9. Recall the explicit expressions for $Q_{1,\ell}$ and $Q_{2,\ell}$ from (6.30), respectively (6.31). Now, conditional on \mathcal{T} and the weights (denoted by \mathcal{T}_ϕ), $\mathcal{P}_{2\ell+1}$ is deterministic, hence

$$\mathbb{E}[Q_{1,\ell} Q_{2,\ell} | \mathcal{T}, \mathcal{T}_\phi] = Q_{1,\ell} \sum_{(u_0, \dots, u_{2\ell+1}) \in \mathcal{P}_{2\ell+1}} \phi_{u_{2\ell+1}} \mathbb{E}[\sigma(u_{2\ell+1}) | \mathcal{T}] = 0,$$

because, $\mathbb{E}[\sigma(u) | \mathcal{T}, \sigma_o] = \left(\frac{a-b}{a+b}\right)^{|u|} \sigma_o$, for a vertex u at distance $|u|$ from the root, by construction of the branching process. □

6.13 Proofs of Section 6.7

Proof of Proposition 6.7.1. The second statement follows from the first after recalling that $(G, e)_\ell = (G', e_2)_\ell$, where G' is the graph G with edge $\{e_1, e_2\}$ removed. Since $e \in \vec{E}$, e_2 then has a biased weight governed by ν^* .

In Section 5.7, we established a coupling between the branching process and the DC-SBM where the spins are drawn *uniformly* from $\{+, -\}$, with error probability $n^{-\frac{1}{2} \log(4/e)}$.

Thus, we are done if we couple the neighbourhoods in the latter graph to the DC-SBM with *deterministic* spins under consideration here.

Now, with probability at least $1 - e^{-\Omega(n^{-1/2})}$ we can couple the graphs such that at most $c_1 n^{\frac{3}{4} \vee (1-\gamma)}$ have unequal spins (call the corresponding set of vertices S) and

all weights are equal. Further, we may assume that the subgraphs obtained after removing S are identical.

The ℓ -neighbourhoods in both graphs are exactly the same if they are both disjoint with S . Conditional on $|S|$ and $|G_\ell|$, this happens with probability at least $1 - c_2 \frac{|G_\ell||S|}{n}$.

From Section 5.7, we know that with probability $1 - n^{-\log(4/e)}$, $|G_\ell| < n^{\frac{1}{8} \wedge \frac{\gamma}{2}}$.

Thus, conditional on the bounds for $|S|$ and $|G_\ell|$, the neighbourhoods are the same with probability at least $1 - c_3 n^{-(\frac{1}{8} \vee \frac{\gamma}{2})}$.

All together, $\mathbb{P}((G, v)_\ell = (T, o)_\ell) \geq 1 - c_4 n^{-(\frac{1}{8} \wedge \frac{\gamma}{2}) \wedge (\frac{1}{2} \log(4/e))}$. \square

Proof of Corollary 6.7.3. This proof follows the proof of Corollary 32 in [17]. Indeed (although with a slightly different probability) the graph neighbourhood $(Y_t(e))_{0 \leq t \leq \ell}$ and branching process $(Z_t)_{0 \leq t \leq \ell}$ coincide again, and moreover, the weights are equal in both processes. \square

Proof of Lemma 6.7.4. As observed in [17], the second statement follows from the first.

Adapting Section 5.7, at step m in the exploration process, the weights of the vertices in $\mathcal{U}(m)$ are independent, and those with spin τ have weight governed by $\nu_\tau^{(m)}$, where

$$d\nu_\tau^{(m)}(\psi) = \frac{g_\tau(\psi)}{\int_{\phi_{\min}}^{\phi_{\max}} g_\tau(\psi') d\nu(\psi')} d\nu(\psi),$$

where $g_\tau(\cdot) = \prod_{i=1}^m \left(1 - \frac{\kappa(x_i, \tau)}{n}\right)$, with $x_u = \sigma_u \phi_u$ the types of the already explored vertices and $\kappa(x, y) = |xy|(1_{\{xy > 0\}}a + 1_{\{xy < 0\}}b)$.

We claim that variables following $\nu_\tau^{(m)}$ are stochastically dominated by variables governed by ν . Indeed, use that for any non-decreasing $f, h : \mathbb{R} \rightarrow \mathbb{R}$ and any random variable X we have $\mathbb{E}[f(X)h(X)] \geq \mathbb{E}[f(X)]\mathbb{E}[h(X)]$. Then, for $\psi \geq 0$,

$$\nu_\tau^{(m)}([0, \psi]) = \frac{\mathbb{E}[-g_\tau(\phi) \cdot -1_{\phi \leq \psi}]}{\mathbb{E}[g_\tau(\phi)]} \geq \frac{\mathbb{E}[g_\tau(\phi)] \mathbb{E}[1_{\phi \leq \psi}]}{\mathbb{E}[g_\tau(\phi)]} = \nu([0, \psi]),$$

with $\phi \sim \nu$.

Secondly, we claim that the weight of a vertex when it is just discovered is stochastically dominated by variables governed by ν^* . To prove this, let $m \geq 0$ and assume the claim to hold for all $l \leq m$. Consider vertex v explored in step $m+1$ (itself discovered in step, say, $l \leq m$) with weight $\phi_v^{*(l)}$. Its children are selected from the set $\mathcal{U}^{(m)}$ in which they have independent weights $(\phi_u^{(m)})_{u \in \mathcal{U}^{(m)}}$ all stochastically dominated by ν . We compare this to a setting \mathcal{S} where a particle with weight $\phi^* \sim \nu^*$ has its children selected following the same rules from a reservoir of $|\mathcal{U}^{(m)}|$ particles with spins as in $\mathcal{U}^{(m)}$ and i.i.d. weights $(\phi_u)_{u \in \mathcal{U}^{(m)}} \sim \nu$. Due to the assumed stochastic domination, there exists a coupling of the exploration process and the setting \mathcal{S} , such that pointwise $\phi_v^{*(l)} \leq \phi^*$ and $\phi_u^{(m)} \leq \phi_u$ for all u . To decide whether $u \in \mathcal{U}^{(m)}$ is selected as a child, we can draw uniformly from $[0, 1]$ a number U_u and include u in the exploration process exactly when $\frac{(1_{\sigma_u = \sigma_v} a + 1_{\sigma_u = -\sigma_v} b) \phi_v^{*(l)} \phi_u^{(m)}}{n} \geq U_u$ and in the setting \mathcal{S} exactly when $\frac{(1_{\sigma_u = \sigma_v} a + 1_{\sigma_u = -\sigma_v} b) \phi^* \phi_u}{n} \geq U_u$. Since by assumption $\phi^* \phi_u \geq \phi_v^{*(l)} \phi_u^{(m)}$, for each u , we conclude that the newly selected particles are also stochastically dominated.

Denote the vertices in S_t by $1, \dots, S_t$ and their weights by $(\widehat{\phi}_v^*)_{v \in S_t}$. We shall use the same strategy as in Lemma 6.6.4 to bound

$$S_{t+1} = \sum_{v=1}^{S_t} \widehat{D}_v^*,$$

where \widehat{D}_v^* is the offspring-size of v . In particular, to use large deviation theory as in (6.80), we shall calculate for $\theta \geq 0$, $\mathbb{E} \left[e^{\theta \sum_{v=1}^{S_t} \widehat{D}_v^*} \middle| S_t \right]$. Caution is needed here as the variables $(\widehat{D}_v^*)_{v \in S_t}$ are *not* independent. Let \mathcal{F}_m be the sigma-algebra generated by the exploration process upto step m (included). If vertex v is explored in step $m+1$, then,

$$\widehat{D}_v^* = \sum_{u \in \mathcal{U}^{(m)}} \text{Ber} \left((1_{\sigma_u = \sigma_v} a + 1_{\sigma_u = -\sigma_v} b) \frac{\widehat{\phi}_v^* \phi_u^{(m)}}{n} \right),$$

where we recall that conditioned on \mathcal{F}_m , $\phi_u^{(m)}$ is stochastically dominated by ν and $\widehat{\phi}_v^*$ by ν^* . Hence, using that $1 + y \leq e^y$ for all $y \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E} \left[e^{\theta \widehat{D}_v^*} \middle| \mathcal{F}_m, \widehat{\phi}_v^* \right] &\leq \mathbb{E} \left[\prod_u \left(1 + \frac{\widehat{\phi}_v^* \phi_u^{(m)}}{n} (1_{\sigma_u = \sigma_v} a + 1_{\sigma_u = -\sigma_v} b) (e^\theta - 1) \right) \middle| \mathcal{F}_m, \widehat{\phi}_v^* \right] \\ &\leq \left(1 + a \frac{\widehat{\phi}_v^* \Phi^{(1)}}{n} (e^\theta - 1) \right)^{n_{\sigma_v}} \left(1 + b \frac{\widehat{\phi}_v^* \Phi^{(1)}}{n} (e^\theta - 1) \right)^{n - n_{\sigma_v}} \\ &\leq e^{r_n \widehat{\phi}_v^* \Phi^{(1)} (e^\theta - 1)}, \end{aligned} \tag{6.117}$$

where $r_n = \max\{\frac{n_+ a + n_- b}{n}, \frac{n_- a + n_+ b}{n}\}$. Thus, if ϕ^* has law ν^* ,

$$\mathbb{E} \left[e^{\theta \widehat{D}_v^*} \middle| \mathcal{F}_m \right] \leq \mathbb{E} \left[e^{r_n \phi^* \Phi^{(1)} (e^\theta - 1)} \right], \tag{6.118}$$

since for $t \geq 0$, $\mathbb{E} [e^{tX}] \leq \mathbb{E} [e^{tY}]$ if $X \stackrel{d}{\leq} Y$. Iterating (6.118), we obtain

$$\mathbb{E} \left[e^{\theta \sum_{v=1}^{S_t} \widehat{D}_v^*} \middle| S_t \right] \leq \left(\mathbb{E} \left[e^{r_n \phi^* \Phi^{(1)} (e^\theta - 1)} \right] \right)^{S_t} = \mathbb{E} \left[e^{r_n \sum_{v=1}^{S_t} \phi_v^* \Phi^{(1)} (e^\theta - 1)} \middle| S_t \right],$$

where $\{\phi_v^*\}_v$ are i.i.d. with law ν^* . Thus, we have

$$\mathbb{E} \left[e^{\theta \sum_{v=1}^{S_t} \widehat{D}_v^*} \right] \leq \mathbb{E} \left[e^{\theta \text{Poi}(\sum_{v=1}^{S_t} r_n \phi_v^* \Phi^{(1)})} \right],$$

compare this to (6.28): the characteristic function of $\sum_{v=1}^{S_t} \widehat{D}_v^*$ is dominated by the characteristic function of the Poisson-mixture in (6.28) if we replace $\frac{a+b}{2}$ with r_n . Hence we can repeat the proof of Lemma 6.6.4, with $\rho_n := r_n \Phi^{(2)}$ instead of ρ . \square

Proof of Lemma 6.7.5. Fix a vertex v . Let $m \geq 0$ be the smallest integer such that all vertices within distance R of v have been revealed at step m of the exploration process. Now, the exploration process constructs a spanning tree \mathcal{T}_m for $G_R(v)$. However, edges between vertices in ∂G_r ($r \leq \ell$) are not inspected, and neither is it verified whether two vertices in ∂G_r share a common neighbour in ∂G_{r+1} ($r \leq R-1$). The number of those uninspected edges is bounded by $|G_r|^2$. Hence, among them at

most $\text{Bin}(|G_r|^2, \frac{c_1}{n})$ are actually present in G_r . Thus, using twice Markov's inequality in conjunction with Lemma 6.7.4, for some $c_2 > 0$,

$$\mathbb{P}(G_r(v) \text{ is not a tree}) \leq \mathbb{E}[|G_r|^2] \frac{c_1}{n} \leq \frac{c_3 \rho^{2\ell}}{n},$$

and,

$$\mathbb{P}\left(\sum_v 1_{G_r(v) \text{ is not a tree}} \geq \rho^{2\ell} \log(n)\right) \leq \frac{c_4}{\log(n)}.$$

For the other claim, if the graph is tangled, then there is a vertex such that among its uninspected edges in the exploration process at step m , at least two are in fact present. Now,

$$\mathbb{P}\left(\text{Bin}\left(|G_r|^2, \frac{c_1}{n}\right) \geq 2\right) \leq \left(\frac{c_1}{n}\right)^4 \mathbb{E}[|G_r|^4] \leq \frac{c_5 \rho^{4\ell}}{n^4}.$$

A union bound over all vertices then gives

$$\mathbb{P}(G \text{ tangled}) \leq \frac{c_6 \rho^{4\ell}}{n^3} = o(1).$$

□

Proof of Proposition 6.7.6. (i) follows from Lemma 6.7.4 and Corollary 6.7.3.

To prove (ii), recall that $B_{\vec{e}\vec{g}}^r$ is the number of non-backtracking paths of length r (i.e., containing $r + 1$ edges) between \vec{e} and \vec{g} . Further, if $G_r(e_2)$ is a tree, then there is exactly one path between e and any edge g on the tree. Hence

$$\langle B^r \chi_k, \delta_e \rangle = \langle g_k, \Psi_r(e) \rangle.$$

An appeal to Corollary 6.7.3 then establishes (ii).

Further, (iii) follows from the fact that G is ℓ -tangle-free with high probability, so that there are at most two non-backtracking walks of length r between any edges \vec{e} and \vec{f} . Thus,

$$|\langle B^r \chi_k, \delta_e \rangle| \leq 2 \|g_k\|_\infty \phi_{\max} S_t(e) \leq \log^2(n) \rho^r,$$

with probability at least $1 - e^{-\Omega(n)}$, due to Lemma 6.7.4. □

Proof of Corollary 6.7.7. We start with the case $\mu_2^2 > \rho$. Using that $\langle B^\ell \chi_k, x \rangle = 0$ and Proposition 6.7.6 (iii), we write,

$$\begin{aligned} |\langle B^r \chi_k, x \rangle| &= \left| \sum_{e \in \vec{E}_\ell} x_e \langle B^r \chi_k, \delta_e \rangle + \sum_{e \notin \vec{E}_\ell} x_e \langle B^r \chi_k, \delta_e \rangle \right. \\ &\quad \left. - \mu_k^{r-\ell} \left(\sum_{e \in \vec{E}_\ell} x_e \langle B^\ell \chi_k, \delta_e \rangle + \sum_{e \notin \vec{E}_\ell} x_e \langle B^\ell \chi_k, \delta_e \rangle \right) \right| \\ &\leq (\log n)^2 \rho^r \sqrt{|\vec{E}_\ell|} + \sum_{e \notin \vec{E}_\ell} |x_e| \|\langle B^r \chi_k, \delta_e \rangle - \mu_k^{r-\ell} \langle B^\ell \chi_k, \delta_e \rangle\| \\ &\quad + \mu_k^{r-\ell} \log(n)^2 \rho^\ell \sqrt{|\vec{E}_\ell|}. \end{aligned} \tag{6.119}$$

Now, $|\mu_k| > 1$ and for $e \notin \vec{E}_\ell$, bound (ii) in Proposition 6.7.6 applies, so that w.h.p.

$$\begin{aligned} |\langle B^r \chi_k, x \rangle| &\leq 2\rho^\ell (\log n)^2 \sqrt{|\vec{E}_\ell|} + \rho^{r/2} (\log n)^4 \sqrt{|E|} \\ &\leq \rho^\ell (\log n)^3 n^{\frac{1}{2} - \frac{\gamma}{4} \wedge \frac{1}{80}} + \rho^{r/2} (\log n)^{\frac{9}{2}} n^{\frac{1}{2}} \\ &\leq \rho^{r/2} (\log n)^5 n^{1/2}, \end{aligned} \quad (6.120)$$

since $\rho^\ell = n^C \ll n^{\frac{\gamma}{4} \wedge \frac{1}{80}}$.

In case $\mu_2^2 \leq \rho$, redefine \vec{E}_ℓ as the set of oriented edges such that $(G, e)_\ell$ is not a tree or $|\langle g_1, \Psi_t(e) \rangle - \rho^{t-\ell} \langle g_1, \Psi_\ell(e) \rangle| > (\log n)^4 \rho^{t/2}$ or $|\langle g_2, \Psi_t(e) \rangle| > (\log n)^4 \rho^{t/2}$. Note that $|\vec{E}_\ell|$ can now be bounded with the same arguments as in the proof of Corollary 6.7.3.

Write $\langle B^r \chi_k, x \rangle = \sum_{e \in \vec{E}_\ell} x_e \langle B^r \chi_k, \delta_e \rangle + \sum_{e \notin \vec{E}_\ell} x_e \langle B^r \chi_k, \delta_e \rangle$. To bound the sum over E_ℓ , use Cauchy-Schwartz inequality and Proposition 6.7.6 (iii), which also holds if $\mu_2^2 \leq \rho$. For the second sum, use that, if $e \notin \vec{E}_\ell$, then $|\langle B^r \chi_k, \delta_e \rangle| \leq (\log n)^4 \rho^{r/2}$, as follows from Theorem 6.6.6 and the coupling result for local neighbourhoods. \square

6.14 Proofs of Section 6.8

Proof of Proposition 6.8.1. We start by using the law of total variance for $Y = \sum_{v=1}^n \tau(G, v)$:

$$\text{Var}(Y) = \mathbb{E}[\text{Var}(Y | \phi_1, \dots, \phi_n)] + \text{Var}(\mathbb{E}[Y | \phi_1, \dots, \phi_n]),$$

and shall apply Efron-Stein's inequality on both terms.

Define the function h for $(\psi_1, \dots, \psi_n) \in [\phi_{\min}, \phi_{\max}]^n$ as $h(\psi_1, \dots, \psi_n) = \mathbb{E}[Y | \phi_1 = \psi_1, \dots, \phi_n = \psi_n]$. We need to bound $|h(\psi_1, \dots, \psi_{k-1}, \psi_k, \psi_{k+1}, \dots, \psi_n) - h(\psi_1, \dots, \psi_{k-1}, \psi'_k, \psi_{k+1}, \dots, \psi_n)|^2$ for arbitrary $\psi'_k \in [\phi_{\min}, \phi_{\max}]$. Denote by $G_{\psi_1, \dots, \psi_k, \dots, \psi_n}$ the random graph G , conditional on $\phi_1 = \psi_1, \dots, \phi_n = \psi_n$. Assume without loss of generality that $\psi_k \geq \psi'_k$. Then, there exists a coupling of $G_{\psi_1, \dots, \psi_k, \dots, \psi_n}$ and $G_{\psi_1, \dots, \psi'_k, \dots, \psi_n}$ such that $G_{\psi_1, \dots, \psi'_k, \dots, \psi_n}$ is a subgraph of $G_{\psi_1, \dots, \psi_k, \dots, \psi_n}$ obtained after removing some edges between k and its neighbours in the latter graph. For this coupling, $|\tau(G_{\psi_1, \dots, \psi_k, \dots, \psi_n}, u) - \tau(G_{\psi_1, \dots, \psi'_k, \dots, \psi_n}, u)|$ is nonzero only if $u \in V(G_{\psi_1, \dots, \psi_k, \dots, \psi_n}, k)_\ell$, and it is bounded by $\max_v \varphi(G_{\psi_1, \dots, \psi_k, \dots, \psi_n}, v) + \max_v \varphi(G_{\psi_1, \dots, \psi'_k, \dots, \psi_n}, v)$. Consequently,

$$\begin{aligned} &|h(\psi_1, \dots, \psi_{k-1}, \psi_k, \psi_{k+1}, \dots, \psi_n) - h(\psi_1, \dots, \psi_{k-1}, \psi'_k, \psi_{k+1}, \dots, \psi_n)|^2 \\ &\leq \mathbb{E} \left[|V(G_{\psi_1, \dots, \psi_k, \dots, \psi_n}, k)_\ell| \left(\max_v \varphi(G_{\psi_1, \dots, \psi_k, \dots, \psi_n}, v) + \max_v \varphi(G_{\psi_1, \dots, \psi'_k, \dots, \psi_n}, v) \right) \right]^2 \\ &\leq \mathbb{E} \left[|V(G_{k, \infty}, k)|^2 | \phi_1 = \psi_1, \dots, \phi_{k-1} = \psi_{k-1}, \phi_{k+1} = \psi_{k+1}, \dots, \phi_n = \psi_n \right] \\ &\quad \cdot 3\mathbb{E} \left[\max_v \varphi^2(G, v) \mid \phi_1 = \psi_1, \dots, \phi_k = \psi_k, \dots, \phi_n = \psi_n \right] \\ &\quad + \mathbb{E} \left[|V(G_{k, \infty}, k)|^2 | \phi_1 = \psi_1, \dots, \phi_{k-1} = \psi_{k-1}, \phi_{k+1} = \psi_{k+1}, \dots, \phi_n = \psi_n \right] \\ &\quad \cdot 3\mathbb{E} \left[\max_v \varphi^2(G, v) \mid \phi_1 = \psi_1, \dots, \phi_k = \psi'_k, \dots, \phi_n = \psi_n \right] \end{aligned} \quad (6.121)$$

where $G_{k, \infty}$ is the random graph G conditioned on $\phi_k = \phi_{\max}$, and where we used Hölder's inequality and the fact that $(x + y)^2 \leq 3(x^2 + y^2)$ for any $x, y \in \mathbb{R}$. Hence,

using again Hölder's inequality, Efron-Stein's inequality becomes

$$\begin{aligned} \text{Var}(\mathbb{E}[Y|\phi_1, \dots, \phi_n]) &\leq \frac{1}{2} \sum_{k=1}^n \mathbb{E}[|h(\phi_1, \dots, \phi_k, \dots, \phi_n) - h(\phi_1, \dots, \phi'_k, \dots, \phi_n)|^2] \\ &\leq 3 \sum_{k=1}^n \sqrt{\mathbb{E}[|V(G_{k,\infty}, k)|_\ell^4]} \sqrt{\mathbb{E}[\max_v \varphi^4(G, v)]}, \end{aligned} \quad (6.122)$$

where $(\phi'_k)_k$ is an i.i.d. copy of $(\phi_k)_k$. Now, due to 6.7.4, $\mathbb{E}[|V(G_{k,\infty}, k)|_\ell^4] \leq \frac{c_1}{2} \rho^{4\ell}$. Thus,

$$\text{Var}(\mathbb{E}[Y|\phi_1, \dots, \phi_n]) \leq c_2 n \rho^{2\ell} \sqrt{\mathbb{E}[\max_v \varphi^4(G, v)]}.$$

To bound $\text{Var}(Y|\phi_1 = \psi_1, \dots, \phi_n = \psi_n)$ we use again Efron-Stein's inequality. Define for $1 \leq k \leq n$, $X_k = \{1 \leq v \leq k : \{v, k\} \in E\}$, where E is the edge set of G . Then, conditioned on the weights $(\phi_u = \psi_u)$, $\{X_k\}_k$ are independent. Let $\{X'_k\}_k$ be an independent copy of $\{X_k\}_k$ and define G_k as the graph on vertex set V with edge set $\cup_{v \neq k} X_v \cup X'_k$. Thus, conditional on the weights, G_k equals G except for the edges in $\{1 \leq v \leq k\}$ which are redrawn independently.

Now, for some function $F_{\psi_1, \dots, \psi_n}$,

$$\sum_{v=1}^n \tau(G, v) = F_{\psi_1, \dots, \psi_n}(X_1, \dots, X_k, \dots, X_n),$$

and hence,

$$\sum_{v=1}^n \tau(G_k, v) = F_{\psi_1, \dots, \psi_n}(X_1, \dots, X'_k, \dots, X_n).$$

Proceeding as above, we obtain

$$\begin{aligned} \text{Var}(Y|\phi_1 = \psi_1, \dots, \phi_n = \psi_n) &\leq \frac{1}{2} \sum_{k=1}^n \mathbb{E}[|F_{\psi_1, \dots, \psi_n}(X_1, \dots, X_k, \dots, X_n) - F_{\psi_1, \dots, \psi_n}(X_1, \dots, X'_k, \dots, X_n)|^2] \\ &\leq \frac{1}{2} \sum_{k=1}^n \sqrt{\mathbb{E}[|V(G, k)|_\ell^4 \cap |V(G_k, k)|_\ell^4]} \sqrt{\mathbb{E}\left[\left(\max_v \varphi(G, v) + \max_v \varphi(G_k, v)\right)^4\right]} \\ &\leq c_3 n \rho^{2\ell} \sqrt{\mathbb{E}[\max_v \varphi^4(G, v)]}. \end{aligned} \quad (6.123) \quad \square$$

Proof of Proposition 6.8.2. We recall that the coupling between neighbourhoods and branching processes is such that, in case of success, the weights are equal in both processes. Therefore, as in the proof of Proposition 36 in [17], we obtain

$$\mathbb{E}\left[\frac{1}{n} \sum_{v=1}^n \tau(G, v)\right] = \mathbb{E}[\tau(T, o)] + \epsilon(n),$$

where

$$\epsilon(n) = \mathcal{O}(n^{-\gamma}) + c_1 n^{-(\frac{\gamma}{2} \wedge \frac{1}{40})} \sqrt{\mathbb{E}[\max_{v \in [n]} \varphi^2(G, v)]} \vee \mathbb{E}[\varphi^2(T, o)].$$

This error stems from the probability for the coupling to fail.

Hence,

$$\mathbb{E} \left[\left| \frac{1}{n} \sum_{v=1}^n \tau(G, v) - \mathbb{E}[\tau(T, o)] \right| \right] \leq \sqrt{\text{Var} \left(\frac{1}{n} \sum_{v=1}^n \tau(G, v) \right)} + \epsilon(n).$$

An appeal to Proposition 6.8.1 then finishes the proof. \square

Proof of Proposition 6.8.3. We give the key steps used to prove Proposition 37 in [17] together with the main differences in the current setting. For (i), consider the branching process defined in Section 6.6, which we denote again by $Z_t(\pm)$. We denote the associated random rooted tree by (T, o) .

Put $\tau(G, v) = \sum_{e \in \vec{E}, e_1=v} \frac{\langle g_k, \Psi_\ell(e) \rangle^2}{\mu_k^{2\ell}}$. Then, $\frac{1}{n} \sum_v \tau(G, v) = \frac{1}{n} \sum_{e \in \vec{E}} \frac{\langle g_k, \Psi_\ell(e) \rangle^2}{\mu_k^{2\ell}}$ and $\tau(G, v) \leq \varphi(G, v) := \phi_{\max}^2 \frac{S_\ell^2(v)}{\rho^\ell}$. It follows from Lemma 6.7.4 that $\mathbb{E}[\max_{v \in [n]} \varphi^4(G, v)] = \mathcal{O}((\log n)^8 \rho^{4\ell})$.

We have $\tau(T, o) = \sum_{v \in Z_1^o} \frac{\langle g_k, \Psi_\ell^v \rangle^2}{\mu_k^{2\ell}}$. Theorem 6.6.3 says that $\left(\frac{\langle g_k, \Psi_\ell^v \rangle}{\mu_k^{t-1}} \right)_{t \geq 1}$ converges in L^2 and so does it conditional on $\|Z_1^o\| = 1$. Hence, $\mathbb{E}[\tau(T, o)]$ converges.

An appeal to Proposition 6.8.2 in conjunction with the triangle inequality then establishes that $\frac{1}{n} \sum_v \tau(G, v)$ converges to a constant, say c'_k .

Statement (ii) follows similarly.

The statements (iii)–(v) follow after properly choosing local functionals. We further use that $\mathbb{E}[\phi_u \phi_v g_1(\sigma_u) g_2(\sigma_v) | \mathcal{T}] = \mathbb{E}[\phi_u \phi_v \frac{1}{2} \sigma_v | \mathcal{T}] = 0$, for any two nodes u, v . Further, on the branching process, $\mathbb{E}[\langle g_k, \Psi_{2\ell} \rangle \langle g_j, \Psi_\ell \rangle | \Psi_\ell] = \langle g_j, \Psi_\ell \rangle \langle g_k, M^\ell \Psi_\ell \rangle = \mu_k^\ell \langle g_k, \Psi_\ell \rangle \langle g_j, \Psi_\ell \rangle$. \square

Proof of Proposition 6.8.4. Starting with (i), we define the local function τ as

$\tau(G, v) = \sum_{e \in \vec{E}, e_1=v} P_{k,\ell}^2(e) \mu_k^{-4\ell}$, for a rooted graph (G, v) . Let

$$M(v) = \max_{0 \leq t \leq \ell} \max_{u \in (G, v)_t} \max_{s \leq 2\ell - t} (S_s(u) / \rho^s).$$

By monotonicity, the statement of Lemma 6.7.4 holds also for $\tilde{S}_{\ell-t-1}(h)$ and $\tilde{S}_t(g)$. We use this fact to bound powers of $M(v)$ in the following calculation:

$$\begin{aligned} \tau(G, v) &\leq \rho^{-2\ell} \sum_{e \in \vec{E}, e_1=v} \left(\sum_{t=0}^{\ell-1} \sum_{f \in \mathcal{Y}_t(e)} \|g_k\|_\infty \phi_{\max} \tilde{S}_{t+1}(f) \tilde{S}_{\ell-t}(f) \right)^2 \\ &\leq c_1 \rho^{-2\ell} \sum_{e \in \vec{E}, e_1=v} \left(\sum_{t=0}^{\ell-1} \sum_{f \in \mathcal{Y}_t(e)} M^2(v) \rho^{t+1} \rho^{\ell-t} \right)^2 \\ &= c_1 (M^2(v) \rho)^2 \sum_{e \in \vec{E}, e_1=v} \left(\sum_{t=0}^{\ell-1} S_t(e) \right)^2 \\ &= c_2 (M^2(v) \rho)^2 \sum_{e \in \vec{E}, e_1=v} (M(v) \rho^\ell)^2 \\ &\leq c_2 M^7(v) \rho^{2\ell}. \end{aligned}$$

We put $\varphi(G, v) = c_2 M^7(v) \rho^{2\ell}$. Then, $\mathbb{E} [\max_v \varphi(G, v)^4] = O((\log n)^{28} \rho^{8\ell})$, and the same bound holds for $\varphi(T, o)$. From Proposition 6.8.2, we then know that

$$\mathbb{E} \left[\left| \frac{1}{n} \sum_{e \in \vec{E}} \frac{P_{k,\ell}^2(e)}{\mu_k^{4\ell}} - \mathbb{E} [\tau(T, o)] \right| \right] \leq c_3 n^{-(\frac{\gamma}{2} \wedge \frac{1}{40})} (\log n)^7 \rho^{2\ell}, \quad (6.124)$$

where

$$\begin{aligned} \tau(T, o) &= \frac{1}{\mu_k^{4\ell}} \sum_{v \in Y_1^o} P_{k,\ell}^2(o \rightarrow v) \\ &= \frac{1}{\mu_k^{4\ell}} \sum_{v \in Y_1^o} (Q_{k,\ell}^v)^2, \end{aligned} \quad (6.125)$$

where $Q_{k,\ell}^v$ is equal to $Q_{k,\ell}$ defined on the subtree of all vertices with common ancestor v .

We need to show that the expectation of $\tau(T, o)$ converges for $\ell \rightarrow \infty$. Conditional on σ_o , and $|Y_1^o|$, $\{Q_{k,\ell}^v\}_{v \in Y_1^o}$ are independent copies of $Q_{k,\ell}$ defined on the branching process in Section 6.6 where the root has spin σ_o with probability $\frac{a}{a+b}$ and random weight governed by the biased law ν^* . The uniform L^2 convergence in Theorem 6.6.7 establishes the claim.

We now prove (ii). Put $\tau(G, v) = \sum_{e \in \vec{E}, e_1=v} (P_{1,\ell}(e) + S_{1,\ell}(e))(P_{2,\ell}(e) + S_{2,\ell}(e))$. We claim that $\mathbb{E} [\tau(T, o)] = 0$. Consider $\tau(T, o) = \sum_{v \in Z_1^o} (P_{1,\ell}(o \rightarrow v) + S_{1,\ell}(o \rightarrow v))(P_{2,\ell}(o \rightarrow v) + S_{2,\ell}(o \rightarrow v))$. Firstly, for $k \in \{1, 2\}$, $P_{k,\ell}(o \rightarrow v) = Q_{k,\ell}^v$. Now, it follows from Theorem 6.6.9, that $\mathbb{E} [Q_{1,\ell}^v Q_{2,\ell}^v] = 0$, since σ_v is drawn uniformly from $\{+, -\}$.

Secondly, $S_{1,\ell}(o \rightarrow v) S_{2,\ell}(o \rightarrow v) = \frac{1}{2} \phi_o^2 \sigma_o S_\ell^2(o \rightarrow v)$ has also zero expectation.

Thirdly,

$$Q_{1,\ell}^v S_{2,\ell}(o \rightarrow v) = \frac{1}{2} \sum_{(u_0, \dots, u_{2\ell+1}) \in \mathcal{P}_{2\ell+1}^v} \phi_{u_{2\ell+1}} \phi_o \sigma_o S_\ell(o \rightarrow v), \quad (6.126)$$

where $\mathcal{P}_{2\ell+1}^v$ is $\mathcal{P}_{2\ell+1}$ (from (6.29)) defined on the subtree of all vertices with common ancestor v . The expectation of $Q_{1,\ell}^v S_{2,\ell}(o \rightarrow v)$ is thus zero since σ_o is independent of all other terms in (6.126).

Lastly, $Q_{2,\ell}^v = \sum_{(u_1, \dots, u_{2\ell+1}) \in \mathcal{P}_{2\ell+1}^v} \sigma_{u_{2\ell+1}}$, is seen to have zero expectation.

Those four statements combined establish $\mathbb{E} [\tau(T, o)] = 0$. As above, we calculate $\mathbb{E} [\max_v \varphi(G, v)^4] = O((\log n)^{28} \rho^{16\ell})$. \square

Proof of Proposition 6.8.5. Put τ as in Proposition 6.8.3 (i), then

$$\tau(T, o) = \sum_{v \in Z_1^o} \frac{\langle g_2, \Psi_\ell^v \rangle^2}{\rho^\ell}. \text{ Now,}$$

$$\mathbb{E} [\langle g_2, \Psi_\ell^v \rangle^2] = \mathbb{E} \left[\langle g_2, \frac{\Phi^{(2)}}{\Phi^{(1)}} Z_\ell^v \rangle^2 \right] + \mathbb{E} \left[\langle g_2, \Psi_\ell^v - \frac{\Phi^{(2)}}{\Phi^{(1)}} Z_\ell^v \rangle^2 \right] \geq \left(\frac{\Phi^{(2)}}{\Phi^{(1)}} \right)^2 \mathbb{E} [\langle g_2, Z_\ell^v \rangle^2].$$

Now, Theorem 2.4 in [70] says that for some random variable X with strictly positive variance, weakly, $\frac{\langle g_2, Z_\ell^v \rangle}{\rho^{\ell/2}} \rightarrow X$, as $\ell \rightarrow \infty$. Because of the weak convergence, we have for any $\theta > 0$, $\mathbb{E} \left[\left(\frac{\langle g_2, Z_\ell^v \rangle}{\rho^{\ell/2}} \right)^2 \wedge \theta \right] \rightarrow \mathbb{E} [X^2 \wedge \theta]$, as $\ell \rightarrow \infty$. Now by Lebesgue's dominated convergence theorem, $\mathbb{E} [X^2 \wedge \theta] \rightarrow \mathbb{E} [X^2] > 0$, as $\theta \rightarrow \infty$. \square

Proof of Proposition 6.8.6. Use τ from Proposition 6.8.4 (i), together with the bound $\mathbb{E} \left[Q_{2,\ell}^2 \right] \leq C\rho^{2\ell}\ell^5$ from Theorem 6.6.8. \square

6.15 Proofs of Section 6.10

6.15.1 Bound on $\|\Delta^{(k)}\|$

We set

$$m = \left\lfloor \frac{\log n}{13 \log(\log n)} \right\rfloor.$$

We bound the norm of $\|\Delta^{(k)}\|$ by using the trace method. Following (36) in [17] (which remains true for the DC-SBM), we obtain

$$\|\Delta^{(k-1)}\|^{2m} \leq \sum_{\gamma \in W_{k,m}} \prod_{i=1}^{2m} \prod_{s=1}^k \underline{A}_{\gamma_{i,s-1}\gamma_{i,s}}, \quad (6.127)$$

where $W_{k,m}$ is the collection containing all sequences of paths $\gamma = (\gamma_1, \dots, \gamma_{2m})$ such that for all i :

- $\gamma_i = (\gamma_{i,0}, \dots, \gamma_{i,k}) \in V^{k+1}$ is a non-backtracking tangle-free path of length k , and,
- $(\gamma_{i,k-1}, \gamma_{i,k}) = (\gamma_{i+1,1}, \gamma_{i+1,0})$,

where we put $\gamma_0 = \gamma_{2m}$.

Recall the notation $G(\gamma) = (V(\gamma), E(\gamma))$. Further introduce the notation $\mathbb{E}_\phi(\cdot) = \mathbb{E}[\cdot | \phi_1, \dots, \phi_n]$. We bound, for a given $\gamma \in W_{k,m}$,

$$\mathbb{E}_\phi \left(\prod_{i=1}^{2m} \prod_{s=1}^k \underline{A}_{\gamma_{i,s-1}\gamma_{i,s}} \right) = \prod_{e \in E(\gamma)} \mathbb{E}_\phi \left(\underline{A}_{e_1 e_2}^{p_{e_1 e_2}^{(\gamma)}} \right), \quad (6.128)$$

where for $e \in E(\gamma)$, $p_{e_1 e_2}^{(\gamma)}$ denotes the number of times the edge e is traversed on the walk γ . In (6.128) we used that \underline{A} is symmetric and that, conditional on the weights, edges are independently present. Note that for any edge uw , and integer p ,

$$\mathbb{E}_\phi \underline{A}_{uw}^p \leq \phi_u \phi_w \frac{W_{\sigma(u)\sigma(w)}}{n}.$$

Below in Lemma 6.15.2, we construct a spanning tree $T(\gamma) = (V(\gamma), E_T(\gamma))$ of γ . In particular, for the $e - (v - 1)$ edges not present in T , we have $\phi_u \phi_w \frac{W_{\sigma(u)\sigma(w)}}{n} \leq \frac{c_1}{n}$, with $c_1 = \phi_{\max}^2(a \vee b)$. Putting this into (6.128), we get

$$\begin{aligned} \prod_{e \in E(\gamma)} \mathbb{E}_\phi \left(\underline{A}_{e_1 e_2}^{p_{e_1 e_2}^{(\gamma)}} \right) &\leq (c_1/n)^{e-v+1} \prod_{e \in E_T(\gamma)} \phi_{e_1} \phi_{e_2} \frac{W_{\sigma(e_1)\sigma(e_2)}}{n} \\ &= (c_1/n)^{e-v+1} \prod_{u \in V(\gamma)} \phi_u^{d_u} \prod_{e \in E_T(\gamma)} \frac{W_{\sigma(e_1)\sigma(e_2)}}{n}, \end{aligned} \quad (6.129)$$

where d_u is the degree of u in the *spanning tree*. Consequently,

$$\mathbb{E} \left(\prod_{e \in E(\gamma)} \underline{A}_{e_1 e_2}^{p_{e_1 e_2}^{(\gamma)}} \right) \leq (c/n)^{e-v+1} \prod_{u \in V(\gamma)} \Phi^{(d_u)} \prod_{e \in E_T(\gamma)} \frac{W_{\sigma(e_1)\sigma(e_2)}}{n}. \quad (6.130)$$

Let $\tau : [v(\gamma)] \mapsto V(\gamma)$ be the bijection describing the order the vertices are visited for the first time. I.e., for $1 \leq u \leq v(\gamma) - 1$, $\tau(u)$ is seen for the first time, before $\tau(u + 1)$.

We shall say that a path γ_c is canonical if $V(\gamma_c) = [v(\gamma_c)]$ and the vertices are first visited in the order $1, \dots, v(\gamma_c)$. With every path γ there corresponds (through the bijection τ) a canonical path γ_c . Consequently, if $\mathbb{W}_{k,m}(v, e)$ denotes the set of canonical paths in $W_{k,m}$ with v vertices and e edges, and I_{γ_c} the set of all injections from $[v(\gamma_c)]$ to $[n]$,

$$\mathbb{E} \left[\sum_{\gamma \in W_{k,m}} \prod_{i=1}^{2m} \prod_{s=1}^k \underline{A}_{\gamma_{i,s-1}\gamma_{i,s}} \right] \leq \sum_{v=3}^{km+1} \sum_{e=v-1}^{km} \sum_{\gamma_c \in \mathbb{W}_{k,m}(v,e)} \sum_{\tau \in I_{\gamma_c}} \mathbb{E} \left[\prod_{e \in E(\gamma_c)} \underline{A}_{\tau(e_1)\tau(e_2)}^{p_{e_1 e_2}^{(\gamma_c)}} \right], \quad (6.131)$$

because any non-backtracking path has at least 3 vertices, and $v - 1 \leq e \leq km$, since (6.128) is non-zero only if each edge is traversed at least twice.

We now bound the term $\sum_{\tau \in I_{\gamma_c}} \mathbb{E} \left[\prod_{e \in E(\gamma_c)} \underline{A}_{\tau(e_1)\tau(e_2)}^{p_{e_1 e_2}^{(\gamma_c)}} \right]$ in (6.131). Using (6.130), we have,

$$\sum_{\tau \in I_{\gamma_c}} \mathbb{E} \left[\prod_{e \in E(\gamma_c)} \underline{A}_{\tau(e_1)\tau(e_2)}^{p_{e_1 e_2}^{(\gamma_c)}} \right] \leq (c_1/n)^{e-v+1} \prod_{u=1}^{v(\gamma_c)} \Phi(d_u) \sum_{\tau \in I_{\gamma_c}} \prod_{e \in E_T(\gamma_c)} \frac{W_{\sigma(\tau(e_1))\sigma(\tau(e_2))}}{n}. \quad (6.132)$$

Our objective is to compare $\prod_{u=1}^{v(\gamma_c)} \Phi(d_u) \sum_{\tau \in I_{\gamma_c}} \prod_{e \in E_T(\gamma_c)} \frac{W_{\sigma(\tau(e_1))\sigma(\tau(e_2))}}{n}$ with $n\rho^{(v-1)}$. We start by analysing the term containing the spins:

Lemma 6.15.1. *For any canonical path $\gamma_c \in \mathbb{W}_{k,m}$,*

$$\sum_{\tau \in I_{\gamma_c}} \prod_{e \in E_T(\gamma_c)} \frac{W_{\sigma(\tau(e_1))\sigma(\tau(e_2))}}{n} \leq (1 + o(1))n \left(\frac{a+b}{2} \right)^{v-1}. \quad (6.133)$$

Proof. Let l be any leaf on the tree with unique neighbour g . Then, writing $\tau_u = \tau(u)$ for $u \in \{1, \dots, v\}$,

$$\sum_{\tau \in I_{\gamma_c}} \prod_{e \in E_T(\gamma_c)} \frac{W_{\sigma(\tau(e_1))\sigma(\tau(e_2))}}{n} \leq \sum_{\tau_1=1}^n \dots \sum_{\tau_v=1}^n \prod_{e \in E_T(\gamma_c)} \frac{W_{\sigma(\tau_{e_1})\sigma(\tau_{e_2})}}{n}.$$

Keeping τ_g fixed,

$$\begin{aligned} \sum_{\tau_1=1}^n \prod_{e \in E_T(\gamma_c)} \frac{W_{\sigma(\tau_{e_1})\sigma(\tau_{e_2})}}{n} &= \prod_{e \in E_T(\gamma_c) \setminus \{g,l\}} \frac{W_{\sigma(\tau_{e_1})\sigma(\tau_{e_2})}}{n} \sum_{\tau_l=1}^n \frac{W_{\sigma(\tau_g)\sigma(\tau_l)}}{n} \\ &= \prod_{e \in E_T(\gamma_c) \setminus \{g,l\}} \frac{W_{\sigma(\tau_{e_1})\sigma(\tau_{e_2})}}{n} \left(\frac{a+b}{2} + \mathcal{O}(n^{-\gamma}) \right), \end{aligned}$$

due to assumption (6.6).

Repeating inductively this procedure (by removing leaves from the tree) proves the assertion. \square

It remains to bound $\prod_{u=1}^{v(\gamma_c)} \Phi(d_u)$. To do so, we note that, since the weights are assumed to be bounded,

$$\Phi(d_u) \leq C_2^{d_u-2} \Phi(2) \left(\Phi(1) \right)^{d_u-2}$$

if $d_u \geq 2$, with $C_2 = \frac{\phi_{\max}}{\Phi^{(1)}} > 1$. Consequently,

$$\begin{aligned} \prod_{u=1}^{v(\gamma_c)} \Phi^{(d_u)} &\leq C_2^{\sum_{u:d_u>2} (d_u-2)} \prod_{u:d_u>2} \Phi^{(2)} \left(\Phi^{(1)} \right)^{d_u-2} \prod_{u:d_u \leq 2} \Phi^{(d_u)} \\ &\leq C_2^{\sum_{u:d_u>2} (d_u-2)} \left(\Phi^{(2)} \right)^{\frac{1}{2} \sum_{u=1}^v d_u} \\ &= C_2^{\sum_{u:d_u>2} (d_u-2)} \left(\Phi^{(2)} \right)^{v-1}, \end{aligned} \quad (6.134)$$

where we used that by Jensen's inequality $(\Phi^{(1)})^2 \leq \Phi^{(2)}$.

Now, the sum $\sum_{u:d_u>2} (d_u - 2)$ is small for a tree spanning a path in $W_{k,m}$:

Lemma 6.15.2. *For any $\gamma \in W_{k,m}$, with v vertices and e edges, there exists a tree spanning γ with degrees $(d_u)_{u=1}^v$ such that:*

$$\sum_{u:d_u>2} (d_u - 2) \leq e - (v - 1) + 2m. \quad (6.135)$$

Proof. We construct a spanning tree, while traversing γ . We denote by $p(t)$ the graph constructed at step $t \geq 0$. Put $p(0) = \{\gamma_{1,0}, \emptyset\}$ and $r = s = 0$ (the meaning of these two *counters* becomes clear in the algorithm below). Consider edge f traversed in step $t + 1$ of the walk: If f or \check{f} has already been traversed, then continue with step $t + 2$. Otherwise, if both f and \check{f} have not yet been traversed, distinguish between the following cases:

1. f_1 is a leaf of $p(t)$ and
 - a) $p(t)$ contains a cycle, then if $f_2 \notin p(t)$, put $p(t + 1) = p(t) \cup f$, otherwise, if $f_2 \in p(t)$, put $p(t + 1) = p(t)$;
 - b) $p(t)$ does not contain a cycle, then put $p(t + 1) = p(t) \cup f$. If $f_2 \in p(t)$, then put $e_c = f$;
2. f_1 is *not* a leaf of $p(t)$ and
 - a) $p(t)$ contains a cycle, then put $p(t + 1) = (p(t) \setminus e_c) \cup f$. If $f_2 \in p(t)$, put $e_c = f$. Increase the value of r with one.
 - b) $p(t)$ does not contain a cycle, then put $p(t + 1) = p(t) \cup f$. If $f_2 \in p(t)$, put $e_c = f$. Otherwise, if $f_2 \notin p(t)$, increase the value of s with one.

Once the path is completely traversed, remove e_c to obtain a spanning tree.

Note that at each stage of the construction, the graph contains at most one cycle and in this case, removing e_c will make the graph into a tree.

Further, cases 1.a and 1.b do not contribute to $\sum_{u:d_u>2} (d_u - 2)$, since the leaf in $p(t)$ becomes a vertex with degree at most 2 in $p(t + 1)$. A cycle formed in step $t + 1$ will *temporarily* increase the degree of the vertex that is merged by the leaf, however this edge e_c will later be removed.

In case 2.a, the degree of vertex f_1 increases with one, however, at the same time an edge is removed. The number of times 2.a happens, r , is thus bounded by the number of times an edge is removed: $r \leq e - (v - 1)$.

In case 2.b, we need only to consider the case where no cycle is formed. But, before arriving at such a vertex considered in 2.b, the path must have made a backtrack. Hence $s \leq 2m$.

(In fact, between two subsequent occurrences of event 2, the walk should at least either make a backtrack or 'get back to the tree' by forming a cycle: giving the same bound for $s + r$).

All together,

$$\sum_{u:d_u>2} (d_u - 2) \leq r + s \leq e - (v - 1) + 2m.$$

□

Finally, we recall the bound on the cardinality of $\mathcal{W}_{k,m}$ from [17]:

Lemma 6.15.3 (Lemma 17 in [17]). *Let $\mathcal{W}_{k,m}(v, e)$ be the set of canonical paths with $v(\gamma) = v$ and $e(\gamma) = e$. We have*

$$|\mathcal{W}_{k,m}(v, e)| \leq k^{2m} (2km)^{6m(e-v+1)}. \quad (6.136)$$

Hence, combining (6.127), (6.131) - (6.136),

$$\begin{aligned} \mathbb{E} \left[\|\Delta^{(k-1)}\|^{2m} \right] &\leq \sum_{v=3}^{km+1} \sum_{e=v-1}^{km} |\mathcal{W}_{k,m}(v, e)| \left(\frac{c}{n}\right)^{e-(v-1)} n C^{e-(v-1)+2m} \rho^{v-1} \\ &\leq n c_5^m \rho^{km} \sum_{v=3}^{km+1} \sum_{e=v-1}^{km} \rho^{2m} \left(\frac{c_7(2\ell m)^{6m}}{n}\right)^{e-(v-1)} \\ &\leq n c_5^m \rho^{km} \rho^{2m} \ell m \sum_{s=0}^{\infty} \left(\frac{c_7(2\ell m)^{6m}}{n}\right)^s \\ &\leq n (c_8 \log n)^m \log^2 n \rho^{km} \\ &\leq (c_9 \log n)^{16m} \rho^{km}, \end{aligned} \quad (6.137)$$

where we used the bound on m , in particular to derive convergence of the series, and the fact that $n^{1/m} = o(\log n)^{14}$.

We finish by using Markov's inequality.

6.15.2 Bound on $\|\Delta^{(k)}\chi_i\|$

We point out the differences with bound (31) in [17]: Here, we have

$$\begin{aligned} \mathbb{E} \left[\|\Delta^{(k-1)}\chi_i\|^2 \right] &= \mathbb{E} \left[\sum_{e,f,g} \Delta_{ef}^{(k-1)} \Delta_{eg}^{(k-1)} \xi_i(f) \xi_i(g) \right] \\ &\leq \phi_{\max}^2 \mathbb{E} \left[\sum_{e,f,g} \Delta_{ef}^{(k-1)} \Delta_{eg}^{(k-1)} \right] \\ &\leq \phi_{\max}^2 \sum_{\gamma \in W''_{k,1}} \mathbb{E} \left[\prod_{i=1}^2 \prod_{s=1}^k \underline{A}_{\gamma_{i,s-1}, \gamma_{i,s}} \right], \end{aligned} \quad (6.138)$$

where $W''_{k,1}$ is defined in [17]. In the latter paper it is also shown that the same bound, Lemma 6.15.3 holds for the cardinality of $W''_{k,1}$). Hence, using the penultimate line of (6.137) with $m = 1$, gives

$$\mathbb{E} \left[\|\Delta^{(k-1)}\chi_i\|^2 \right] \leq c_1 n \log^3(n) \rho^k.$$

6.15.3 Bound on $\|R_k^{(\ell)}\|$

Put

$$m = \left\lfloor \frac{\log n}{25 \log(\log n)} \right\rfloor.$$

We apply the same strategy as above: for $0 \leq k \leq \ell - 1$, we have the bound

$$\begin{aligned} \|R_k^{(\ell-1)}\|^{2m} &\leq \text{tr} \left\{ \left(R_k^{(\ell-1)} R_k^{(\ell-1)*} \right)^m \right\} \\ &= \sum_{\gamma \in T'_{\ell,m,k}} \prod_{i=1}^{2m} \prod_{s=1}^k A_{\gamma_{i,s-1}\gamma_{i,s}} \phi_{\gamma_{i,k}} \phi_{\gamma_{i,k+1}} W_{\sigma(\gamma_{i,k})\sigma(\gamma_{i,k+1})} \prod_{s=k+2}^{\ell} A_{\gamma_{i,s-1}\gamma_{i,s}} \\ &\leq c_1^m \sum_{\gamma \in T'_{\ell,m,k}} \prod_{i=1}^{2m} \prod_{s=1}^k A_{\gamma_{i,s-1}\gamma_{i,s}} \prod_{s=k+2}^{\ell} A_{\gamma_{i,s-1}\gamma_{i,s}}, \end{aligned} \quad (6.139)$$

where $c_1 = \phi_{\max}^4 (a \vee b)^2$, and where $T'_{\ell,m,k}$ is the collection containing all sequences of paths $\gamma = (\gamma_1, \dots, \gamma_{2m})$ such that

- for all i : $\gamma_i = (\gamma_i^1, \gamma_i^2)$, where $\gamma_i^1 = (\gamma_{i,0}, \dots, \gamma_{i,k})$ and $\gamma_i^2 = (\gamma_{i,k+1}, \dots, \gamma_{i,\ell})$ are non-backtracking tangle-free;
- for all odd i : $(\gamma_{i,0}, \gamma_{i,1}) = (\gamma_{i-1,0}, \gamma_{i-1,1})$ and $(\gamma_{i,\ell-1}, \gamma_{i,\ell}) = (\gamma_{i+1,\ell-1}, \gamma_{i+1,\ell})$,

with the convention that $\gamma_0 = \gamma_{2m}$.

To calculate the expectation of $\|R_k^{(\ell-1)}\|^{2m}$, we note that

$$\mathbb{E} \left[\prod_{i=1}^{2m} \prod_{s=1}^k A_{\gamma_{i,s-1}\gamma_{i,s}} \prod_{s=k+2}^{\ell} A_{\gamma_{i,s-1}\gamma_{i,s}} \right]$$

is non-zero only if (for i fixed) each edge $\{\gamma_{i,s-1}, \gamma_{i,s}\}$ for $1 \leq s \leq k$ appears more than once in the $2(\ell-1)m$ pairs $\{\{\gamma_{j,s-1}, \gamma_{j,s}\}\}_{j=1, s \neq k+1}^{j=2m}$. Hence,

$$\mathbb{E} \left[\|R_k^{(\ell-1)}\|^{2m} \right] \leq c_1^m \sum_{\gamma \in T_{\ell,m,k}} \mathbb{E} \left[\prod_{i=1}^{2m} \prod_{s=1}^k A_{\gamma_{i,s-1}\gamma_{i,s}} \prod_{s=k+2}^{\ell} A_{\gamma_{i,s-1}\gamma_{i,s}} \right], \quad (6.140)$$

where

$$T_{\ell,m,k} = \{\gamma \in T'_{\ell,m,k} \mid v(\gamma) \leq e(\gamma) \leq km + 2m(\ell-1-k)\}. \quad (6.141)$$

Similarly as in establishing the bound on $\|\Delta^{(k)}\|$, we say that a path γ_c is *canonical* if $V(\gamma_c) = [v(\gamma_c)]$ and the vertices are first visited in order. We denote by $\mathcal{T}_{\ell,m,k}(v, e)$ the set of canonical paths in $T_{\ell,m,k}$ with v vertices and e edges. Then:

$$\begin{aligned} &\mathbb{E} \left[\|R_k^{(\ell-1)}\|^{2m} \right] \\ &\leq c_1^m \sum_{v=1}^{m(2\ell-2-k)} \sum_{e=v}^{m(2\ell-2-k)} \sum_{\gamma_c \in \mathcal{T}_{\ell,m,k}(v,e)} \sum_{\tau \in I_{\gamma_c}} \mathbb{E} \left[\prod_{e \in E(\gamma_c)} A_{\tau(e_1)\tau(e_2)}^{p_{e_1 e_2}^{(\gamma_c)}} A_{\tau(e_1)\tau(e_2)}^{p_{e_1 e_2}^{(\gamma_c)}} \right], \end{aligned} \quad (6.142)$$

where I_{γ_c} is defined as above, $p_{e_1 e_2}^{(\gamma_c)}$ is the number of times edge $\{e_1, e_2\}$ occurs in $\{\{\gamma_{j,s-1}, \gamma_{j,s}\}\}_{s=1, j=1}^{s=k, j=2m}$ and $p_{e_1 e_2}^{(\gamma_c)}$ denotes the number of times edge $\{e_1, e_2\}$ occurs in the remainder of the collection of edges, $\{\{\gamma_{j,s-1}, \gamma_{j,s}\}\}_{s=k+2, j=1}^{s=\ell, j=2m}$.

Now, again,

$$\mathbb{E} \left[\underline{A}_{\tau(e_1)\tau(e_2)}^{p_{e_1e_2}^{(\gamma_c)}} \underline{A}_{\tau(e_1)\tau(e_2)}^{p_{e_1e_2}^{(\gamma_c)}} \right] \leq \phi_{\tau(e_1)} \phi_{\tau(e_2)} \frac{W_{\sigma(\tau(e_1))\sigma(\tau(e_2))}}{n}.$$

Below we construct a spanning forest $F = (V(\gamma), E_F(\gamma))$ of γ (i.e., F is the disjoint union of trees, each spanning another component of $G(\gamma)$).

Let $n_c \leq m$ denote the number of components of $G(\gamma)$. Then,

$$\mathbb{E} \left[\prod_{e \in E(\gamma)} \underline{A}_{e_1e_2}^{p_{e_1e_2}^{(\gamma)}} \underline{A}_{e_1e_2}^{p_{e_1e_2}^{(\gamma)}} \right] \leq (c/n)^{e-(v-n_c)} \prod_{u \in V(\gamma)} \Phi^{(d_u)} \prod_{e \in E_F(\gamma)} \frac{W_{\sigma(e_1)\sigma(e_2)}}{n}, \quad (6.143)$$

with d_u the degree of vertex u in the forest F , compare to (6.130).

Now, this time,

Lemma 6.15.4. *For any canonical path $\gamma_c \in \mathcal{T}_{\ell,m,k}(v, e)$,*

$$\sum_{\tau \in I_{\gamma_c}} \prod_{e \in E_F(\gamma_c)} \frac{W_{\sigma(\tau(e_1))\sigma(\tau(e_2))}}{n} \leq (1 + o(1)) n^{n_c} \left(\frac{a+b}{2} \right)^{v-n_c}. \quad (6.144)$$

Proof. Apply Lemma 6.15.1 subsequently to the different components of F . \square

Further, applying (6.134) to different components in F gives

$$\prod_{u=1}^{v(\gamma)} \Phi^{(d_u)} \leq C_2^{\sum_{u:d_u>2} (d_u-2)} \left(\Phi^{(2)} \right)^{v-n_c}. \quad (6.145)$$

Together,

$$\sum_{\tau \in I_{\gamma_c}} \mathbb{E} \left[\prod_{e \in E(\gamma_c)} \underline{A}_{\tau(e_1)\tau(e_2)}^{p_{e_1e_2}^{(\gamma_c)}} \underline{A}_{\tau(e_1)\tau(e_2)}^{p_{e_1e_2}^{(\gamma_c)}} \right] \leq (c/n)^{e-v} C_2^{\sum_{u:d_u>2} (d_u-2)} \rho^{v-n_c}. \quad (6.146)$$

Again, we bound $\sum_{u:d_u>2} (d_u-2)$:

Lemma 6.15.5. *For any $\gamma \in \mathcal{T}_{\ell,m,k}$, with v vertices and e edges, there exists a forest spanning γ with degrees $(d_u)_{u=1}^v$ such that:*

$$\sum_{u:d_u>2} (d_u-2) \leq 18m + e - (v - n_c). \quad (6.147)$$

Proof. As in Lemma 6.15.2, we construct the spanning forest, while traversing γ . Again $p(t)$ denotes the graph constructed at step $t \geq 0$, with $p(0) = \{\gamma_{1,0}, \emptyset\}$. Further, we introduce three counters: $r = s = q = 0$, together with $e_c = \emptyset$ (below, e_c is either equal to \emptyset or it is an edge such that $p(t)$ contains one cycle, but $p(t) \setminus e_c$ is a forest). At any step t , we let $C_1, \dots, C_{\#\text{components}}$ be the components of $p(t)$.

Consider step $t+1$ of the walk: if the step consists in jumping to a vertex w , then put $p(t+1) = (p(t) \setminus e_c) \cup \{w\}$.

Else, if the step consists in traversing an edge $f = f_1 f_2$, then: If f or \check{f} has already been traversed, continue with step $t+2$. Otherwise, if both f and \check{f} have not yet been traversed, distinguish between the following cases:

1. f_1 is a leaf or an isolated vertex of component C_i of $p(t)$ and
 - a. C_i does not contain a cycle, then put $p(t+1) = p(t) \cup f$. Further, distinguish between the following cases:
 - i) $f_2 \notin p(t)$;

- ii) $f_2 \in C_i$, then put $e_c = f$;
 - iii) $f_2 \in C_{j \neq i}$, then increase the value of s with one.
- b. C_i contains a cycle, then distinguish between the following cases:
- i) $f_2 \notin p(t)$, then put $p(t+1) = p(t) \cup f$;
 - ii) $f_2 \in C_i$, then put $p(t+1) = p(t)$;
 - iii) $f_2 \in C_{j \neq i}$, then put $p(t+1) = p(t) \cup f$ and increase the value of s with one.
2. f_1 in component C_i has degree at least 2 in $p(t)$, then distinguish between the following cases:
- a. C_i does not contain a cycle, then put $p(t+1) = p(t) \cup f$. Further, distinguish between the following cases:
- i) $f_2 \notin p(t)$, then increase the value of q with one;
 - ii) $f_2 \in C_i$, then put $e_c = f$;
 - iii) $f_2 \in C_{j \neq i}$, then increase the value of s with two.
- b. C_i contains a cycle, then put $p(t+1) = (p(t) \setminus e_c) \cup f$. Further, distinguish between the following cases:
- i) $f_2 \notin p(t)$, then increase the value of r with one;
 - ii) $f_2 \in C_i$, then put $e_c = f$;
 - iii) $f_2 \in C_{j \neq i}$, then increase the value of s with two.

Once the path is completely traversed, remove e_c to obtain a spanning tree.

The only cases that contribute to $\sum_{u:d_u > 2} (d_u - 2)$ are 1.a.iii, 1.b.iii, 2.a.i, 2.a.iii, 2.b.i and 2.b.iii.

Now, s counts the contribution of 1.a.iii, 1.b.iii, 2.a.iii and 2.b.iii. But, in all those 4 cases, two components are merged, hence $s \leq 6 \# \text{merges} \leq 12m$.

By definition of the event 2.b.i, r is an upper bound for the number of edges that are removed: $r \leq e - (v - n_C)$.

To bound q (which counts the occurrence of 2.a.i), note that between two subsequent occurrences of the event 2.a.i, the walk makes at least one of the following: a backtrack, a jump or a merge. Hence $q \leq 2m + 2m + 2m = 6m$.

Adding the bounds for r, q and s establishes (6.147). \square

Returning to (6.146), we get, since $n_C \leq 2m$:

$$\begin{aligned} \sum_{\tau \in I_{\gamma_c}} \mathbb{E} \left[\prod_{e \in E(\gamma_c)} \frac{p(\gamma_c)}{A_{\tau(e_1)\tau(e_2)}^{e_1 e_2}} A_{\tau(e_1)\tau(e_2)}^{p(\gamma_c)} \right] &\leq (c_1/n)^{e-v} C_2^{18m+e-v+n_C} \rho^{v-n_C} \\ &\leq \left(\frac{c_3}{n}\right)^{e-v} c_4^m \rho^{v-n_C}. \end{aligned} \quad (6.148)$$

Putting this into (6.142), we obtain

$$\mathbb{E} \left[\|R_k^{(\ell-1)}\|^{2m} \right] \leq c_5^m \sum_{v=1}^{m(2\ell-2-k)} \sum_{e=v}^{m(2\ell-2-k)} \sum_{\gamma_c \in \mathcal{T}_{\ell,m,k}(v,e)} \left(\frac{c_3}{n}\right)^{e-v} \rho^{m(2\ell-k)}. \quad (6.149)$$

Now the cardinality of $\mathcal{T}_{\ell,m,k}(v, e)$ is bounded in the following lemma:

Lemma 6.15.6 (Lemma 18 in [17]). *Let $\mathcal{T}_{\ell,m,k}(v, e)$ be the set of canonical paths in $\mathcal{T}_{\ell,m,k}$ with $v(\gamma) = v$ and $e(\gamma) = e$. We have*

$$|\mathcal{T}_{\ell,m,k}(v, e)| \leq (4\ell m)^{12m(e-v+1)+8m}.$$

Hence,

$$\begin{aligned}
\mathbb{E} \left[\|R_k^{(\ell-1)}\|^{2m} \right] &\leq c_5^m \rho^{m(2\ell-k)} \sum_{v=1}^{m(2\ell-2-k)} \sum_{e=v}^{m(2\ell-2-k)} (4\ell m)^{12m(e-v+1)+8m} \left(\frac{c_3}{n}\right)^{e-v} \\
&\leq \rho^{m(2\ell-k)} c_5^m (4\ell m)^{20m} \sum_{v=1}^{m(2\ell-2-k)} \sum_{s=0}^{\infty} \left(\frac{c_3(4\ell m)^{12m}}{n}\right)^s \\
&\leq \rho^{m(2\ell-k)} c_5^m (4\ell m)^{20m} 2\ell m \cdot \mathcal{O}(1) \\
&\leq \rho^{m(2\ell-k)} (c_5 \log(n))^{42m}.
\end{aligned} \tag{6.150}$$

We used that, due to our choice of m , $(4\ell m)^{12m} \leq n^{24/25}$.

We use (6.150) together with Markov's inequality:

$$\begin{aligned}
\mathbb{P} \left(\|R_k^{(\ell)}\| > (\log(n))^{25} \rho^{\ell-k/2} \right) &\leq \frac{\mathbb{E} \left[\|R_k^{(\ell)}\|^{2m} \right]}{(\log(n))^{50m} \rho^{m(2\ell-k)}} \\
&\leq (c_6 \log(n))^{-8m} \rightarrow 0.
\end{aligned} \tag{6.151}$$

6.15.4 Bound $\|KB^{(k)}\|$

Put

$$m = \left\lfloor \frac{\log n}{13 \log(\log n)} \right\rfloor. \tag{6.152}$$

We have, with the convention that $e_{2m+1} = e_1$,

$$\begin{aligned}
\|KB^{(k-2)}\|^{2m} &\leq \text{Tr} \left\{ \left(KB^{(k-2)} KB^{(k-2)*} \right)^m \right\} \\
&= \sum_{e_1, \dots, e_{2m}} \prod_{i=1}^m (KB^{(k-2)})_{e_{2i-1}, e_{2i}} (KB^{(k-2)})_{e_{2i+1}, e_{2i}}.
\end{aligned} \tag{6.153}$$

Now,

$$\begin{aligned}
\left(KB^{(k-2)} \right)_{ef} &= \sum_g K_{eg} B_{gf}^{(k-2)} \\
&= \sum_g 1_{e \rightarrow g} \phi_{e_1} \phi_{e_2} W_{\sigma(e_1)\sigma(e_2)} \sum_{\gamma \in F_{gf}^{k-1}} \prod_{s=0}^{k-2} A_{\gamma_s \gamma_{s+1}} \\
&\leq c_1 \sum_g 1_{e \rightarrow g} \sum_{\gamma \in F_{gf}^{k-1}} \prod_{s=0}^{k-2} A_{\gamma_s \gamma_{s+1}}.
\end{aligned} \tag{6.154}$$

Hence,

$$\begin{aligned}
&\|KB^{(k-2)}\|^{2m} \\
&\leq c_2^m \sum_{e_1, \dots, e_{2m}} \prod_{i=1}^m \left(\sum_g 1_{e_{2i-1} \rightarrow g} \sum_{\gamma \in F_{ge_{2i}}^{k-1}} \prod_{s=0}^{k-2} A_{\gamma_s \gamma_{s+1}} \right) \left(\sum_g 1_{e_{2i+1} \rightarrow g} \sum_{\gamma \in F_{ge_{2i}}^{k-1}} \prod_{s=0}^{k-2} A_{\gamma_s \gamma_{s+1}} \right) \\
&= c_2^m \sum_{\gamma \in \bar{W}_{k,m}} \prod_{i=1}^m \prod_{s=2}^k A_{\gamma_{2i-1, s-1} \gamma_{2i-1, s}} \prod_{s=1}^{k-1} A_{\gamma_{2i, s-1} \gamma_{2i, s}},
\end{aligned} \tag{6.155}$$

where $\overline{W}_{k,m}$ is the collection containing all sequences of paths $\gamma = (\gamma_1, \dots, \gamma_{2m})$ with $\gamma_i = (\gamma_{i,0}, \dots, \gamma_{i,k}) \in V^{k+1}$ is non-backtracking such that

- for all i : $(\gamma_{i,k-1}, \gamma_{i,k}) = (\gamma_{i+1,1}, \gamma_{i+1,0})$,
- for all *odd* i : $(\gamma_{i,1}, \dots, \gamma_{i,k})$ is tangle-free,
- for all *even* i : $(\gamma_{i,0}, \dots, \gamma_{i,k-1})$ is tangle-free,

with the convention that $\gamma_{2m+1} = \gamma_1$.

Recall the definition of $W_{k,m}$ and note that $W_{k,m} \subset \overline{W}_{k,m}$. Fix $\overline{\gamma} \in \overline{W}_{k,m} \setminus W_{k,m}$ and let $S_{\overline{\gamma}}$ be the set of all $\hat{\gamma} \in \overline{W}_{k,m} \setminus W_{k,m}$ such that for all odd i : $(\hat{\gamma}_{i,1}, \dots, \hat{\gamma}_{i,k}) = (\overline{\gamma}_{i,1}, \dots, \overline{\gamma}_{i,k})$ and for all even i : $(\hat{\gamma}_{i,0}, \dots, \hat{\gamma}_{i,k-1}) = (\overline{\gamma}_{i,0}, \dots, \overline{\gamma}_{i,k-1})$. Then $|S_{\overline{\gamma}}| \leq k^m$. Indeed, if for odd i , $\hat{\gamma}_i$ is *not* tangle-free then necessarily $\hat{\gamma}_{i,0} \in \{\hat{\gamma}_{i,1}, \dots, \hat{\gamma}_{i,k}\}$, i.e., $\hat{\gamma}_{i,0}$ can be chosen in at most k different ways. A similar argument works in case i is even.

Now, there always exists $\gamma \in W_{k,m}$ such that for all odd i : $(\gamma_{i,1}, \dots, \gamma_{i,k}) = (\overline{\gamma}_{i,1}, \dots, \overline{\gamma}_{i,k})$ and for all even i : $(\gamma_{i,0}, \dots, \gamma_{i,k-1}) = (\overline{\gamma}_{i,0}, \dots, \overline{\gamma}_{i,k-1})$.

As a consequence of these two observations, we have

$$\|KB^{(k-2)}\|^{2m} \leq c_2^m (1+k^m) \sum_{\gamma \in W_{k,m}} \prod_{i=1}^m \prod_{s=2}^k A_{\gamma_{2i-1,s-1}\gamma_{2i-1,s}} \prod_{s=1}^{k-1} A_{\gamma_{2i,s-1}\gamma_{2i,s}}. \quad (6.156)$$

To proceed following the method used to bound $\Delta^{(k)}$, note that the product in (6.156) is taken over a path, consisting of $2m$ non-backtracking tangle-free subpaths of length $k-1$, that makes at most $2m$ backtracks. Hence Lemma's 6.15.1 and 6.15.2 may be adapted to the current setting (for instance the right hand side of (6.135) becomes $e - (v-m-1) + 2m$), entailing

$$\begin{aligned} & \mathbb{E} \left[\|KB^{(k-2)}\|^{2m} \right] \\ & \leq c_2^m (1+k^m) \sum_{v=3}^{2km+1} \sum_{e=v-1}^{2km} |\mathcal{W}_{k,m}| \left(\frac{c_3}{n}\right)^{e-(v-1)-m} c_4^{e-(v-m-1)+2m} n \rho^{v-1} \\ & \leq c_5^m (1+k^m) n^{m+1} \sum_{v=3}^{2km+1} \sum_{e=v-1}^{2km} |\mathcal{W}_{k,m}| \left(\frac{c_6}{n}\right)^{e-(v-1)} \rho^{v-1} \\ & \leq c_7^m (1+k^m) n^{m+1} \rho^{2km} \ell^{2m} \ell m \sum_{s=0}^{\infty} \left(\frac{c_6(2\ell m)^{6m}}{n}\right)^s \\ & \leq c_8^m (\ell m)^2 \ell^{3m} n^{m+1} \rho^{2km} \\ & \leq (c_9 \log n)^{19m} n^m \rho^{2km}, \end{aligned} \quad (6.157)$$

where we used our choice for m several times. An appeal to Markov's inequality finishes the proof.

6.15.5 Bound on $\|S_k^{(l)}\|$

This proof follows almost line-to-line the proof used in [17] to establish bound (34) there. We restrict ourselves here to the differences:

Observe that $L_{ef} = 0$ unless $e \xrightarrow{2} f$ does not hold, that is $e = f$, $e \rightarrow f$, $f^{-1} \rightarrow e$ or $e \rightarrow f^{-1}$, in which cases $L_{ef} = -\phi_{e_2} \phi_{f_1} W_{\sigma(e_2)\sigma(f_1)}$. Hence, we have the decomposition

$$L = -I^* - K^*,$$

where $(I^*)_{ef} = 1_{e=f}\phi_{e_1}\phi_{e_2}W_{\sigma(e_1)\sigma(e_2)}$, and where $(K^*)_{ef} = \phi_{e_2}\phi_{f_1}W_{\sigma(e_2)\sigma(f_1)}$ if $e \rightarrow f$, $f^{-1} \rightarrow e$ or $e \rightarrow f^{-1}$ and $(K^*)_{ef} = 0$ otherwise.

Thus

$$\|S_k^{(\ell)}\| \leq \phi_{\max}^2(a \vee b) \left(\|\Delta^{(k-1)}\| \|B^{(\ell-k-1)}\| + \|\Delta^{(\ell-1)}K'\| \|B^{(\ell-k-1)}\| \right),$$

where K' is defined in [17]. The rest of the proof follows after applying the arguments used in [17] and following the procedure set out above to obtain the bound on $KB^{(k)}$.

6.16 Proofs of Section 6.11

Proof of Lemma 6.11.1. Since $\hat{\sigma}(v) = +$ if and only if $F(v) = 1$, it follows that

$$\frac{1}{n} \sum_{v=1}^n 1_{\sigma(v)=+} 1_{\hat{\sigma}(v)=\sigma(v)} = \frac{1}{n} \sum_{v=1}^n 1_{\sigma(v)=+} F(v) \rightarrow \frac{f(+)}{2},$$

and

$$\frac{1}{n} \sum_{v=1}^n 1_{\sigma(v)=-} 1_{\hat{\sigma}(v)=\sigma(v)} = \frac{1}{n} \sum_{v=1}^n 1_{\sigma(v)=-} (1 - F(v)) \rightarrow \frac{1 - f(-)}{2}.$$

Consequently,

$$\frac{1}{n} \sum_{v=1}^n 1_{\hat{\sigma}(v)=\sigma(v)} \rightarrow \frac{1 + f(+)}{2} - \frac{f(-)}{2} > \frac{1}{2},$$

because $f(+)$ is greater than $f(-)$ by assumption. \square

Proof of Lemma 6.11.2. We use Proposition 6.8.2 with

$$\tau(G, v) = 1_{\sigma(v)=i} 1_{I_\ell(v)\mu_2^{-2\ell} - \hat{c}g_2(i) \geq t}.$$

Denote by (T, o) the branching process defined in Section 6.6 where the root has spin σ_o uniformly drawn from $\{+, -\}$. Denote the number of offspring of the root by D and let $Q_\ell(v)$ be equal to $Q_{2,\ell}$ defined on the tree T^v obtained after removing the subtree attached to v from T . Then,

$$\tau(T, o) = 1_{\sigma_o=i} 1_{J_\ell\mu_2^{-2\ell} - \hat{c}g_2(i) \geq t},$$

where

$$J_\ell = \sum_{v=1}^D Q_\ell(v) = (D-1)Q_{2,\ell} - L_{2,\ell}^o, \quad (6.158)$$

with $L_{2,\ell}^o$ defined in (6.34).

We need to calculate $\lim_{\ell \rightarrow \infty} \mathbb{E}[\tau(T, o)]$. To this end, we first show that, conditional on $\sigma_o = i$, $\frac{J_\ell}{\mu_2^{2\ell}} - \hat{c}g_2(i)$ converges in probability to some *centered* random variable \widehat{Y}_i .

We first calculate $\mathbb{E}_i[J_\ell | \phi_o]$, where $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot | \sigma_o = i]$. Put $r_o = \frac{a+b}{2}\Phi^{(1)}\phi_o$, then

$$\begin{aligned} \mathbb{E}_i[J_\ell | \phi_o] &= \sum_{n=0}^{\infty} \mathbb{E}_i[J_\ell | D = n, \phi_o] \mathbb{P}(D = n | \phi_o) \\ &= \sum_{n=0}^{\infty} n \mathbb{E}_i[Q_{2,\ell} | D = n-1, \phi_o] \frac{r_o^n e^{-r_o}}{n!} \\ &= r_o \sum_{n=1}^{\infty} \mathbb{E}_i[Q_{2,\ell} | D = n-1, \phi_o] \frac{r_o^{n-1} e^{-r_o}}{(n-1)!} \\ &= r_o \mathbb{E}_i[Q_{2,\ell} | \phi_o]. \end{aligned} \quad (6.159)$$

Recall from Theorem 6.6.7 that, uniformly for all ψ_o ,

$$\mathbb{E}_i \left[\frac{Q_{2,\ell}}{\mu_2^{2\ell}} \middle| \phi_o = \psi_o \right] \rightarrow \frac{\Phi^{(3)}}{\Phi^{(2)}} \frac{\rho}{\mu_2^2 - \rho} \mu_{2,\psi_o} g_2(i)$$

as $n \rightarrow \infty$. Hence, $\sup_{n,\psi_o} \mathbb{E}_i \left[\frac{Q_{2,\ell}}{\mu_2^{2\ell}} \middle| \phi_o = \psi_o \right] < \infty$, so that we can apply Lebesgue's dominated convergence theorem:

$$\frac{\mathbb{E}_i [J_{2,\ell}]}{\mu_2^{2\ell}} = \mathbb{E}_i \left[r_o \mathbb{E}_i \left[\frac{Q_{2,\ell}}{\mu_2^{2\ell}} \middle| \phi_o \right] \right] \rightarrow \widehat{c}g_2(i), \quad (6.160)$$

as $n \rightarrow \infty$.

We now combine the right hand side of (6.158), (6.160), and Theorem 6.6.7 (and in particular (6.111) which implies that $L_{2,\ell}^o/\mu_2^{2\ell} \rightarrow 0$ as $n \rightarrow \infty$) to establish the claim that, conditional on $\sigma_o = i$, $\frac{J_\ell}{\mu_2^{2\ell}} - \widehat{c}g_k(i)$ converges in probability to some *centered* random variable \widehat{Y}_i .

In particular, conditional on $\sigma_o = i$, $\frac{J_\ell}{\mu_2^{2\ell}} - \widehat{c}g_2(i)$ converges in distribution to \widehat{Y}_i . So that, for t as in the statement,

$$\mathbb{E} [\tau(T, o)] = \frac{1}{2} \mathbb{P} \left(\frac{J_\ell}{\mu_2^{2\ell}} - \widehat{c}g_2(i) \geq t \middle| \sigma_o = i \right) \rightarrow \frac{1}{2} \mathbb{P} \left(\widehat{Y}_i \geq t \right),$$

as $n \rightarrow \infty$.

Finally, noting that the error term in Proposition 6.8.2 is $\mathcal{O} \left(n^{-\left(\frac{\gamma}{2} \wedge \frac{1}{40}\right)} \right) = o(1)$ finishes the proof. \square

Proof of Lemma 6.11.3. This follows after repeating the proof in [17] in conjunction with Lemma 6.11.2 established here. \square

Chapter 7

Tests on real networks

7.1 Normalized Adjacency Matrix

We have tested our method from Chapter 4 on three real networks, namely, Zachary's karate club [124], the dolphin social network [81] and the political blogs dataset [3], see Section 1.4. The error rate for Zachary's karate club is 2/34 and for the dolphin social network 0/62.

The error rate for the political blogs dataset is 230/1221 when thresholding the Frobenius eigenvector. We restricted to the giant component of 1221 nodes, as is common in most other works (the original data contained 1490 blogs). Our clustering is worse than obtained by SCORE (where the error rate is 58/1221), but similar to the non-backtracking matrix (where around 15 percent of the nodes are misclassified [72]).

We observed that the leading eigenvectors are concentrated on a few nodes, due to the presence of certain problematic structures (such as two low-degreed vertices connected by an edge). However, the value of the Frobenius eigenvector on the remaining vertices is still correlated with their community-membership as can be observed in Figures 7.1 and 7.2.

Figure 7.1 is a histogram of the Frobenius eigenvector restricted to the roughly 600 nodes that have corresponding value in the interval $[0, 10^{-9}]$. The nodes seem to concentrate around two centres according to their community. However, this phenomenon is only weakly visible (note that our theory does not apply for *sparse* graphs).

In Figure 7.2 we have sorted the 1221 indices of the Frobenius eigenvector according to an increasing corresponding value: the community structure becomes then clear.

We further observed that thresholding the eight-est eigenvector leads to only 160 misclassified vertices. Interestingly, if we inflate the degrees by replacing $H = \frac{A_{uv}}{\overline{D}_u \overline{D}_v}$ by $H_{\text{inflated}} = \frac{A_{uv}}{\max\{\overline{D}_u \overline{D}_v, 200\}}$, we obtain an error rate of 74/1221 by thresholding its second eigenvector. This suggests that initial misclassifications are indeed due to low degree nodes (the average degree is 27, but there are also many leafs present).

7.2 Non-backtracking Matrix

In Figure 7.3, the non-backtracking spectrum of the political blogs graph is shown. There are two eigenvalues that are clearly separated from the bulk of eigenvalues. However some other outliers are also present, indicating that the political blogs graph has in fact more than two communities - corresponding to the intuition that the orientation of a political blog is in reality often in between the two extremes.

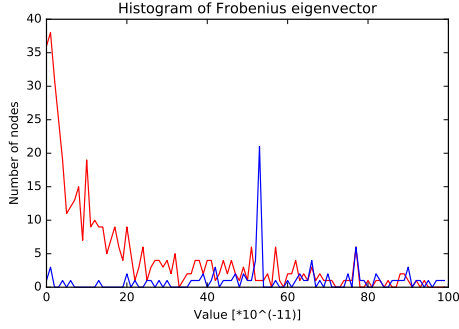


Figure 7.1: Histogram of the Frobenius eigenvector restricted to the roughly 600 nodes that have corresponding value in the interval $[0, 10^{-9}]$. The colors represent the communities.

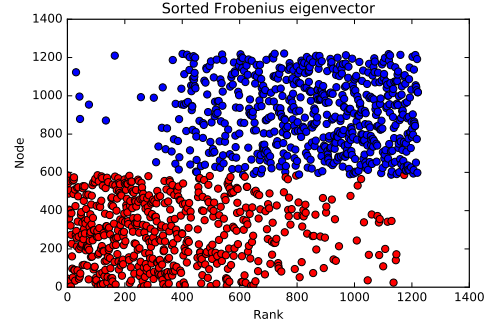


Figure 7.2: Ranking of the 1221 indices of the Frobenius eigenvector according to an increasing corresponding value. Rank 1 is the node with smallest value in the eigenvector and rank 1221 the node with largest value. Colors indicate community-membership.

We have fitted (in a crude way) a DC-SBM to the political blogs graph as follows: We estimated ρ and μ to be the two largest eigenvalues in the spectrum in Figure 7.3 and put $a = 1$ and $b = \frac{\rho - \mu}{\rho + \mu}$. Further, for every vertex u , we put σ_u to be the hand-labelled political orientation (i.e., the estimated ground-truth) and put $\phi_u = \hat{D}_u / \sqrt{\frac{\rho}{\rho + \mu} \bar{D}}$, with \hat{D}_u the observed degree of vertex u and \bar{D} the average observed degree in the graph. We plotted in Figure 7.4 the non-backtracking spectrum of a typical realisation of this DC-SBM.

For comparison, we plotted in Figure 7.5 the spectrum of a DC-SBM where $a = 1, b = 0.1098$ and where the weights are drawn uniformly from the interval $[5, 30]$. The spectrum behaves as expected (see the theory in the previous chapter).

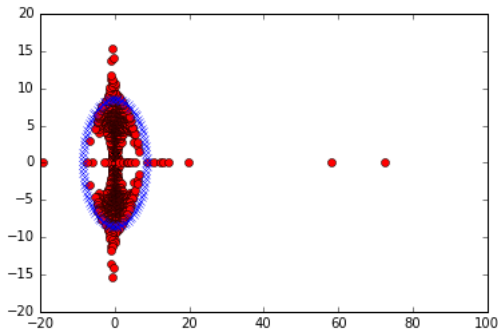


Figure 7.3: Non-backtracking spectrum of the political blogs graph.

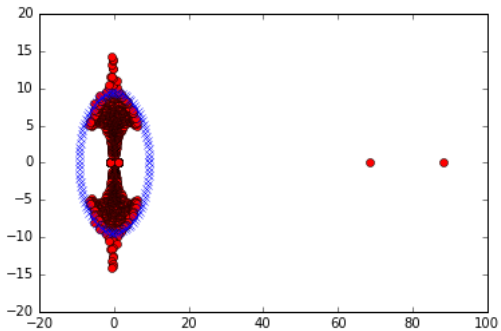


Figure 7.4: Non-backtracking spectrum of a DC-SBM fit to the political blogs graph. Parameter estimation: $a = 1, b = \frac{\rho - \mu}{\rho + \mu}$. $D_u \simeq \phi_u \sqrt{\frac{\rho}{\rho + \mu} \bar{D}}$.

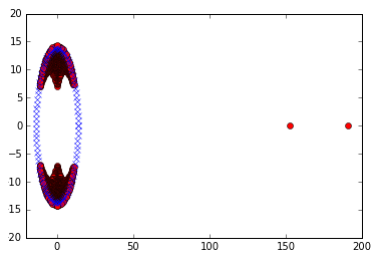


Figure 7.5: Non-backtracking spectrum of a DC-SBM with 1500 nodes and parameters $a = 1, b = 0.1098$, where the weights are drawn uniformly from the interval $[5, 30]$.

Part II

Adaptive Matching

Chapter 8

Adaptive Matching for Expert Systems with Uncertain Task Types

Acknowledgement: This part contains joint work with Laurent Massoulié, Milan Vojnović, Nidhi Hegde and, most notably, **Virag Shah**. Further, Section 8.8.1 is inspired by a discussion of [104] with Anastasia Podosinnikova.

Parts of this chapter will appear in the conference proceedings of Allerton 2017 [112].

8.1 Introduction

Online platforms that enable matches between trading partners in two-sided markets have recently blossomed in many areas: LinkedIn and Upwork facilitate matches between employers and employees; Uber allows matches between passengers and car drivers; Airbnb and Booking.com connect travelers and housing facilities; Quora and Stack Exchange facilitate matches between questions and either answers, or experts able to provide them.

All these systems crucially rely on the ability to propose adequate matches based on imperfect knowledge of the characteristics of the two parties to be matched. For example, in the context of online labour platforms, there is uncertainty about both the skill sets of candidate employees and the job requirements. Similarly, in the context of online Q&A platforms, there is uncertainty about both question types and users' ability to provide answers.

This naturally leads to the following question: *which matching recommendation algorithms can, in the presence of such uncertainty, lead to efficient platform operation?* A natural measure of efficiency is the throughput that the platform achieves, i.e. the rate of successful matches it allows. To address this question, one thus needs first to characterize fundamental limits on the achievable throughput.

In this chapter, we progress towards answering these questions as follows.

First, we propose a simple model of such platforms, which features a static collection of servers, or experts on the one hand, and a continuous stream of arrivals of tasks, or jobs, on the other hand. In our model, the platform's operation consists of servers iteratively attempting to solve tasks. After being processed by some server, a task leaves the system if solved; otherwise it remains till successfully treated by some server. To model uncertainty about task types, we assume that for each incoming task we are given the prior distribution of this task's "true type". Servers' abilities are then represented via the probability that each server has to solve a task of given type after one attempt at it.

In a Q&A platform scenario, tasks are questions, and servers are experts; a server processing a task corresponds to an expert providing an answer to a question. A task being solved corresponds to an answer being accepted. In an online labour platform, tasks could be job offers, and a server may be a pool of workers with similar abilities. A server processing a task then corresponds to a worker being interviewed for a job, and the task is solved if the interview leads to a hire. We could also consider the dual interpretation when the labour market is constrained by workers rather than job offers. Then a task is a worker seeking work, while a server is a pool of employers looking for hires.

An important feature of our model consists in the fact that when a task's processing does not succeed, it does however affect uncertainty about the task's type. Indeed, the a posteriori distribution of the task's type after a failed attempt on it by some server differs from its prior distribution. For instance in a Q&A scenario, a question which an expert in Calculus failed to answer either is not about Calculus, or is very hard.

For our model, we then determine necessary and sufficient conditions for an incoming stream of task arrivals to be manageable by the servers, or in other words, determine achievable throughputs of the system. In the process we introduce candidate policies, in particular the greedy policy according to which a server chooses to serve tasks for which its chance of success is highest. This scheduling strategy is both easy to implement and is based on a natural motivation. Surprisingly perhaps, we show that it is not optimal in the throughput it can handle. In contrast, we introduce a so-called backpressure policy inspired from the wireless networking literature [114], which we prove to be throughput-optimal.

We summarize contributions of this chapter as follows:

- We propose a new model of a generic task-expert system that allows for uncertainty of task types, heterogeneity of skills, and recurring attempts of experts in solving tasks.
- We provide a full characterization of the stability region, or sustainable throughputs, of the task-expert system under consideration. We establish that a particular backpressure policy is throughput-optimal, in the sense that it supports maximum task arrival rate under which the system is stable.
- We show that there exist instances of task-expert systems under which simple matching policies such as a natural greedy policy and a random policy can only support a much smaller maximum task arrival rate, than the backpressure policy.
- We report the results of empirical analysis of the popular Math.StackExchange Q&A platform which establish heterogeneity of skills of experts, with experts knowledgeable across different types of tasks and others specialized in particular types of tasks. We also show numerical evaluation results that confirm the benefits of the backpressure policy on greedy and random matchmaking policies.

The remainder of the chapter is structured as follows. Section 8.2 presents our system model. In Section 8.3, we present results for two baseline matchmaking policies, namely Greedy and Random. Section 8.4 presents the characterization of task arrival rates that can be supported under which the system is stable and prove the superiority of backpressure policy over Random and Greedy. In Section 8.5, we present our experimental results. Related work is discussed in Section 8.6. We conclude in Section 8.7. Further, in Section 8.8, we give some directions to obtain the prior distribution of a task's true type and the server's abilities. Proofs of the results are provided in Section 8.9.

8.2 Problem Setting

Let $C = \{c_1, \dots, c_k\}$ be the set of task types. Each task in the system is of a particular type in C . Let $S = \{s_1, \dots, s_m\}$ be the set of servers (or experts) present in the system. When a server $s \in S$ attempts to resolve a task of type $c \in C$, the outcome is 1 (a success) with probability $p_{s,c}$ and it is 0 (a failure) with probability $1 - p_{s,c}$. Upon success we say that the task is resolved. In the context of online hiring platform, this is equivalent to successful hiring of an employee for a job. In the context of Q&A platform, this is equivalent to an answer by an expert being accepted by the asker of the question.

We consider a Bayesian setting where we have a prior distribution $z = \{z_c\}_{c \in C} \in \mathcal{C}$ for a task's type, where \mathcal{C} is the set of all distributions. Note, different tasks may have different prior distributions. Clearly, if server s processes a task with prior distribution z then the probability that it fails is given by

$$\psi_s(z) = \sum_{c \in C} z_c(1 - p_{s,c}). \quad (8.1)$$

Further, upon failure, the posterior distribution of task's type is given by

$$\phi_s(z) = \left\{ \frac{z_c(1 - p_{s,c})}{\psi_s(z)} \right\}_{c \in C}. \quad (8.2)$$

Note that the posterior distribution of a task's type upon failure by a subset of servers does not depend on the sequence in which these servers resolve the task, i.e., for each $s, s' \in S$ we have $\phi_s \circ \phi_{s'} = \phi_{s'} \circ \phi_s$. At any point in time a task is associated with a 'mixed-type' which is defined as the posterior distribution of its type given the past attempts.

We allow a task to be attempted sequentially by multiple servers until it is resolved. We would like to resolve the tasks as quickly as possible.

8.2.1 Single Task Scenario

Before considering the setting of online task arrivals, for ease of exposition we first consider a toy scenario with single task. Suppose that time $t \in \mathbb{Z}_+$ is discrete. A task arrives at time $t = 0$. Let the prior distribution of its type upon arrival (equivalently, its mixed-type at time $t = 0$) be z . At a time, only one server attempts to resolve a task. Consider the problem of designing a sequence of servers $(s(t) : 0 \leq t \leq \tau)$ such that the probability that the task is resolved within a fixed time τ is maximized. Let $z(0) = z$, and for each $t \geq 1$ let $z(t) = \phi_{s(t-1)}(z(t-1))$, i.e., $z(t)$ is the mixed-type of the task at time t given that it was not resolved upon previous attempts. Then the probability that the task is resolved by time τ is given as $f(\tau) = 1 - \prod_{t=0}^{\tau-1} \psi_{s(t)}(z(t))$

Contrast this with the Bayesian active learning setting in [49, 62] where the goal is to reduce uncertainty in true hypothesis via outcome from several experiments. Using a diminishing returns property called adaptive submodularity the authors in [49] obtain a policy which is competitive with the optimal. In our setting, $f(\tau)$ is a submodular function. Thus a greedy policy where $s(t)$ for each t is chosen to be from $\arg \min_s \psi_s(z(t))$ is $1 - 1/e$ -competitive [100].

The feedback assumed in [49] is more general than the binary feedback assumed in our case. However we keep the binary feedback assumption since for hiring platforms the outcome of a job-employee match is binary and is notified to the platform by the agents involved. Further feedback on the performance of the agents may also be reported, but such a feedback is not reliable as it is often biased towards the larger values [31]. Even for a Q&A setting the model is relevant since only one answer from the available responses may be accepted by the asker, such as in StackOverflow.

Further, in this paper we add an extra dimension to the problem which was not considered in the [49, 62], namely, we consider the setting of online task arrivals

where tasks of different mixed-types may compete for the servers resources before they leave upon being resolved. We design throughput optimal policies under such a setting.

8.2.2 Online Task Arrivals

We consider a continuous time setting, i.e., $t \in \mathbb{R}_+$. Tasks arrive at a rate of λ per time unit on average. The subsequent mixed types Z_i of incoming tasks are assumed i.i.d., taking values in a countable subset \mathcal{Z} of \mathcal{C} , and we let $\pi_z = P(Z_i = z)$ for all $z \in \mathcal{Z}$. Finally, the time for server $s \in S$ to complete an attempt on a job takes on average $1/\mu_s$ time units, and such attempt durations are i.i.d.. All involved sources of randomness are independent.

We assume that \mathcal{Z} is closed under $\phi_s(\cdot)$, i.e., for each $z \in \mathcal{Z}$, $\phi_s(z) \in \mathcal{Z}$. This loses no generality, as the closure of a countable set with respect to a finite number of maps ϕ_s remains countable.

We assume that a given task may be inspected several times by a given server and assume that the outcomes success / failure are independent at each inspection. This can be justified if a label s in fact represents a collection of experts with similar abilities, in which case multiple processings by s correspond to processing by distinct individual experts.

For such a setting we would like to minimize the expected sojourn time of a typical task, i.e., the expected time between the arrival and the resolving of a typical task. Recall that the success probabilities $p_{s,c}$ are assumed to be arbitrary. Under such a heterogeneous setting minimizing expected sojourn time is a hard problem. In fact, this is true even when there is no uncertainty in task types. As a proxy to sojourn time optimal policies, we will strive for throughput optimal policies. In particular, we will characterize the arrival rates λ for which the system can be *stabilized*, i.e. for which there exists a scheduling policy which induces a time-stationary regime of the system's behavior. Indeed for a stable system the long term task resolution rate coincides with the task arrival rate λ , and thus throughput-optimal policies must make the system stable whenever this is possible. Note that for an unstable system the number of outstanding tasks accumulate over time and the expected sojourn time tends to infinity.

Finally, for simplicity we assume more specifically that the tasks arrive at the instants of a Poisson process with intensity λ , and that the time for server s to complete an attempt at a task follows an Exponential distribution with parameter μ_s . These are continuous time analog of i.i.d. arrivals and independent departures per time slot in discreet time setting. These assumptions will imply that the system state at any given time t can be represented as a Markov process, which simplifies throughput analysis. The system throughput is often insensitive to such statistical assumptions on arrival and service times [122].

We close the section with some additional assumptions and notations which will aid our analysis.

At any time t let $N_z(t)$ represent the number of tasks of mixed-type z present in the system and $N(t) = \{N_z(t)\}_{z \in \mathcal{Z}}$. We also let $z(s, t)$ denote the mixed type of the task that server s works on at time t . For strategies such that the servers select which task to handle based uniquely on the vector $N(t)$, the process $\{N(t)\}_{t \geq 0}$ forms a continuous-time Markov chain (CTMC) [18, 73]. The policies considered in this paper are studied by analyzing the associated CTMC.

We allow a task to be assigned to multiple experts at a given time. Further, we allow pre-emptive service, i.e., an expert may drop service of a task should a new task arrive into a system or an existing task receive a response.

8.3 Baseline Policies

Following the discussion in Section 8.2.1, could it be that a greedy approach may work well even under the online setting? From throughput perspective, a natural

Greedy policy is one where each expert is assigned a task which best suits its skills, i.e., among the outstanding tasks, an expert s is assigned a task of a mixed-type z which minimizes $\psi_s(z)$. We now show a surprising result that for a class of asymmetric systems the Greedy policy is as suboptimal as a Random policy where each server chooses a task uniformly at random. Let us first formally define these policies.

Definition 8.3.1 (Greedy Policy). *A policy is Greedy if given the system state each expert is assigned an outstanding task which maximizes its success probability, i.e., for each N such that $|N| > 0$ we have*

$$z(s) \in A_s(N) = \arg \min_{z: N_z > 0} \psi_s(z),$$

where ties could be broken uniformly at random among this set.

Definition 8.3.2 (Random Policy). *A policy is Random if each expert s is assigned a task chosen uniformly at random from the pool of outstanding tasks.*

We will consider the following class of task-expert system.

Definition 8.3.3 (Asymmetric(a) System). *Suppose that there are two task types $C = \{c_1, c_2\}$ and two experts $S = \{s_1, s_2\}$. Each arrival is equally likely to be of both types, i.e., $\pi_{z'} = 1$ where z' satisfies $z'_c = 1/2$ for each $c \in C$, and $\pi_z = 0$ if $z \neq z'$. Both experts provide responses at unit rate, i.e., $\mu_s = 1$ for each s . Further, for class c_1 we have $p_{s,c_1} = 1$ for each $s \in S$, and for class c_2 we have $p_{s_1,c_2} = a < 1$, and $p_{s_2,c_2} = 0$. We refer to such a task-expert system as a Asymmetric(a) system with parameter a .*

For this class of systems, if a task of mixed-type z' receives a failure from either of the experts, then its mixed type becomes z'' where $z''_{c_1} = 0$ and $z''_{c_2} = 1$. Thus, it is sufficient to assume that $\mathcal{Z} = \{z', z''\}$ where $z'_c = \frac{1}{2}$ for each $c \in C$, and $z''_c = 1_{\{c=c_2\}}$. Further, it is easy to check that $\psi_{s_1}(z') = (1-a)/2$, $\psi_{s_1}(z'') = 1-a$, $\psi_{s_2}(z') = 1/2$, and $\psi_{s_2}(z'') = 1$. We then have the following result (its proof, as that of the other stability results in the article, is established through the identification of suitable Lyapunov functions, and given in the Section 8.9).

Theorem 8.3.4. *For a given $0 < a < 1$, consider the Asymmetric(a) system as defined in Definition 8.3.3. The following statements hold:*

1. *The Greedy stabilizes the system if and only if $\lambda < 4a/(2+a)$.*
2. *The Random stabilizes the system if and only if $\lambda < 4a/(2+a)$.*
3. *There exists a policy which stabilizes the system if $\lambda < \min\{3a/(a+1), 2a\}$.*
4. *If $\lambda > \min\{3a/(a+1), 2a\}$ then no policy can stabilize the system.*

In particular, the stability threshold for task arrival rates under optimal policy can be up 25% higher (namely, when $a = 1/2$) than that under either Greedy or Random.

While it is intuitive that Random may perform poorly as compared to an optimal policy, it is counterintuitive that Greedy may perform as sub-optimally as Random. The reason for the poor performance of Greedy can be explained as follows. In the Asymmetric(a) system, we have a flexible expert, i.e. an expert for tasks of all pure-types, and a specialized expert, i.e. an expert only for pure-type c_1 . Under Greedy policy, all experts prioritize the newly arriving tasks as it optimizes the probability of achieving success in the short term. However, a larger long-term throughput could be achieved if the flexible expert could focus more on the lagging tasks, i.e., the tasks of pure-type c_2 .

8.4 Optimal Stability

Main goal of this section is to provide necessary and sufficient conditions for stability of the system. We provide a policy, called backpressure policy, which stabilizes the system when the sufficient conditions are satisfied.

We obtain stability conditions via capacity constraints and flow conservation constraints which capture the flow of tasks from one type to another upon service by an expert. For instance, if $\nu_{s,z}$ represents the flow of tasks of mixed-type z served by expert s , a fraction $1 - \psi_s(z)$ of it leaves the system due to success while the rest gets converted into a flow of type $\phi_s(z)$. The total arrival rate of flow of mixed-type z , i.e., $\lambda\pi_z + \sum_{s \in S, z' \in \phi_s^{-1}(z)} \nu_{s,z'} \psi_s(z')$, must match the total service rate, i.e., $\sum_{s \in S} \nu_{s,z}$. Further, the total flow service rate at expert s , i.e., $\sum_{z \in \mathcal{Z}} \nu_{s,z}$, must be less than its service capacity μ_s . The following is the main result of this section. For a proof see Section 8.9.3

Theorem 8.4.1. *Suppose there exists s such that $\min_c p_{s,c} > 0$. If there exist non-negative real numbers $\nu_{s,z}$, for $s \in S$, and $z \in \mathcal{Z}$ and positive real numbers δ_s , for $s \in S$ such that the following hold:*

$$\forall z \in \mathcal{Z}, \quad \lambda\pi_z + \sum_{s \in S, z' \in \phi_s^{-1}(z)} \nu_{s,z'} \psi_s(z') = \sum_{s \in S} \nu_{s,z}, \quad (8.3)$$

$$\forall s \in S, \quad \sum_{z \in \mathcal{Z}} \nu_{s,z} + \delta_s \leq \mu_s, \quad (8.4)$$

then there exists a policy under which the system is stable. If there does not exist non-negative real numbers $\nu_{s,z}$, for $s \in S$, $z \in \mathcal{Z}$ and non-negative real numbers δ_s for $s \in S$ such that the above constraints hold, then the system cannot be stable.

We use the condition of existence of an expert s such that $\min_c p_{s,c} > 0$ only for a technical reason to simplify our proof. We believe that the result holds even when this is not true.

We need some more notation to describe the policy. Consider a set $\mathcal{Y} \subset \mathcal{Z}$. Let $X(t)$ be the number of tasks in the system at time t which have been of type $z \in \mathcal{Z} \setminus \mathcal{Y}$. For $z \in \mathcal{Y}$, let $\tilde{X}_z(t)$ and $\tilde{N}_z(t)$ be the tasks of mixed-type z which have and have not had mixed-type in $\mathcal{Z} \setminus \mathcal{Y}$. Also, for convenience, for each $z \in \mathcal{Z} \setminus \mathcal{Y}$, let \tilde{X}_z be the tasks of mixed-type z , i.e., $N_z = \tilde{X}_z$ for each $z \in \mathcal{Z} \setminus \mathcal{Y}$. Thus, we have $X = \sum_z \tilde{X}_z$. Consider the following policy.

Definition 8.4.2 (Backpressure(\mathcal{Y}) policy). *For a given \mathcal{Y} , let X , and $\{\tilde{N}_z\}_{z \in \mathcal{Y}}$ be as defined above. For each $s \in S$, $z \in \mathcal{Y}$ let*

$$w_{s,z}(\tilde{N}, X) = \begin{cases} \tilde{N}_z - \psi_s(z) \tilde{N}_{\phi_s(z)}, & \text{if } \phi_s(z) \in \mathcal{Y} \\ \tilde{N}_z - \psi_s(z) X, & \text{if } \phi_s(z) \in \mathcal{Z} \setminus \mathcal{Y} \end{cases}.$$

Define

$$B_s(\tilde{N}, X) = \arg \max_{z' \in \mathcal{Y}: \tilde{N}_{z'} > 0} w_{s,z}(\tilde{N}, X).$$

If

$$\sum_s \mu_s \max_{z \in \mathcal{Y}: \tilde{N}_z > 0} w_{s,z}(\tilde{N}, X) \geq X \min_{c \in C} \sum_s \mu_s p_{s,c}$$

then each expert chooses a task in \tilde{N}_z where $z \in B_s(\tilde{N}, X) \subset \mathcal{Y}$ with ties broken arbitrarily. Else, each expert serves a task in X chosen uniformly at random.

The following theorem follows from the proof of Theorem 8.4.1.

Theorem 8.4.3. *Suppose there exists s such that $\min_c p_{s,c} > 0$. If the sufficient conditions for stability as given by Theorem 8.4.1 are satisfied, then there exists a finite subset \mathcal{Y} of \mathcal{Z} such that the policy Backpressure(\mathcal{Y}) stabilizes the system.*

Table 8.1: Skills of experts estimated by using data from the Math.Stack-Exchange Q&A platform.

Tag	Clusters									
	1	2	3	4	5	6	7	8	9	10
calculus	.32	.39	.30	.35	.37	.47	.28	.16	.26	.41
real-analysis	.17	.41	.25	.32	.23	.49	.40	.10	.10	.44
linear-algebra	.46	.29	.05	.36	.14	.48	.26	.31	.07	.43
probability	.07	.49	.02	.33	.02	.50	.06	.02	.46	.04
abstract-algebra	.02	.05	.03	.32	.02	.38	.23	.50	.01	.27
integration	.09	.43	.05	.19	.44	.45	.03	.01	.06	.37
sequences-and-series	.05	.32	.16	.31	.20	.45	.09	.04	.06	.33
general-topology	.02	.10	.03	.16	.02	.43	.50	.07	.02	.31
combinatorics	.03	.14	.06	.43	.04	.37	.02	.06	.19	.05
matrices	.27	.15	.02	.31	.02	.44	.06	.11	.02	.34
complex-analysis	.02	.19	.08	.16	.14	.50	.09	.05	.01	.44
Size	165	188	313	200	179	183	231	187	178	176

In particular, the Backpressure(\mathcal{Z}) policy is optimally stable for Asymmetric(a) system as defined in Definition 8.3.3.

Unlike backpressure policy proposed in [114] under a different setting, which was agnostic to system arrival rates, a set \mathcal{Y} such that Backpressure(\mathcal{Y}) policy stabilizes the system may depend on the value of λ . While the policy as stated may be complex to implement, it allows us to develop implementable heuristics which significantly outperform greedy policy. We demonstrate this in the next section.

8.5 Experimental Results

In this section, we present our empirical results obtained by using data from Math.Stack-Exchange Q&A platform. In this platform, users post tagged questions that are answered by other users. The asker may select one of the submitted answers as the best answer for the given question, which is made public information in the platform. We will use this data to estimate the success probabilities of experts in answering questions, and use these parameters in simulations to compare the throughputs that can be achieved by greedy, random, and backpressure policies. As we will see, a substantially larger throughput can be achieved by backpressure policy than greedy and random.

Dataset The dataset consists of around 702,286 questions and 994,138 answers. It was retrieved on February 2nd, 2017. The top 11 most common tags are given in Table 8.1 in decreasing order of popularity. Among these tags, the most common is *calculus* which covers 61,184 questions, and the least common is *complex analysis* which covers 22,813 questions. In our analysis, we used only questions that are tagged with at least one of the 11 most popular tags, which amounts to a total of 381,239 questions and 544,267 answers.

Estimated skill sets The success probabilities of answering questions are estimated as follows. For a given user-tag pair, the success probability is estimated by the empirical frequency of the accepted answers by this user for questions of given tag, conditional on that the user had at least 5 accepted answers for questions of the given tag, and otherwise we estimate the success probability is set to be equal to zero. These success probabilities are estimated for 2,000 users with the most accepted answers. Among these users, the user with the most accepted answers had 4,665 accepted answers, and the user with the least number of accepted answers had 13 accepted answers. 712 users had more than 50 answers accepted. In order to form clusters of users with similar success probabilities for different tags, we ran the k-means clustering algorithm.

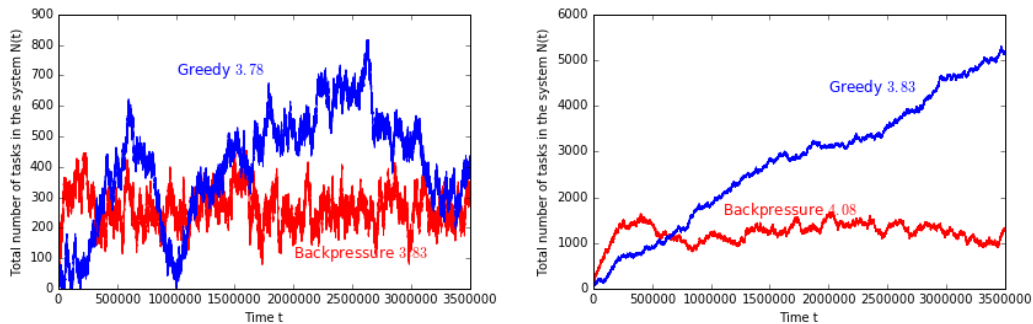


Figure 8.1: Total number of tasks in the system over time for the greedy and backpressure policy. The task arrival rates are as indicated in the figures.

The estimated success probabilities are shown in Table 8.1. The columns correspond to different centroids of the clusters and give average success probabilities for different tags. In the bottom row, we give the sizes of the corresponding clusters. For instance, the 165 persons in cluster 1 have on average 32% of their *calculus*, and 46% of their *linear algebra* answers accepted.

There is a pronounced heterogeneity in user expertise. We highlighted the success probabilities of value at least 35%. A subset of users, namely cluster 6, are good at all topics whereas other clusters are good at a different subsets of topics.

There is also a prevalence of questions with different combinations of tags, that is, mixed types. When a question of mixed type arrives with multiple tags, we associated with it a mixed-type which is the uniform distribution of the associated tags. We kept only those combinations of tags that occur for at least 1% of the total number of questions. This results in 16 tag combinations among which 11 are singletons and 5 are combinations of 2 tags. These are the mixed types z with positive π_z . We observed that roughly 19% of the questions are tagged with multiple tags, showing the relevance of our model.

Simulation setup We assumed the experts to have unit service rates. We make this approximation as we do not have the information about times at which experts begin to respond a question. We examined the system for increasing values of task arrival rates. We simulate our CTMC via a custom discrete event simulator.

For the backpressure policy we define the set \mathcal{Y} to consist of all 11 pure types, the 5 most frequently seen mixed types upon arrival as described above, and the mixed types that result from an attempt by an expert exactly once. Note that pure types can be attempted multiple times without changing its type. We thus have $|\mathcal{Y}| = 16 + 5 \cdot 10 = 66$. Our choice of \mathcal{Y} is a result of a compromise between performance and complexity. Choosing a larger set of \mathcal{Y} may increase the stability region by a small fraction, but may significantly increase the complexity of the Backpressure policy.

Performance comparison of different policies We examined the evolution of the number of tasks in the system waiting to be served over time for greedy and backpressure policy for respective task arrival rates, 3.78 and 3.83 (Figure 8.1 left) and respective task arrival rates 3.83 and 4.08 (Figure 8.1 right). Even under stability, backpressure may outperform greedy. Further, while greedy is unstable at $\lambda = 3.83$, the backpressure is stable even at $\lambda = 4.08$.

We evaluated the time-averaged number of tasks in the system for different task arrival rates for the three policies under our consideration. By Little’s law, the time-averaged number of tasks is a good proxy for the expected sojourn time of a typical task. These results are shown in Figure 8.2. We observe that the task arrival rates at which random, greedy, and backpressure become unstable are, 2.2, 3.80 and

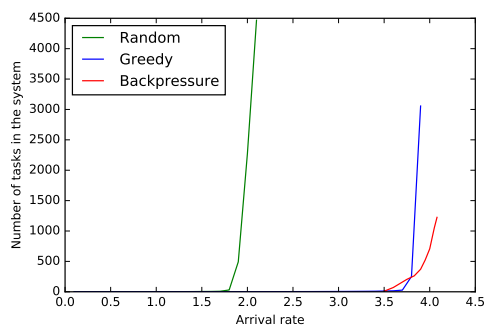


Figure 8.2: Comparison of Random, Greedy, and Backpressure policies.

4.10, respectively. Thus, the backpressure policy achieves throughput improvement of about 8% over the greedy policy. While greedy may perform better than backpressure at lower arrival rates, the backpressure significantly outperforms greedy at higher arrival rates.

8.6 Related Work

Bayesian Active Learning [49, 62, 22, 45] aims at learning true hypothesis by adaptively selecting sequence of experiments. In [22] labels are obtained for a batch of items at a time. In [45] a stream based budgeted setting is considered where a finite number of items arrive in a random order. In contrast we allow infinite stream of tasks and interested in the task resolution throughput. The crowdsourcing works such as [68, 111, 127, 47] consider task assignment problems for classification with unknown ground truths, however they consider a static model.

A related line of work is that on *stochastic online matching*, e.g., [88, 89, 55]. The stochastic online matching can be interpreted as a task-expert system where each expert is associated with a budget constraint that allows to solve at most one task. Unlike our work where the task types are uncertain, uncertainty in these models come from the arbitrariness of the future task arrivals and the monotonically decreasing available resource budgets.

Another related literature is that of constrained queueing systems, where arriving tasks are to be served by heterogeneous servers subject to resource constraints, e.g., [114, 99, 48, 123, 19, 5, 64, 29, 83, 113]. In particular, our backpressure matching policy is of a flavor similar to the policy proposed in [114]. The main difference from our work is that all these works assume that the task types are known.

In [12], the authors considered a constrained queueing system where the hardness of a task could be unknown and the task upon service can only become progressively harder. In [66], authors consider a setting where each task can be divided into a large number of subtasks of the same type, a small amount of which could be used to accurately learn the task's type.

8.7 Conclusion

We studied matching of tasks and experts in a system with uncertain task types. We established a complete characterization of the stability region of the system, i.e. the set of task arrival rates that can be supported by a matching policy such that the expected number of tasks waiting to be served is finite. We showed that any task arrival rate in the stability region can be supported by a back-pressure matching policy. We also compared with two baseline matching policies, and identified instances under which there is a substantial gap between the maximum task arrival

rates that can be supported by these policies and that of the optimum back-pressure matching policy.

8.8 Preprocessing phase

Above we assumed that all parameters are known: the success-probabilities $\{p_{sc}\}_c$ for each server s , the distribution $\{z_c\}_c$ and corresponding arrival rate $\pi_z\lambda$ for each mixed-type $z \in \mathcal{Z}$. In practice we need to estimate those.

We detail here some directions in case the tasks are of textual form, e.g., questions on a forum or curricula vitae of job applicants. In this setting, $\{z_c\}_c$ could be obtained as the output of a text-classifier. We discuss two common approaches and assume that the text underwent standard preprocessing:

- All letters are converted to lower-case.
- All symbols except $\{a, b, \dots, z\}$, spaces and "." are removed.
- All words are stemmed (for instance using Porter's algorithm [105]).

The *topic model* or *Latent Dirichlet Allocation (LDA)* [13] is an unsupervised classification method that outputs a topic distribution for every document in a collection with vocabulary V . A document of length N in this model is represented as a *bag of words*, that is, a vector $\mathbf{w} \in \mathbb{N}^{|V|}$ such that $\mathbf{w}(v)$ counts the number of times word $v \in V$ occurs in the document. Note that with this representation, all structural information (such as word order) of the document is lost.

The core assumption in this model is that every document is generated in the following way, given parameters $\xi > 0$, number of topics k , vector $\mathbf{a} = (a_1, \dots, a_k)$ (all elements > 0), and right stochastic matrix β of dimensions $k \times |V|$:

- Draw the document length N from $\text{Poi}(\xi)$;
- Draw the topic distribution $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ of this document from $\text{Dir}(\mathbf{a})$;
- For each word w_n ($n \in \mathbb{N}$), choose a topic $z_n \sim \text{Multinomial}(\boldsymbol{\theta})$ and, *given this topic*, choose the actual word with respective probabilities $\{\beta_{i1}, \dots, \beta_{i|V|}\}$, where $i = z_n$.

In this way we obtain a document of (random) length N that consists for a fraction θ_i of topic $i \in \{1, \dots, k\}$ and the frequency of words for that topic are on average $\{\beta_{i1}, \dots, \beta_{i|V|}\}$.

Given a (large) collection of M documents $\{\mathbf{w}_1, \dots, \mathbf{w}_M\}$ (e.g., questions or curricula vitae), we need to estimate the parameters \mathbf{a} and β . A standard approach is maximizing the log-likelihood of the data:

$$\ell(\mathbf{a}, \beta) = \sum_{u=1}^N \log(\mathbb{P}(\mathbf{w}_u | \mathbf{a}, \beta)).$$

In practice, this is too computationally expensive and we resort to the approach of variational inference described in Section 5.3 of [13]. Then, after estimating \mathbf{a} and β , we estimate for a given document the posterior distribution of its hidden variables (i.e., $\boldsymbol{\theta}$ and z_1, \dots, z_n) using the method described in Section 5.1 of [13].

Artificial neural networks are *supervised* classifiers, that could be used if part of the documents have been labelled [10, 67, 126].

8.8.1 Estimation of success-probabilities

Here we explain how one could estimate the success-probabilities and the distribution $\{\hat{\pi}_c\}_c$ of the pure types. Note that for $c \in C$, we have $\hat{\pi}_c = \sum_{z \in \mathcal{Z}} z_c \pi_z$. We assume that $m := |C| < \infty$.

Given a question of type c , let $X_s = \text{Ber}(p_s^c)$ be the indicator function of server s answering this question successfully.

We shall use *tensor decompositions* to learn the parameters $\{\hat{\pi}_c\}_{c=1}^m$ and $\{p_s^c\}_{c,s}$. To this end, we start with some standard tensor-terminology. For given vectors $v_1, v_2, v_3 \in \mathbb{R}^m$ we define the tensor-products $v_1 \otimes v_2$ and $v_1 \otimes v_2 \otimes v_3$ as

$$v_1 \otimes v_2 = v_1 v_2^T,$$

and, for $i, j, k \in [m]$,

$$(v_1 \otimes v_2 \otimes v_3)_{i,j,k} = v_1(i)v_2(j)v_3(k).$$

Thus $v_1 \otimes v_2$ is a matrix and $v_1 \otimes v_2 \otimes v_3$ is a *cube*. Put $y = (X_1, X_2, \dots, X_{m-1}, 1)^T$, i.e., the outcome vector of the first $m-1$ servers on a *single* question. Conditioned on c , let y_1, y_2 and y_3 be i.i.d. copies of y . The latter vectors represent i.i.d. attempts of *simultaneously* solving the same question (recall that we assumed a server to represent a homogeneous collection of experts). We then have the tensors

$$T = \mathbb{E}[y_1 \otimes y_2 \otimes y_3] = \sum_{c=1}^m \hat{\pi}_c \mathbb{E}[y|c] \otimes \mathbb{E}[y|c] \otimes \mathbb{E}[y|c] = \sum_{c=1}^m \hat{\pi}_c \boldsymbol{\mu}_c \otimes \boldsymbol{\mu}_c \otimes \boldsymbol{\mu}_c,$$

and

$$M = \mathbb{E}[y_1 \otimes y_2] = \sum_{c=1}^m \hat{\pi}_c \boldsymbol{\mu}_c \otimes \boldsymbol{\mu}_c = \sum_{c=1}^m \hat{\pi}_c \boldsymbol{\mu}_c \boldsymbol{\mu}_c^T,$$

where

$$\boldsymbol{\mu}_c = \mathbb{E}[y|c] = (p_1^c, p_2^c, \dots, p_{m-1}^c, 1)^T.$$

For the sequel, we shall assume that we have sufficiently accurate estimations of both T and M . In practice this is a strong assumption, because it asks that $m-1$ servers often attempt simultaneously the same questions. We shall also make the technical assumption that $\{\boldsymbol{\mu}_c\}_{c=1}^m$ is a *linearly independent set*.

Let us first study the case where T is known, and see if we can retrieve $\{\hat{\pi}_c\}_c$ and $\{\boldsymbol{\mu}_c\}_c$. In [7] a generalization of the matrix-diagonalization method is developed for tensors. It follows that the tensor T admits a *unique orthogonal decomposition*, that is, a collection of orthonormal vectors $\{v_1, v_2, \dots, v_m\}$ together with corresponding positive scalars $\lambda_j > 0$ such that

$$T = \sum_{j=1}^m \lambda_j v_j \otimes v_j \otimes v_j. \quad (8.5)$$

Our objective is thus to decompose the tensor

$$T = \sum_{i,j,k} T_{i,j,k} e_i \otimes e_j \otimes e_k$$

as in (8.5). To this end, we need the notion of an eigenvector/eigenvalue pair of T , which is defined in terms of the vector-valued map

$$u \mapsto T(I, u, u),$$

where $T(I, u, u)$ is in *standard tensor-terminology* defined as

$$T(I, u, u) = \sum_{i,j,k} T_{i,j,k} (e_j^T u)(e_k^T u)e_i.$$

An *eigenvector* u of T with corresponding *eigenvalue* λ is then a vector such that $T(I, u, u) = \lambda u$. Now, since the mapping $u \mapsto T(I, u, u)$ is non-linear, there are differences with the eigenvalue/eigenvector notion for matrices. However, as pointed out in Theorem 4.1 in [7], considering *robust eigenvectors* of T resolves the issues. A unit vector u is called a *robust eigenvector* of T if there exists $\epsilon > 0$ such that for all $\theta \in \{u' \in \mathbb{R}^m : \|u' - u\| \leq \epsilon\}$, repeatedly iterating the map

$$\bar{\theta} \mapsto \frac{T(I, \bar{\theta}, \bar{\theta})}{\|T(I, \bar{\theta}, \bar{\theta})\|} \quad (8.6)$$

starting from θ converges to u . It is *very interesting* to note that the robust eigenvectors of T are precisely the vectors $\{v_1, v_2, \dots, v_m\}$, implying that the orthogonal-decomposition is unique:

Theorem 8.8.1 (Theorem 4.1 in [7]). *Let T have an orthogonal decomposition as in (8.5), then*

- *The set of $\theta \in \mathbb{R}^m$ which do not converge to some v_i under repeated iteration of (8.6) has measure zero.*
- *The set of robust eigenvectors of T is equal to $\{v_1, v_2, \dots, v_m\}$.*

We explain now how the robust eigenvectors of $T(W, W, W)$, for a special matrix W (defined shortly), can be related to $\{\hat{\pi}_c\}_c$ and $\{\boldsymbol{\mu}_c\}_c$.

In Section 4.3.1 of [7] and also in [104] the following reduction method is described: Let U be the matrix of orthonormal eigenvectors of M (note that M is a full-rank matrix due to the assumption that the span of $\{\boldsymbol{\mu}_c\}_{c=1}^m$ has rank m) and D be its corresponding diagonal matrix of positive eigenvalues. Put $W = UD^{-1/2}$. Then,

$$W^T M W = I.$$

Note that there are standard diagonalization methods available to obtain U and D for the symmetric positive-definite matrix M .

Let $\bar{\boldsymbol{\mu}}_i = \sqrt{\hat{\pi}_i} W^T \boldsymbol{\mu}_i$. Then, $\{\bar{\boldsymbol{\mu}}_i\}$ are orthonormal vectors. Indeed,

$$I = W^T M W = \sum_{c=1}^m W^T (\sqrt{\hat{\pi}_c} \boldsymbol{\mu}_c) (\sqrt{\hat{\pi}_c} \boldsymbol{\mu}_c^T) W = \sum_{c=1}^m \bar{\boldsymbol{\mu}}_c \bar{\boldsymbol{\mu}}_c^T.$$

Note that we have access to $T(W, W, W) := \sum_{i,j,k} T_{i,j,k} (W^T e_i) \otimes (W^T e_j) \otimes (W^T e_k)$.

But,

$$\begin{aligned}
& \sum_{i,j,k} T_{i,j,k} (W^T e_i) \otimes (W^T e_j) \otimes (W^T e_k) \\
&= \sum_{i,j,k} T_{i,j,k} (W^T e_i) \otimes (W^T e_j) \otimes (W^T e_k) \\
&= \sum_{i,j,k} \left(\sum_c \hat{\pi}_c \boldsymbol{\mu}_c(i) \boldsymbol{\mu}_c(j) \boldsymbol{\mu}_c(k) \right) (W^T e_i) \otimes (W^T e_j) \otimes (W^T e_k) \\
&= \sum_c \hat{\pi}_c \sum_{i,j,k} (\boldsymbol{\mu}_c(i) W^T e_i) \otimes (\boldsymbol{\mu}_c(j) W^T e_j) \otimes (\boldsymbol{\mu}_c(k) W^T e_k) \\
&= \sum_c \hat{\pi}_c \left(\sum_i \boldsymbol{\mu}_c(i) W^T e_i \right) \otimes \left(\sum_j \boldsymbol{\mu}_c(j) W^T e_j \right) \otimes \left(\sum_k \boldsymbol{\mu}_c(k) W^T e_k \right) \\
&= \sum_c \hat{\pi}_c (W^T \boldsymbol{\mu}_c) \otimes (W^T \boldsymbol{\mu}_c) \otimes (W^T \boldsymbol{\mu}_c) \\
&= \sum_c \frac{1}{\sqrt{\hat{\pi}_c}} (\bar{\boldsymbol{\mu}}_c) \otimes (\bar{\boldsymbol{\mu}}_c) \otimes (\bar{\boldsymbol{\mu}}_c).
\end{aligned}$$

Hence, applying the above theorem to the tensor $T(W, W, W)$ gives that $\{\bar{\boldsymbol{\mu}}_c\}_c$ are its robust eigenvectors with corresponding eigenvalues $\{\frac{1}{\sqrt{\hat{\pi}_c}}\}_c$.

Given T , a simple algorithm would then be to compute W and then for many random starting vectors u , iterate the map $u \mapsto S(I, u, u) / \|S(I, u, u)\|$ until convergence, where $S := T(W, W, W)$.

Alternatively, there are other methods available (for instance the Tensor Power Method see [7]) to obtain from $T(W, W, W)$ its orthogonal decomposition. However, in practice we observe a tensor $\hat{T} = T + E$, where $E = \sum_{i,j,k} \epsilon_{i,j,k} e_i \otimes e_j \otimes e_k$, with $|\epsilon_{i,j,k}| \leq \epsilon$ and $e_i(l) = \delta_{il}$ (coordinate vectors). The symmetric tensor E represents noise due to the fact that neither our model nor moment estimates are perfect. Theorem 5.1 in [7] tells us that we can in this case use an algorithm that gives us approximations $\{\hat{v}_c\}_c$ and $\{\hat{\lambda}_c\}_c$ such that for all c , $\|\hat{v}_c - v_c\| = \mathcal{O}(\epsilon)$ and $\|\hat{\lambda}_c - \lambda_c\| = \mathcal{O}(\epsilon)$. See also [104] for a method that in practice outperforms existing methods.

How many observations do we need? To apply the above approach, we need all first-, second- and third order cross moments for servers $1, \dots, m-1$. In particular, we need approximations for $\sum_c \pi_c (p_s^c)^3$. Meaning that we need to be able to pose the same question three times to a server, and obtain three independent responses. If this is too restrictive, one could employ the method in [6]: one needs then all second order cross-moments between $3(m-1)$ different servers.

Note that when $|C| = m$, we need at least $m-1$ servers. In particular, for $|C| = 2$, we need only a single server. This makes intuitively sense as we then have three estimations, namely for $\pi p_1 + (1-\pi)p_2$, $\pi p_1^2 + (1-\pi)p_2^2$ and $\pi p_1^3 + (1-\pi)p_2^3$, while we have only three unknowns.

Estimating the parameters of the remaining servers (outside the set $\{1, \dots, m\}$) is easier, since $\hat{\pi}_1, \dots, \hat{\pi}_m$ are now known. Indeed, let A be the matrix containing the vectors $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m$ as its columns. Then, we can observe

$$A \begin{pmatrix} \pi_1 p_s^1 \\ \vdots \\ \pi_{m-1} p_s^{m-1} \\ \pi_m p_s^m \end{pmatrix} = \begin{pmatrix} \sum_c \pi_c p_1^c p_s^c \\ \vdots \\ \sum_c \pi_c p_{m-1}^c p_s^c \\ \sum_c \pi_c p_s^c \end{pmatrix}.$$

We carry out an *heuristic* error analysis, under the assumption that we obtained estimates for $\{\frac{1}{\sqrt{\hat{\pi}_c}}\}_c$ and $\bar{\boldsymbol{\mu}}_1, \dots, \bar{\boldsymbol{\mu}}_m$ containing errors of order $\mathcal{O}(\epsilon)$. Then, the error in $\hat{\pi}_c$ is of order $\mathcal{O}(\epsilon)$, but the error in $\bar{\boldsymbol{\mu}}_m$ is of order $\mathcal{O}\left(\frac{1}{\sqrt{\hat{\pi}_c}}\|UD^{1/2}\|\epsilon\right) = \mathcal{O}\left(\sqrt{\frac{d_{\max}}{\pi_{\min}}}\epsilon\right)$, where d_{\max} is the largest eigenvalue of M (note that U is orthonormal).

Therefore, instead of A , we observe $A + \delta A$, where $\|\delta A\| = \mathcal{O}\left(\sqrt{\frac{d_{\max}}{\pi_{\min}}}\epsilon\right)$. Now, consider the perturbation-equation

$$(A + \Delta A)(x + \delta x) = b + \delta b,$$

where $Ax = b$. Then, *heuristically*,

$$\begin{aligned} x + \delta x &= (A + \Delta A)^{-1}(b + \delta b) \\ &= (I + A^{-1}\Delta A)^{-1}A^{-1}(b + \delta b) \\ &= (I + \mathcal{O}(\|A^{-1}\Delta A\|))(x + A^{-1}\delta b) \end{aligned}$$

Since $\|A\| \geq 1$, we have,

$$\|\delta x\| \leq \mathcal{O}\left(\sqrt{\frac{d_{\max}}{\pi_{\min}}}\epsilon\right)(\|\Delta b\| + \|x\|) + \|\Delta b\|.$$

8.9 Proofs

8.9.1 Proof of Theorem 8.3.4

Proof of Part (i): Let us first describe the the transition rates for the CTMC under greedy policy. Let $q(n, n')$ be the transition rate from state n to state n' . Let e_z denote the vector with all coordinates equal to 0 except for the z -coordinate which equals 1. Fix a state n . For each $z \in \mathcal{Z}$ we have

$$\begin{aligned} q(n, n + e_z) &= \lambda \pi_z, \\ q(n, n - e_z) &= \sum_{s: z \in A_s(N)} \mu_s (1 - \psi_s(z)) \frac{1}{|A_s(N)|}, \\ q(n, n - e_z + e_{\phi_s(z)}) &= \sum_{s: z \in A_s(N)} \mu_s \psi_s(z) \frac{1}{|A_s(N)|}. \end{aligned}$$

Transition rate $q(n, n')$ for each (n, n') which is not as given above is equal to 0.

We first show that if $\lambda < 4a/(2 + a)$ then the system is stable. For each t let $t + \tau(t)$ be the time at which the first event (arrival or completion of a response) occurs after time t . Let $\tau_n = E[\tau(t)|N(t) = n]$, i.e., given that $N(t) = n$ at time t , τ_n is the expected time at which the first event occurs after time t . For example, for $n = 0$ we have $\tau_n = 1/\lambda$.

A common approach to show system stability is to use Lyapunov-Foster theorem, see e.g., Proposition I.5.3 on page 21 in [8]. The idea is to construct a function $L(\cdot)$ such that $L(n)$ tends to infinity as $|n| \rightarrow \infty$ and that has a strictly negative ‘drift’ for all but finite values of n , i.e., there exists a constant $\epsilon > 0$ such that

$$\mathbb{E} [L(N(t + \tau(t))) - L(N(t)) | N(t) = n] \leq -\epsilon \tau_n,$$

for all but finite values of n . Intuitively, negative drift condition implies that as $L(N(t))$ becomes large (i.e., as $N(t)$ becomes large) the system dynamics is such that

$L(N(t))$ tends to decrease in expectation. This prevents the $L(t)$ from blowing up to ∞ as t increases and thus keeps the system stable. Roughly, the Lyapunov-Foster theorem states that the negative drift condition is indeed sufficient to ensure that that system is positive recurrent and thus stable. We will use a variant of Lyapunov-Foster theorem, provided below, which follows from Theorem 8.13 in [108].

Theorem 8.9.1. *Consider an irreducible CTMC $N(t)$ that takes values in a countable state-space. Let $\tau(t)$ and τ_n be as defined above. If there exists a function $L(\cdot)$, and constants $K > 0$ and $\epsilon > 0$ such that for $L(n) > K$ we have*

$$\mathbb{E} [L(N(t + \tau(t))) - L(N(t)) | N(t) = n] \leq -\epsilon\tau_n,$$

and if $\{n : L(n) \leq K\}$ is finite, then $N(t)$ is positive recurrent.

Now suppose that $\lambda < 4a/(2 + a)$. Then, it can be checked that $\frac{2-a}{2(2-\lambda)}\lambda < a$. Thus, there exists $\delta > 0$ such $\frac{2-a}{2(2-\lambda)}\lambda = (1 + \delta)a$. Now, consider the following Lyapunov function: for each n , we have

$$\frac{1}{\tau_n}L(n) = (1 + \delta)\frac{2-a}{2(2-\lambda)}n_{z'} + n_{z''},$$

where δ is a constant obtained as above.

Consider states n such that $n_{z'} > 0$. For these states, we obtain

$$\begin{aligned} \frac{1}{\tau_n}L(n) &= (1 + \delta)\frac{2-a}{2(2-\lambda)}(\lambda - \mu_{s_1} - \mu_{s_2}) + (\mu_{s_1}\psi_{s_1}(z') + \mu_{s_2}\psi_{s_2}(z')) \\ &= (1 + \delta)\frac{2-a}{2(2-\lambda)}(\lambda - 2) + \frac{1-a}{2} + \frac{1}{2} = -\delta\frac{2-a}{2} < 0. \end{aligned}$$

Now, consider states n such that $n_{z'} = 0$ and $n_{z''} > 0$. For these states

$$\frac{1}{\tau_n}L(n) = \delta\frac{2-a}{2(2-\lambda)}\lambda - \mu_{s_1}a = -\delta a < 0.$$

Thus, the conditions of Theorem 8.9.1 are satisfied with $K = L((1, 1))/\tau_{(1,1)}$ and $\epsilon = \min(\delta a, \delta(2-a)/2)$. Hence, $N(t)$ is positive recurrent if $\lambda < 4a/(2 + a)$. We now show the only if part. Suppose that $\lambda \geq 4a/(2 + a)$. Then, the δ used in the above argument is greater than or equal to 0. Thus, drift is non-negative for all but finite values of n . Further, since $L(\cdot)$ is bounded, the maximum change in $L(\cdot)$ upon an arrival or a departure is also bounded, using Proposition I.5.4 on page 22 in [8], we establish the only if part.

Proof of Part (ii): Part (ii) is a special case of Theorem 8.9.2 stated and proven below.

Proof of Part (iii) and (iv): We use Theorem 8.4.1 from Section 8.4 in this chapter. For Asymmetric(a) system we have $\mathcal{Z} = \{z', z''\}$ where $z'_c = \frac{1}{2}$ for each $c \in C$, and $z''_c = 1_{\{c=c_2\}}$. The flow conservation constraints in Theorem 8.4.1 can be given as follows:

$$\lambda = \sum_s \nu_{s,z'}, \text{ and } \sum_s \nu_{s,z'}\psi_s(z') + \sum_s \nu_{s,z''}\psi_s(z'') = \sum_s \nu_{s,z''}$$

Suppose $a \geq \frac{1}{2}$. There exists an $\epsilon > 0$ such that $\lambda = \frac{3a(1-\epsilon)}{a+1}$. It can be checked that $\{\nu_{sz}\}_{s,z}$ where

$$\nu_{s_2,z'} = 1 - \epsilon, \nu_{s_2,z''} = 0, \nu_{s_1,z'} = \frac{2a-1}{a+1}(1-\epsilon), \nu_{s_1,z''} = \frac{2-a}{a+1}(1-\epsilon)$$

and $\{\delta_s\}_{s \in S}$ where $\delta_s = \epsilon$ for each s satisfies sufficient conditions of Theorem 8.4.1.

Now suppose $a < \frac{1}{2}$. There exists an $\epsilon > 0$ such that $\lambda = 2a(1 - \epsilon)$. It can be checked that $\{\nu_{s,z}\}_{s,z}$ where

$$\nu_{s_2,z'} = 2a(1 - \epsilon), \nu_{s_2,z''} = 0, \nu_{s_1,z'} = 0, \nu_{s_1,z''} = (1 - \epsilon)$$

and $\{\delta_s\}_{s \in S}$ where $\delta_s = \epsilon$ for each s satisfies sufficient conditions of Theorem 8.4.1.

The result then follows from the proof of Theorem 8.4.1 by taking \mathcal{Y} as \mathcal{Z} \square

8.9.2 Stability Threshold under Random Policy for an Arbitrary System

Theorem 8.9.2. *Under Random policy, a system is stable if and only if it satisfies the following:*

$$\lambda < \left(\sum_{c \in C} \frac{\sum_{z \in \mathcal{Z}} z_c \pi_z}{\sum_{s \in S} \mu_s p_{s,c}} \right)^{-1},$$

Proof:

Note that the system under random policy is equivalent to the one where pure-type of a task is revealed upon arrival, i.e., there is no uncertainty in task types. This is true since the random policy does not use the information of type (pure or mixed). We thus assume that the pure-type is indeed revealed upon arrival. Let $X_c(t)$ be the number of tasks in the system of pure-type c . Let $\bar{X}(t) = \{X_c(t)\}_c$. For each $c \in C$, the arrival rate into queue $X_c(t)$ is equal to

$$\lambda_c := \sum_{z \in \mathcal{Z}} \lambda z_c \pi_z.$$

We first show the if part of the result. Suppose that we have $\frac{\sum_{c \in C} \lambda_c}{\sum_{s \in S} \mu_s p_{s,c}} < 1$. We use the fluid limit approach developed in [110, 30, 84]. Roughly, given initial condition $X(0) = x$, the fluid trajectories of the state process $X(t)$ can be obtained by scaling initial conditions, speeding time, and then studying the rescaled process; i.e., letting $\lim_{k \rightarrow \infty} \frac{1}{k} X(0) = x$, and studying $\lim_{k \rightarrow \infty} \frac{1}{k} X(kt)$.

Using arguments similar to those used in [84], the fluid limits for the number of tasks in each class can be shown to satisfy the following at almost all times t : for each $c \in C$ and $X_c > 0$ we have

$$\frac{d}{dt} X_c = \lambda_c - \sum_{s \in S} \mu_s p_{s,c} \frac{X_c}{\sum_{c'} X_{c'}}. \quad (8.7)$$

Define a function L on \mathbb{R}^C as

$$L(X) = \sum_c X_c \log \left(\frac{X_c}{\gamma_c \sum_{c'} X_{c'}} \right), \quad (8.8)$$

where $\gamma_c := \frac{\lambda_c}{\sum_{s \in S} \mu_s p_{s,c}}$.

Further, by following the arguments similar to [84], if we have that $L(X) \rightarrow \infty$ and for some $\alpha > 0$, $\frac{d}{dt} L(X) \leq -\alpha$ as $|X| \rightarrow \infty$ under fluid limits then the stability of the original system follows. We show below that both these limits hold.

Using (8.7) and (8.8), we obtain

$$\frac{d}{dt}L(X) = \sum_c \left(\frac{d}{dt}X_c \right) \log \left(\frac{X_c}{\gamma_c \sum_{c'} X_{c'}} \right), \quad (8.9)$$

$$= \sum_c \left(\lambda_c - \sum_{s \in S} \mu_s p_{s,c} \frac{X_c}{\sum_{c'} X_{c'}} \right) \left(\log \frac{X_c}{\sum_{c'} X_{c'}} - \log \gamma_c \right), \quad (8.10)$$

$$= \sum_c \left(\sum_s \mu_s p_{s,c} \right) (\gamma_c - Y_c) \log(Y_c/\gamma_c), \quad (8.11)$$

where $Y_c := \frac{X_c}{\sum_{c'} X_{c'}}$. Now, (8.11) is negative and strictly bounded away from zero. This can be seen as follows. Firstly, all terms in the sum are non-positive. Therefore, it suffices to show that there exists a $\delta > 0$ such that there always exists a c for which $(\gamma_c - Y_c) \log(Y_c/\gamma_c) \leq -\delta$. Since, $\sum_c Y_c = 1$ and, for some fixed $\epsilon > 0$, $\sum_c \gamma_c = 1 - \epsilon$, it follows that there exists c such that $\gamma_c - Y_c \leq -\epsilon$. For this c , we thus also have $Y_c/\gamma_c \geq 1 + \epsilon$. Consequently, $(\gamma_c - Y_c) \log(Y_c/\gamma_c) \leq -\epsilon \log(1 + \epsilon)$.

Let $\theta = 1/\sum_c \gamma_c$ and $\hat{\gamma}_c = \gamma_c \theta$ for each $c \in C$. Since $\sum_c \gamma_c < 1$, we have $\theta > 1$. Let $D(p||q)$ be the Kullback-Leibler divergence between two discrete random variables with probability distributions p and q . Now, we can write

$$L(X) = \sum_c X_c \log \left(\frac{\theta X_c}{\hat{\gamma}_c \sum_{c'} X_{c'}} \right) \quad (8.12)$$

$$= \sum_c X_c \log \theta + \sum_c X_c \log \left(\frac{X_c}{\sum_{c'} X_{c'}} \cdot \frac{1}{\hat{\gamma}_c} \right) \quad (8.13)$$

$$= \sum_c X_c \log \theta + \left(\sum_c X_c \right) D \left(\left\{ \frac{X_c}{\sum_{c'} X_{c'}} \right\}_{c \in C} \left\| \left\{ \hat{\gamma}_c \right\}_{c \in C} \right) \quad (8.14)$$

which converges to ∞ as $|X|$ grows large.

Hence, the if part of the result follows. The same line of argument can also be used to show that if $\frac{\sum_{c \in C} \lambda_c}{\sum_{s \in S} \mu_s p_{s,c}} \geq 1$ the drift is non-negative for all but finite number of states. Further, since $L(X)$ is bounded, the maximum change in $L(X)$ upon an arrival or a departure is also bounded, using Proposition I.5.4 on page 22 in [8], we get the only if part. \square

8.9.3 Proof of Theorem 8.4.1

We first show stability under sufficient conditions. In networked systems, e.g. see [114, 48], a standard approach towards proving stability of a backpressure type policy is to design a ‘static’ policy using flow variables $\{\nu_{sz}\}_{s,z}$ and the slacks $\{\delta_s\}_s$ which provides a fixed service rate to each queue N_z such that its drift is sufficiently negative for each. However, in our setup the total number of queues $\{N_z\}_{z \in \mathcal{Z}}$ could be countable, while the total available slack is finite. Thus, it is not possible to design a static policy such that the drift in each individual queue is bounded from above by a negative constant. This is unlike any finite-server queueing system considered in the previous literature.

We thus take a different approach, which can be explained roughly as follows. Since the total exogenous arrival rate λ , and the total endogenous arrival rate, i.e. arrival into a queue due to failure at another queue, are both finite (they are bounded from above by $\sum_s \mu_s$), there exists a finite set $\mathcal{Y} \subset \mathcal{Z}$ such that the total arrival rate into $\mathcal{Z} \setminus \mathcal{Y}$ is less than $\min_{c \in C} \sum_{s \in S} \frac{\delta_s}{4} p_{s,c}$. Each task which enters a queue N_z

where $z \in \mathcal{Z} \setminus \mathcal{Y}$ is instead sent to a virtual queue X , and stays there until there is a success. If X is ‘large’ compared to the other queues then all the servers focus on X . The finite number of remaining queues are operated via a backpressure policy which accounts for the ‘expected backlog’ seen in these queues.

More formally, consider $\{\nu_{s,z}\}_{s,z}$ and positive constants $\{\delta_s\}_s$ as postulated in the theorem. Without loss of generality, assume that there exists a constant $0 < \epsilon < 1$ such that $\delta_s = \epsilon \mu_s$ for each $s \in S$. Let \mathcal{Y} be a finite subset of \mathcal{Z} such that

$$\sum_{z \in \mathcal{Z} \setminus \mathcal{Y}} \left(\lambda \pi_z + \sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \nu_{s,z'} \psi_s(z') \right) \leq \min_{c \in C} \sum_{s \in S} \frac{\delta_s}{4} p_{s,c}.$$

Since $\lambda + \sum_{s \in S, z \in \mathcal{Z}} \nu_{s,z} \leq 2 \sum_s \mu_s$, such a \mathcal{Y} exists.

Let X be the number of tasks in the system which are or have been in past of type $z \in \mathcal{Z} \setminus \mathcal{Y}$. Once a task enters queue X it does not leave it until success. There may be tasks in it with mixed-type in \mathcal{Y} . Note, our policy will depend on X and thus $\{z(s, t)\}_s$ will not be $N(t)$ measurable. In turn, $N(t)$ will not be a CTMC. For $z \in \mathcal{Y}$, let \tilde{X}_z and \tilde{N}_z be the tasks of mixed-type z which have and have not had mixed-type in $\mathcal{Z} \setminus \mathcal{Y}$. Also, for convenience for each $z \in \mathcal{Z} \setminus \mathcal{Y}$, let \tilde{X}_z be the tasks of mixed-type z , i.e., $N_z = \tilde{X}_z$ for each $z \in \mathcal{Z} \setminus \mathcal{Y}$. We now formally define $\sigma \left(\{\tilde{X}_z\}_{z \in \mathcal{Z}}, \{\tilde{N}_z\}_{z \in \mathcal{Y}} \right)$ -measurable backpressure policy. Thus, $\left(\{\tilde{N}_z\}_{z \in \mathcal{Y}}, \{\tilde{X}_z\}_{z \in \mathcal{Z}} \right)$ is a CTMC.

We now show stability of the system under this policy for Backpressure(\mathcal{Y}) as given in Definition 8.4.2. Below we will assume that the ties in selecting z from $B_s(\tilde{N}, X)$ are broken uniformly at random for simplicity of exposition. The proof can be easily extend to any other tie breaking approach. Consider the following Lyapunov function.

$$L(\tilde{N}, \tilde{X}) = \sum_{z \in \mathcal{Y}} \tilde{N}_z^2 + \left(\sum_{z \in \mathcal{Z}} X_z \right)^2 = \sum_{z \in \mathcal{Y}} \tilde{N}_z^2 + X^2.$$

For each t , let $t + \tau(t)$ be the time at which the first event (arrival or completion of a response) occurs after time t . Clearly, $\tau(t)$ is a stopping time. Further, let $\tau_{\tilde{n}, \tilde{x}}(t) = \mathbb{E} \left[\tau(t) | (\tilde{N}, \tilde{X}) = (\tilde{n}, \tilde{x}) \right]$.

Let

$$D(\tilde{n}, \tilde{x}) := \frac{1}{\tau_{\tilde{n}, \tilde{x}}} \mathbb{E} \left[L(\tilde{N}(t + \tau), \tilde{X}(t + \tau)) - L(\tilde{N}(t), \tilde{X}(t)) | \tilde{N}(t) = \tilde{n}, \tilde{X}(t) = \tilde{x} \right].$$

$D(\tilde{n}, \tilde{x})$ is called drift in state n . We would like to show that there exists a positive integer K and positive constant ϵ such that

$$D(\tilde{n}, \tilde{x}) \leq -\epsilon \quad \forall (\tilde{n}, \tilde{x}) \text{ s.t. } \max(|\tilde{n}|_\infty, x) \geq K.$$

Let for each $s \in S$ and $z \in \mathcal{Y}$ let

$$\nu_{s,z}^* = \mathbb{1}_{\{x \min_{c \in C} \sum_s \mu_s p_{s,c} > \sum_s \mu_s \max_{z \in \mathcal{Y}: \tilde{n}_z > 0} w_{s,z}(\tilde{n}, x)\}} \mathbb{1}_{\{z \in B_s(n)\}} \frac{1}{|B_s(n)|}.$$

Then, one can check that

$$\begin{aligned}
& \frac{1}{\tau_{\tilde{n}, \tilde{x}}} \mathbb{E} \left[\tilde{N}_z(t + \tau)^2 - \tilde{N}_z(t)^2 \mid \tilde{N}(t) = \tilde{n}, \tilde{X}(t) = \tilde{x} \right] \\
&= (2\tilde{n}_z + 1) \left(\lambda\pi_z + \sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \nu_{s,z'}^* \psi_s(z') \right) \\
&+ (-2\tilde{n}_z + 1) \sum_s \nu_{s,z}^*.
\end{aligned} \tag{8.15}$$

Further, let

$$\nu^* = 1_{\{x \min_{c \in C} \sum_s \mu_s p_{s,c} > \sum_s \mu_s \max_{z \in \mathcal{Y}: \tilde{n}_z > 0} w_{s,z}(\tilde{n}, x)\}}.$$

Then, we have that

$$\begin{aligned}
& \frac{1}{\tau_{\tilde{n}, \tilde{x}}} \mathbb{E} \left[X(t + \tau)^2 - X(t)^2 \mid \tilde{N}(t) = \tilde{n}, \tilde{X}(t) = \tilde{x} \right] \\
&\leq (2x + 1) \sum_{z \in \mathcal{Z} \setminus \mathcal{Y}} \left(\lambda\pi_z + \sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \nu_{s,z'}^* \psi_s(z') \right) + (-2x + 1) \nu^* \min_c \sum_s \mu_s p_{s,c}
\end{aligned}$$

Thus, we get

$$\begin{aligned}
D(\tilde{n}, \tilde{x}) &\leq \sum_{z \in \mathcal{Y}} (2\tilde{n}_z + 1) \left(\lambda\pi_z + \sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \nu_{s,z'}^* \psi_s(z') \right) + (-2\tilde{n}_z + 1) \sum_s \mu_s \nu_{s,z}^* \\
&+ (2x + 1) \sum_{z \in \mathcal{Z} \setminus \mathcal{Y}} \left(\lambda\pi_z + \sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \nu_{s,z'}^* \psi_s(z') \right) + (-2x + 1) \nu^* \min_c \sum_s \mu_s p_{s,c}.
\end{aligned}$$

Upon arranging terms, we obtain

$$\begin{aligned}
D(\tilde{n}, \tilde{x}) &\leq \sum_{z \in \mathcal{Y}} 2\tilde{n}_z \left(\lambda\pi_z + \sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \nu_{s,z'}^* \psi_s(z') - \sum_s \nu_{s,z}^* \right) \\
&+ 2x \left(\sum_{z \in \mathcal{Z} \setminus \mathcal{Y}} \left(\lambda\pi_z + \sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \nu_{s,z'}^* \psi_s(z') \right) - \nu^* \min_c \sum_s \mu_s p_{s,c} \right) \\
&+ \left(\lambda + \sum_{z \in \mathcal{Z}} \sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \nu_{s,z'}^* \psi_s(z') + \sum_{z \in \mathcal{Y}} \sum_s \nu_{s,z}^* + \nu^* \min_c \sum_s \mu_s p_{s,c} \right)
\end{aligned}$$

The last of the above three terms can be bounded by a constant, say $\alpha_1 = 10 \sum_s \mu_s$. For each $s \in S$ and $z \in \mathcal{Y}$ let $\hat{\nu}_{s,z}^* = (\mu_s - 3\delta_s/4)\nu_{s,z}^*$ and $\tilde{\nu}_{s,z}^* = (\delta_s/4)\nu_{s,z}^*$. Further,

let $\hat{\nu}^* = \min_c \sum_s (\mu_s - 3\delta_s/4)p_{s,c}\nu^*$ and $\tilde{\nu}^* = \min_c \sum_s (\delta_s/4)p_{sc}\nu^*$. Then,

$$\begin{aligned} D(\tilde{n}, \tilde{x}) &\leq \sum_{z \in \mathcal{Y}} 2\tilde{n}_z \left(\lambda\pi_z + \sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \hat{\nu}_{sz'}^* \psi_s(z') - \sum_s \hat{\nu}_{s,z}^* \right) \\ &\quad + 2x \left(\sum_{z \in \mathcal{Z} \setminus \mathcal{Y}} \left(\lambda\pi_z + \sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \hat{\nu}_{s,z'}^* \psi_s(z') \right) - \hat{\nu}^* \right) + \alpha_1 \\ &+ \sum_{z \in \mathcal{Y}} 2\tilde{n}_z \left(\sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \tilde{\nu}_{s,z}^* \psi_s(z') - \sum_s \tilde{\nu}_{sz}^* \right) + 2x \left(\sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \tilde{\nu}_{s,z'}^* \psi_s(z') - \tilde{\nu}^* \right) \end{aligned}$$

Consider the following lemma. Its proof is given in Section 8.9.4.

Lemma 8.9.3. *Recall the $\{\nu_{s,z}\}_{s,z}$ as postulated by the theorem.*

For $\Theta = \{\theta_{s,z}\}_{s \in S, z \in \mathcal{Y}} \cup \theta$, where θ and $\theta_{s,z}$ for each s, z are reals, let

$$\begin{aligned} f(\Theta) &= \sum_{z \in \mathcal{Y}} 2\tilde{n}_z \left(\lambda\pi_z + \sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \theta_{sz'} \psi_s(z') - \sum_s \theta_{s,z} \right) \\ &\quad + 2x \left(\sum_{z \in \mathcal{Z} \setminus \mathcal{Y}} \left(\lambda\pi_z + \sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \theta_{sz'} \psi_s(z') \right) - \theta \right). \end{aligned} \tag{8.16}$$

Then,

$$f(\{\hat{\nu}_{s,z}^*\}_{s \in S, z \in \mathcal{Y}} \cup \hat{\nu}^*) \leq f(\{\nu_{s,z}\}_{s \in S, z \in \mathcal{Y}} \cup \{\min_s \delta_s/4\}).$$

From definition of $\nu_{s,z}$, we get that the first term in $f(\{\nu_{s,z}\}_{s \in S, z \in \mathcal{Y}} \cup \{\min_s \delta_s/4\})$ is equal to 0 and that the second term is less than or equal to 0.

Thus, we obtain

$$\begin{aligned} D(\tilde{n}, \tilde{x}) &\leq \alpha_1 + \sum_{z \in \mathcal{Y}} 2\tilde{n}_z \left(\sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \tilde{\nu}_{s,z}^* \psi_s(z') - \sum_s \tilde{\nu}_{s,z}^* \right) \\ &\quad + 2x \left(\sum_{s \in S} \sum_{z' \in \phi_s^{-1}(z) \cap \mathcal{Y}} \tilde{\nu}_{s,z'}^* \psi_s(z') - \tilde{\nu}^* \right). \end{aligned} \tag{8.17}$$

Rearranging, we get

$$\begin{aligned} D(\tilde{n}, \tilde{x}) &\leq \alpha_1 - 2 \sum_{s \in S} \sum_{z \in \mathcal{Y}: \phi_s(z) \in \mathcal{Y}} \tilde{\nu}_{sz}^* (\tilde{n}_z - \psi_s(z) \tilde{n}_{\phi_s(z)}) \\ &\quad - 2 \sum_{s \in S} \sum_{z \in \mathcal{Y}: \phi_s(z) \in \mathcal{Z} \setminus \mathcal{Y}} \tilde{\nu}_{s,z}^* (\tilde{n}_z - \psi_s(z) X) - 2x \tilde{\nu}^*. \end{aligned} \tag{8.18}$$

Fix $\epsilon > 0$. We now show that there exist a positive integer K such that if $x > K$ or if $|\tilde{n}|_\infty > K$ then $D(\tilde{n}, \tilde{x}) \leq -\epsilon$. Upon rearranging terms, we obtain

$$\begin{aligned}
D(\tilde{n}, \tilde{x}) &\leq \alpha_1 - 2 \sum_{s \in S} \sum_{z \in \mathcal{Y}: \phi_s(z) \in \mathcal{Y}} \tilde{\nu}_{s,z}^* (\tilde{n}_z - \psi_s(z) \tilde{n}_{\phi_s(z)}) \\
&\quad - 2 \sum_{s \in S} \sum_{z \in \mathcal{Y}: \phi_s(z) \in \mathcal{Z} \setminus \mathcal{Y}} \tilde{\nu}_{s,z}^* (\tilde{n}_z - \psi_s(z) X) - 2x \tilde{\nu}^* \\
&= \alpha_1 - 2 \sum_{s \in S} \sum_{z \in \mathcal{Y}} \tilde{\nu}_{s,z}^* w_{s,z}(\tilde{n}, x) - 2\tilde{\nu}^* x, \\
&\leq \alpha_1 - \max \left(2 \sum_{s \in S} \sum_{z \in \mathcal{Y}} \tilde{\nu}_{s,z}^* w_{s,z}(\tilde{n}, x), 2\tilde{\nu}^* x \right)
\end{aligned} \tag{8.19}$$

Thus we get,

$$D(\tilde{n}, \tilde{x}) \leq \alpha_1 - x \min_{c \in C} \sum_{s \in S} \frac{\delta_s}{4} p_{s,c}.$$

Hence, for any (\tilde{n}, x) such that $x > (\alpha_1 + \epsilon) \min_{c \in C} \sum_{s \in S} \frac{\delta_s}{4} p_{s,c}$, we have $D(\tilde{n}, \tilde{x}) \leq -\epsilon$.

We also have that

$$D(\tilde{n}, \tilde{x}) \leq \alpha_1 - 2 \sum_{s \in S} \frac{\delta_s}{4} \max_{z \in \mathcal{Y}} w_{s,z}(\tilde{n}, x).$$

Thus,

$$D(\tilde{n}, \tilde{x}) \leq \alpha_1 - 2 \left(\min_{s \in S} \frac{\delta_s}{4} \right) \sum_{s \in S} \max_{z \in \mathcal{Y}} w_{s,z}(\tilde{n}, x) \leq \alpha_1 - 2 \left(\min_{s \in S} \frac{\delta_s}{4} \right) \max_{z \in \mathcal{Y}} \sum_{s \in S} w_{s,z}(\tilde{n}, x).$$

Now suppose that $x \leq \alpha_2 := (\alpha_1 + \epsilon) \min_{c \in C} \sum_{s \in S} \frac{\delta_s}{4} p_{s,c}$. Then, if we show that $\max_{z \in \mathcal{Y}} \sum_{s \in S} w_{s,z}(\tilde{n}, x) \rightarrow \infty$ as $|\tilde{n}|_\infty \rightarrow \infty$, then we have that $D(\tilde{n}, \tilde{x}) \leq -\epsilon$ a positive integer K' such that $|\tilde{n}|_\infty > K'$. We now show that $\sum_{s \in S} \max_{z \in \mathcal{Y}} w_{s,z}(\tilde{n}, x) \rightarrow \infty$ as $|\tilde{n}|_\infty \rightarrow \infty$.

Let $z^* = \arg \max_{z \in \mathcal{Y}} n_z$. Then we have

$$\begin{aligned}
\sum_{s \in S} w_{s,z^*}(\tilde{n}, x) &\leq \sum_s (n_{z^*} - \psi_s(z) \min(\alpha_2, n_{z^*})) \\
&= |S| n_{z^*} - \min(\alpha_2, n_{z^*}) \sum_s \psi_s(z)
\end{aligned}$$

which tends to infinity because

$$\begin{aligned}
\sum_s \psi_s(z) &= \sum_s \sum_c z_c^* (1 - p_{s,c}) = |S| - \sum_s \sum_c z_c^* p_{s,c} \\
&\leq |S| - \max_c \sum_s z_c^* p_{s,c} \leq |S| - \max_c z_c^* \min_c \sum_s p_{s,c} \\
&\leq |S| - \frac{1}{|C|} \min_c \sum_s p_{s,c} < |S|.
\end{aligned}$$

Thus, there exist positive constants K and ϵ such that if $x > K$ or if $|\tilde{n}|_\infty > K$ then $D(\tilde{n}, \tilde{x}) \leq -\epsilon$.

Let $\mathcal{A} := \{(\tilde{n}, \tilde{x}) : \max(|\tilde{n}|_\infty, x) \leq K\}$. Then, using a version of Lyapunov-Foster Theorem from [116], we have that, from any state (\tilde{n}, \tilde{x}) such that $|\tilde{n}| + x < \infty$, the expected time to return to \mathcal{A} , i.e., $\tau_{\mathcal{A}}(\tilde{n}, \tilde{x})$ is finite. Further,

$$T := \sup_{(\tilde{n}, \tilde{x}) \in \mathcal{A}} \tau_{\mathcal{A}}(\tilde{n}, \tilde{x}) < \infty.$$

Thus, starting with any state in \mathcal{A} , we return to it in a finite expected time. We will be done if we show that expected time to return to state $(0, 0)$ is also finite. We do this as follows. Fix a constant $\beta > 0$. Since there exists s such that $\min_c p_{s,c} > 0$, we have that for any interval of time of size β the probability that no arrival happens in the this interval and that a task leaves the system is finite.

Suppose that system is in a state $(\tilde{n}, \tilde{x}) \in \mathcal{A}$ at time $t = 0$. Now consider renewal times $T_i, i = 0, 1, 2, \dots$, where $T_0 = 0$ and for each $i > 0$, T_i is defined as follows: T_i is equal to $T_{i-1} + \beta$ if indeed no arrival happens and a task leaves the system in the interval $[T_{i-1}, T_i)$, else T_i is the first time of return to \mathcal{A} after T_{i-1} . Clearly $E[T_i] < \infty$ since T as defined above is finite. Further the probability that a task leaves the system in time $T_i - T_{i-1}$ is non-zero, say α . Thus, the time for the system emptying after first reaching \mathcal{A} can be upper-bounded by a sum of K geometric random variables with rate α . Thus expected time to return to state $(0, 0)$ is finite. Hence, the system is stable.

Now suppose that the system is stable. Then, the necessary conditions can be shown to hold by the ergodicity of the system, and letting $\nu_{s,z}$ for each s, z to be the long-term fraction of times a server s attempts a task in N_z . \square

8.9.4 Proof of Lemma 8.9.3

Upon rearrangement of terms in the expression of $f(\Theta)$ we obtain

$$f(\Theta)/2 = - \sum_s \sum_{z \in \mathcal{Y}: \phi_s(z) \in \mathcal{Y}} \theta_{s,z}(n_z - \psi_s(z')) n_{\phi_s(z)} - \sum_s \sum_{z \in \mathcal{Y}: \phi_s(z) \in \mathcal{Z} \setminus \mathcal{Y}} \theta_{s,z}(n_z - \psi_s(z')) x - x\theta.$$

By using the definition of weights $w_{s,z}$, we obtain

$$f(\Theta)/2 = - \sum_s \sum_{z \in \mathcal{Y}} \theta_{s,z} w_{s,z}(\tilde{n}, x) - x\theta \geq - \sum_s \left(\max_{z \in \mathcal{Y}} w_{s,z}(\tilde{n}, x) \right) \sum_{z \in \mathcal{Y}} \theta_{s,z} - x\theta.$$

Thus,

$$\begin{aligned} & f(\{\nu_{sz}\}_{s \in S, z \in \mathcal{Y}} \cup \{\min_s \delta_s/4\})/2 \\ & \geq - \sum_s \left(\max_{z \in \mathcal{Y}} w_{s,z}(\tilde{n}, x) \right) \sum_{z \in \mathcal{Y}} \nu_{s,z} - x \min_c \sum_{s \in S} (\delta_s/4) p_{s,c} \\ & \geq - \sum_s (\mu_s - \delta_s/2) \max_{z \in \mathcal{Y}} w_{sz}(\tilde{n}, x) - x \min_c \sum_{s \in S} (\delta_s/4) p_{s,c} \\ & \geq - 1_{\{\sum_s \max_{z \in \mathcal{Y}} w_{s,z}(\tilde{n}, x) (\mu_s - 3\delta_s/4) \geq x (\min_c \sum_s (\mu_s - 3\delta_s/4) p_{s,c})\}} \sum_s \max_{z \in \mathcal{Y}} w_{s,z}(\tilde{n}, x) (\mu_s - 3\delta_s/4) \\ & \quad - 1_{\{\sum_s \max_{z \in \mathcal{Y}} w_{s,z}(\tilde{n}, x) (\mu_s - 3\delta_s/4) < x (\min_c \sum_s (\mu_s - 3\delta_s/4) p_{s,c})\}} x \min_c \sum_s (\mu_s - 3\delta_s/4) p_{s,c} \\ & = f(\{\hat{\nu}_{sz}^*\}_{s \in S, z \in \mathcal{Y}} \cup \hat{\nu}^*)/2. \end{aligned}$$

Hence, the lemma holds. \square

Bibliography

- [1] E. Abbe, A Bandeira, and G. Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, Jan 2016.
- [2] E. Abbe and C. Sandon. Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic bp, and the information-computation gap. *arXiv preprint 1512.09080*, 2016.
- [3] L. Adamic and N. Glance. The political blogosphere and the 2004 u.s. election: Divided they blog. pages 36–43, 2005.
- [4] William Aiello, Fan Chung, and Linyuan Lu. A random graph model for power law graphs. *Experiment. Math.*, 10(1):53–66, 2001.
- [5] M. Alresaini, M. Sathiamoorthy, B. Krishnamachari, and M.J. Neely. Back-pressure with adaptive redundancy (bwar). In *Proc. IEEE INFOCOM*, pages 2300–2308, March 2012.
- [6] A Anandkumar, D. Foster, D. Hsu, S. Kakade, and Y.-K. Liu. A spectral algorithm for latent dirichlet allocation. *Advances in Neural Information Processing Systems*, 25, 2012.
- [7] A. Anandkumar, R. Ge, D. Hsu, S. Kakade, and M. Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15(1):2773–2832, January 2014.
- [8] Søren Asmussen. *Applied probability and queues*. Springer Science & Business Media, 2nd edition, 2003.
- [9] B. Ball, B. Karrer, and M. Newman. Efficient and principled method for detecting communities in networks. *Phys. Rev. E*, 84:036103, Sep 2011.
- [10] Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin. A neural probabilistic language model. *Journal of Machine Learning Research*, 3:1137–1155, March 2003.
- [11] S. Bernstein. *The Theory of Probabilities*. Gastehizdat Publishing House, Moscow, 1946.
- [12] Kostas Bimpikis and Mihalis G Markakis. Learning and hierarchies in service systems. *Unpublished manuscript*, 2015.
- [13] D. Blei, A Ng, and M. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:903–1022, 2003.
- [14] V. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008.
- [15] Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Struct. Algorithms*, 31(1):3–122, August 2007.

- [16] I. Bomze, M. Budinich, P. Pardalos, and M. Pelillo. *Handbook of Combinatorial Optimization*. Kluwer Academic Publishers, 1999.
- [17] C. Bordenave, M. Lelarge, and L. Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. *arXiv preprint 1501.06087*, 2015.
- [18] Pierre Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, volume 31. Springer Science & Business Media, 2013.
- [19] L.X. Bui, R. Srikant, and A. Stolyar. A novel architecture for reduction of delay and queueing structure complexity in the back-pressure algorithm. *IEEE/ACM Transactions on Networking*, 19(6):1597–1609, Dec 2011.
- [20] T.N. Bui, S. Chaudhuri, F.T. Leighton, and M. Sipser. Graph bisection algorithms with good average case behavior. *Combinatorica*, 7(2):171–191, 1987.
- [21] K. Chaudhuri, F. Chung, and A. Tsiatas. Spectral clustering of graphs with general degrees in the extended planted partition model. 23:35.1–35.23, 2012.
- [22] Yuxin Chen and Andreas Krause. Near-optimal batch mode active learning and adaptive submodular optimization. In *ICML’13*, pages 160–168, 2013.
- [23] F. Chung, L. Lu, and V. Vu. The spectra of random graphs with given expected degrees. *Internet Math.*, 1(3):257–275, 2003.
- [24] F. Chung and M. Radcliffe. On the spectra of general random graphs. *Electr. J. Comb.*, 18(1), 2011.
- [25] Fan Chung, Linyuan Lu, and Van Vu. The spectra of random graphs with given expected degrees. *Internet Math.*, 1(3):257–275, 2003.
- [26] A. Clauset, M. E. J. Newman, and C. Moore. Finding community structure in very large networks. *Phys. Rev. E*, 70, 2004.
- [27] A. Coja-Oghlan and A. Lanka. Finding planted partitions in random graphs with general degree distributions. *SIAM J. Discret. Math.*, 23(4):1682–1714, November 2009.
- [28] Amin Coja-Oghlan and Andr Lanka. Partitioning random graphs with general degree distributions. *SIAM Journal on Discrete Mathematics*, 273:127–141, 2008.
- [29] Y. Cui, E.M. Yeh, and R. Liu. Enhancing the delay performance of dynamic backpressure algorithms. *IEEE/ACM Transactions on Networking*, 2015.
- [30] J. G. Dai. On positive harris recurrence of multiclass queueing networks: A unified approach via fluid limit models. *The Annals of Applied Probability*, 5(1):49–77, 1995.
- [31] M Daltayanni, L De Alfaro, and P Papadimitriou. Workerrank: Using employer implicit judgements to infer worker reputation. In *WSDM*, 2015.
- [32] A. Dasgupta, J.E. Hopcroft, and F. McSherry. Spectral analysis of random graphs with skewed degree distributions. *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 602–610, 2004.
- [33] A. Dasgupta, J.E. Hopcroft, and F. McSherry. Spectral analysis of random graphs with skewed degree distributions. pages 602–610, Oct 2004.
- [34] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84:066106, Dec 2011.

- [35] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Inference and phase transitions in the detection of modules in sparse networks. *Phys. Rev. Lett.*, 107:065701, Aug 2011.
- [36] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.
- [37] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*, volume 31 of *Applications of Mathematics*. Springer, corrected 2nd edition, 1997.
- [38] W. E. Donath and A. J. Hoffman. Lower bounds for the partitioning of graphs. *IBM J. Res. Dev.*, 17(5):420–425, September 1973.
- [39] M. E. Dyer and A. M. Frieze. The solution of some random np-hard problems in polynomial expected time. *J. Algorithms*, 10(4):451–489, 1989.
- [40] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.
- [41] William Evans, Claire Kenyon, Yuval Peres, and Leonard J. Schulman. Broadcasting on trees and the ising model. *Ann. Appl. Probab.*, 10(2):410–433, 05 2000.
- [42] U. Feige and E. Ofek. *Random Struct. Algorithms*, 27(2):251–275, 2005.
- [43] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2):298–305, 1973.
- [44] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(35):75 – 174, 2010.
- [45] Kaito Fujii and Hisashi Kashima. Budgeted stream-based active learning via adaptive submodular maximization. In *NIPS’16*, pages 514–522, 2016.
- [46] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.
- [47] Chao Gao, Yu Lu, and Dengyong Zhou. Exact exponent in optimal rates for crowdsourcing. In *ICML’16*, pages 603–611, 2016.
- [48] Leonidas Georgiadis, Michael J. Neely, and Leandros Tassiulas. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends in Networking*, 1(1):1–144, 2006.
- [49] Daniel Golovin, Andreas Krause, and Debajyoti Ray. Near-optimal bayesian active learning with noisy observations. In *NIPS’10*, pages 766–774, 2010.
- [50] L. Gulikers, M. Lelarge, and L. Massoulié. An impossibility result for reconstruction in a degree-corrected planted-partition model. *arXiv preprint 1511.00546*, 2015.
- [51] L. Gulikers, M. Lelarge, and L. Massoulié. Non-backtracking spectrum of degree-corrected stochastic blockmodels. *ITCS 2017*, january 2017.
- [52] L. Gulikers, M. Lelarge, and L. Massoulié. A spectral method for community detection in moderately-sparse degree-corrected stochastic block models. *Advances in Applied Probability*, september 2017.
- [53] M. B. Hastings. Community detection as an inference problem. April 2006.
- [54] S. Hata and H. Nakao. Localization of laplacian eigenvectors on random networks. *Nature (Scientific Reports)*, April 2017.

- [55] Chien-Ju Ho and Jennifer Wortman Vaughan. Online task assignment in crowdsourcing markets. In *AAAI'12*, pages 45–51, 2012.
- [56] P. Hoff, A. Raftery, and M. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97:1090–1098, 2001.
- [57] P. Holland, K. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, June 1983.
- [58] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.
- [59] M. Horton, H. Stark, and T. Terras. What are zeta functions of graphs and what are they good for? *Contemporary Mathematics, Quantum Graphs and Their Applications*, 415:173–190, 2006.
- [60] K. i. Hashimoto. Zeta functions of finite graphs and representations of p-adic groups. In *Automorphic Forms and Geometry of Arithmetic Varieties*, volume 15 of *Advanced Studies in Pure Mathematics*, pages 211 – 280. Academic Press, 1989.
- [61] Y. Iba. The nishimori line and bayesian statistics. *Journal of Physics A: Mathematical and General*, 32(21):3875, 1999.
- [62] Shervin Javdani, Yuxin Chen, Amin Karbasi, Andreas Krause, Drew Bagnell, and Siddhartha S Srinivasa. Near optimal bayesian active learning for decision making. In *AISTATS'14*, pages 430–438, 2014.
- [63] Mark Jerrum and Gregory B. Sorkin. The metropolis algorithm for graph bisection. *Discrete Applied Mathematics*, 82(13):155 – 175, 1998.
- [64] Bo Ji, Changhee Joo, and N.B. Shroff. Delay-based back-pressure scheduling in multihop wireless networks. *IEEE/ACM Transactions on Networking*, 21(5):1539–1552, Oct 2013.
- [65] J. Jin. Fast community detection by score. *Ann. Statist.*, 43(1):57–89, 02 2015.
- [66] Ramesh Johari, Vijay Kamble, and Yash Kanoria. Know your customer: Multi-armed bandits with capacity constraints. abs/1603.04549, 2016.
- [67] R. Johnson and T. Zhang. Effective use of word order for text categorization with convolutional neural networks. *The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 103–112, 2015.
- [68] David R. Karger, Sewoong Oh, and Devavrat Shah. Budget-optimal task allocation for reliable crowdsourcing systems. *Operations Research*, 62(1):1–24, 2014.
- [69] B. Karrer and M. Newman. Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, 83:016107, Jan 2011.
- [70] H. Kesten and B. P. Stigum. Additional limit theorems for indecomposable multidimensional galton-watson processes. *Ann. Math. Statist.*, 37(6):1463–1481, 12 1966.
- [71] H. Kesten and B. P. Stigum. A limit theorem for multidimensional galton-watson processes. *Ann. Math. Statist.*, 37(5):1211–1223, 10 1966.
- [72] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.

- [73] Anurag Kumar. Discrete event stochastic processes: Lecture notes for an engineering curriculum. 2014.
- [74] R Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Trawling the web for emerging cyber-communities. *Computer Networks*, 1999.
- [75] C. Le and R. Vershynin. Concentration and regularization of random graphs. *arXiv preprint arXiv:1506.00669*, 2015.
- [76] J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *arXiv preprint arXiv:1312.2050*, 2013.
- [77] G. Linden, B. Smith, and J. York. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, Jan 2003.
- [78] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2), 1982.
- [79] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.
- [80] F. Luccio and M. Sami. Ieee transactions on circuit theory. *Experiment. Math.*, 16, 1969.
- [81] D. Lusseau, K. Schneider, O. Boisseau, P. Haase, E. Slooten, and S. Dawson. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4):396–405, 2003.
- [82] Russell Lyons. Random walks and percolation on trees. *Ann. Probab.*, 18(3):931–958, 07 1990.
- [83] Siva Theja Maguluri, Sai Kiran Burle, and R. Srikant. Optimal heavy-traffic queue length scaling in an incompletely saturated switch. *SIGMETRICS Perform. Eval. Rev.*, 44(1):13–24, June 2016.
- [84] Laurent Massoulié. Structural properties of proportional fairness: Stability and insensitivity. *Annals of Applied Probability*, 17(3):809–839, 2007.
- [85] Laurent Massoulié. Community detection thresholds and the weak ramanujan property. *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 694–703, 2014.
- [86] F. McSherry. Spectral partitioning of random graphs. *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 529–537, 2001.
- [87] F. McSherry. Spectral partitioning of random graphs. *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 529–537, Oct 2001.
- [88] A. Mehta and D. Panigrahi. Online matching with stochastic rewards. In *IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 728–737, Oct 2012.
- [89] Aranyak Mehta, Bo Waggoner, and Morteza Zadimoghaddam. *Online Stochastic Matching with Unequal Probabilities*, pages 1388–1404. 2015.
- [90] M. Mihail and C. Papadimitriou. On the eigenvalue power law. In *Proceedings of the 6th International Workshop on Randomization and Approximation Techniques*, RANDOM '02, pages 254–262, London, UK, UK, 2002. Springer-Verlag.

- [91] C. Moore. The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *arXiv preprint arXiv:1702.00467*, 2017.
- [92] E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461, 2014.
- [93] E. Mossel, J. Neeman, and A. Sly. Consistency thresholds for the planted bisection model. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing, STOC '15*, pages 69–75. ACM, 2015.
- [94] E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *Combinatorica*, 2017.
- [95] E. Mossel and Y. Peres. Information flow on trees. *The Annals of Applied Probability*, 13(3):817–844, 08 2003.
- [96] R. Murty. Ramanujan graphs. *J. Ramanujan Math. Soc.*, 18(1):1–20, 2003.
- [97] M. Mzard, G. Parisi, and M. Virasoro. Sk model: The replica solution without replicas. *EPL (Europhysics Letters)*, 1(2):77, 1986.
- [98] R. Nadakuditi and M. Newman. Graph spectra and the detectability of community structure in networks. *Phys. Rev. Lett.*, 108:188701, May 2012.
- [99] M. J. Neely, E. Modiano, and C. E. Rohrs. Dynamic power allocation and routing for time varying wireless networks. In *Proc. IEEE INFOCOM*, 2003.
- [100] George L Nemhauser and Laurence A Wolsey. *Integer and combinatorial optimization*, volume 18. Wiley, 1988.
- [101] M. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2):026113, February 2004.
- [102] M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Phys. Rev. E*, 69, Jun 2004.
- [103] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23), 2006.
- [104] A. Podosinnikova, F. Bach, and S. Lacoste-Julien. Rethinking lda: Moment matching for discrete ica. *Advances in Neural Information Processing Systems 28*, pages 514–522, 2015.
- [105] M. Porter. Readings in information retrieval. pages 313–316, 1997.
- [106] T. Qin and K. Rohe. Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems*, pages 3120–3128, 2013.
- [107] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences*, 101(9):2658, 2004.
- [108] Philippe Robert. Stochastic networks and queues, stochastic modelling and applied probability series, vol. 52, 2003.
- [109] K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 08 2011.
- [110] Aleksandr Nikolaevich Rybko and Alexander L Stolyar. Ergodicity of stochastic processes describing the operation of open queueing networks. *Problemy Peredachi Informatsii*, 28(3):3–26, 1992.

- [111] Nihar B. Shah, Dengyong Zhou, and Yuval Peres. Approval voting and incentives in crowdsourcing. In *ICML '15*, pages 10–19, 2015.
- [112] V. Shah, L. Gulikers, L. Massoulié, and M. Vojnović. Adaptive matching for expert systems with uncertain task types. *Allerton 2017*, October 2017.
- [113] Virag Shah, Gustavo de Veciana, and George Kesidis. A stable approach for routing queries in unstructured p2p networks. *IEEE/ACM Transactions on Networking*, 24(5):3136–3147, Oct 2016.
- [114] L. Tassiulas and A. Ephremides. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control*, 37:1936–1948, 1992.
- [115] D. Tomozei and L. Massoulié. *Stochastic Systems*, 4(1):1–43, 2014.
- [116] R. L. Tweedie. Criteria for classifying general markov chains. *Advances in Applied Probability*, 8(4):737–771, 1976.
- [117] U. von Luxburg. A tutorial on spectral clustering. *Stat. Comput.*, 17(4):395–416, 2007.
- [118] K. Wakita and T. Tsurumi. Finding community structure in mega-scale social networks. In *Proceedings of the 16th International Conference on World Wide Web*, pages 1275–1276. ACM, 2007.
- [119] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *The Annals of Mathematics*, pages 548–564, 1955.
- [120] Eugene P. Wigner. On the Distribution of the Roots of Certain Symmetric Matrices. *The Annals of Mathematics*, 67:325–327, 1958.
- [121] J. Yang, J. McAuley, and J. Leskovec. Community detection in networks with node attributes. pages 1151–1156, Dec 2013.
- [122] Heng-Qing Ye, Jihong Ou, and Xue-Ming Yuan. Stability of data networks: Stationary and bursty models. *Operations Research*, 53(1):107–125, 2005.
- [123] Lei Ying, Sanjay Shakkottai, Aneesh Reddy, and Shihuan Liu. On combining shortest-path and back-pressure routing over multihop wireless networks. *IEEE/ACM Transactions on Networking*, 19(3):841–854, June 2011.
- [124] W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33:452–473, 1977.
- [125] P. Zhang, F. Krzakala, J. Reichardt, and L. Zdeborova. Comparative study for inference of hidden classes in stochastic block models. *Journal of Statistical Mechanics: Theory and Experiment*, Dec 2012.
- [126] X. Zhang, J. Zhao, and Y. Lecun. Character-level convolutional networks for text classification. *Advances in Neural Information Processing Systems*, 2015-January:649–657, 2015.
- [127] Yuchen Zhang, Xi Chen, Dengyong Zhou, and Michael I. Jordan. Spectral methods meet em: A provably optimal algorithm for crowdsourcing. *J. Mach. Learn. Res.*, 17(1):3537–3580, January 2016.

Appendices

Appendix A

Algebraic Preliminaries

We shall make use of the following fact about the spectral radius:

Lemma A.0.4. *If $|X| \leq Y$ holds entry-wise for two real symmetric matrices X and Y , then*

$$\rho(X) \leq \rho(Y).$$

Proof. Due to the Rayleigh-Ritz theorem, we have

$$\rho(X) = \max_{\|z\|=1} \|Xz\|.$$

Hence,

$$\rho(X) = \max_{\|z\|=1} \|Xz\| \leq \max_{\|z\|=1} \|Y|z|\| = \max_{\|z\|=1} \|Yz\| = \rho(Y).$$

□

The following lemma could be of independent interest as a simple alternative to the commonly used David-Kahan theorem:

Lemma A.0.5. *Let $A, \delta A$ be two $n \times n$ symmetric matrices. Let $\lambda_1 \geq \dots \geq \lambda_n$ be the eigenvalues of $A + \delta A$ and $\mu_1 \geq \dots \geq \mu_n$ be the eigenvalues of A . Let $\Delta = \min\{|\mu_i - \mu_j| : \mu_i \neq \mu_j, \mu_i, \mu_j \text{ eigenvalue of } A\}$. Assume that $\rho(\delta A) < \frac{\Delta}{2}$. Let \mathbf{v}_i be a normed eigenvector of $A + \delta A$ corresponding to eigenvalue λ_i , for any $i = 1, \dots, n$. Then,*

1. $|\lambda_i - \mu_i| \leq \rho(\delta A)$,
2. the dimension of the eigenspace E_i of $A + \delta A$ corresponding to the eigenvalue λ_i is no larger than the dimension of the eigenspace of A corresponding to the eigenvalue μ_i ,
3. there exists a normed eigenvector $\hat{\mathbf{v}}_i$ of A corresponding to eigenvalue μ_i such that

$$\mathbf{v}_i \cdot \hat{\mathbf{v}}_i \geq \sqrt{1 - \left(\frac{\rho(\delta A)}{\Delta/2}\right)^2}.$$

Proof. (i) is due to Weyl's inequality (see for instance [58]).

To prove (ii), let d be the dimension of E_i and write $\lambda_i = \lambda_{i+1} = \dots = \lambda_{i+d-1}$. Since $|\lambda_i - \mu_i| \leq \rho(\delta A)$, we have $|\mu_i - \mu_{i+1}| \leq 2\rho(\delta A) < \Delta$. Thus $\mu_i = \mu_{i+1}$, and similarly for the other eigenvalues.

To prove (iii), we start with some notation: Let m be the number of distinct eigenvalues of A , denote those distinct numbers as $\gamma_1 > \dots > \gamma_m$. Define $S_i = \{u \in \{1, \dots, n\} : \mu_u = \gamma_i\}$, the set of indices of eigenvalues that are all equal to γ_i . For $u \in \{1, \dots, n\}$, define $\tau_u \in \{1, \dots, m\}$ as the unique index such that $u \in S_{\tau_u}$. Write

$$\mathbf{v}_i = \sum_j \alpha_j \mathbf{w}_j,$$

where $\{\mathbf{w}_j\}_j$ are orthonormal eigenvectors of A with associated eigenvalues $\{\mu_j\}_j$. Then,

$$(A + \delta A)\mathbf{v}_i = \sum_j \alpha_j \mu_j \mathbf{w}_j + (\delta A)\mathbf{v}_i.$$

Hence,

$$(\delta A)\mathbf{v}_i = \sum_{j \notin S_{\tau_i}} \alpha_j (\lambda_i - \mu_j) \mathbf{w}_j + \sum_{j \in S_{\tau_i}} \alpha_j (\lambda_i - \mu_j) \mathbf{w}_j.$$

Taking norms on both sides,

$$(\rho(\delta A))^2 \geq \sum_{j \notin S_{\tau_i}} \alpha_j^2 (\lambda_i - \mu_j)^2 \geq \sum_{j \notin S_{\tau_i}} \alpha_j^2 (\Delta - \Delta/2)^2 = \left(1 - \sum_{j \in S_{\tau_i}} \alpha_j^2\right) (\Delta/2)^2,$$

because, by definition $|\mu_i - \mu_j| \geq \Delta$ if $\tau_i \neq \tau_j$, and our observation $|\lambda_i - \mu_i| \leq \rho(\delta A) < \Delta/2$. Put

$$\hat{\mathbf{v}}_i = \frac{1}{\sqrt{\sum_{j \in S_{\tau_i}} \alpha_j^2}} \sum_{j \in S_{\tau_i}} \alpha_j \mathbf{w}_j,$$

then

$$\mathbf{v}_i \cdot \hat{\mathbf{v}}_i = \sqrt{\sum_{j \in S_{\tau_i}} \alpha_j^2} \geq \sqrt{1 - \left(\frac{\rho(\delta A)}{\Delta/2}\right)^2}.$$

□

Lemma A.0.6. Consider a square $n \times n$ symmetric zero-diagonal random matrix A such that its elements $A_{uv} = A_{vu}$ are independent Bernoulli random variables with parameters

$$\mathbb{E}[A_{uv}] = a_{uv} \frac{\hat{\omega}(n)}{n},$$

where the a_{uv} are constants independent of n and $\hat{\omega}(n) = \Omega(\log(n))$. Then, with probability larger than $1 - \mathcal{O}\left(\frac{1}{n^2}\right)$, the spectral radius of $A - \mathbb{E}[A]$ satisfies

$$\rho(A - \mathbb{E}[A]) \leq \mathcal{O}\left(\sqrt{\hat{\omega}(n)}\right).$$

Proof. This is precisely Lemma 2 in [115], where we quantified the term *with high probability*. We did this by choosing $c_1 > 3$ in its proof. Note that the latter proof builds further on results by Feige and Ofek [42]. □

Lemma A.0.7 (Bernstein's inequality). Let X_1, \dots, X_n be zero-mean independent random variables all bounded from above by one. Put $\sigma^2 = \frac{1}{n} \sum_{u=1}^n \text{var}(X_u)$. Then,

$$\mathbb{P}\left(\frac{1}{n} \sum_{u=1}^n X_u > \epsilon\right) \leq \exp\left(-\frac{n\epsilon^2}{2(\sigma^2 + \epsilon/3)}\right).$$

Proof. See [11].

□

Note that Bernstein's lemma can easily be extended to the case of non-centred random variables.

Résumé

Dans cette thèse, nous étudions deux problèmes d'apprentissage automatique: (I) la détection des communautés et (II) l'appariement adaptatif.

I) Il est bien connu que beaucoup de réseaux ont une structure en communautés. La détection de ces communautés nous aide à comprendre et exploiter des réseaux de tout genre. Cette thèse considère principalement la détection des communautés par des méthodes spectrales utilisant des vecteurs propres associés à des matrices choisies avec soin. Nous faisons une analyse de leur performance sur des graphes artificiels. Au lieu du modèle classique connu sous le nom de « Stochastic Block Model » (dans lequel les degrés sont homogènes) nous considérons un modèle où les degrés sont plus variables: le « Degree-Corrected Stochastic Block Model » (DC-SBM). Dans ce modèle les degrés de tous les noeuds sont pondérés - ce qui permet de générer des suites des degrés hétérogènes. Nous étudions ce modèle dans deux régimes: le régime dense et le régime « épars », ou « dilué ». Dans le régime dense, nous prouvons qu'un algorithme basé sur une matrice d'adjacence normalisée réussit à classer correctement tous les noeuds sauf une fraction négligeable. Dans le régime épars il existe un seuil en termes de paramètres du modèle en-dessous lequel n'importe quel algorithme échoue par manque d'information. En revanche, nous prouvons qu'un algorithme utilisant la matrice « non-backtracking » réussit jusqu'au seuil - cette méthode est donc très robuste. Pour montrer cela nous caractérisons le spectre des graphes qui sont générés selon un DC-SBM dans son régime épars. Nous concluons cette partie par des tests sur des réseaux sociaux.

II) Les marchés d'intermédiation en ligne tels que des plateformes de Question-Réponse et des plateformes de recrutement nécessitent un appariement basé sur une information incomplète des deux parties. Nous développons un modèle de système d'appariement entre tâches et serveurs représentant le comportement de telles plateformes. Pour ce modèle nous donnons une condition nécessaire et suffisante pour que le système puisse gérer un certain flux de tâches. Nous introduisons également une politique de « back-pressure » sous lequel le débit gérable par le système est maximal. Nous prouvons que cette politique atteint un débit strictement plus grand qu'une politique naturelle « gloutonne ». Nous concluons en validant nos résultats théoriques avec des simulations entrainées par des données de la plateforme Stack-Overflow.

Mots Clés

Apprentissage automatique, détection des communautés, réseaux sociaux, degree-corrected stochastic block models, méthodes spectrales, matrices aléatoires, matrice non-backtracking, graphes aléatoires, appariement adaptatif, apprentissage par renforcement, théorie des files d'attente.

Abstract

In this thesis, we study two problems of machine learning: (I) community detection and (II) adaptive matching.

I) It is well-known that many networks exhibit a community structure. Finding those communities helps us understand and exploit general networks. In this thesis we focus on community detection using so-called spectral methods based on the eigenvectors of carefully chosen matrices. We analyse their performance on artificially generated benchmark graphs. Instead of the classical Stochastic Block Model (which does not allow for much degree-heterogeneity), we consider a Degree-Corrected Stochastic Block Model (DC-SBM) with weighted vertices, that is able to generate a wide class of degree sequences. We consider this model in both a dense and sparse regime. In the dense regime, we show that an algorithm based on a suitably normalized adjacency matrix correctly classifies all but a vanishing fraction of the nodes. In the sparse regime, we show that the availability of only a small amount of information entails the existence of an information-theoretic threshold below which no algorithm performs better than random guess. On the positive side, we show that an algorithm based on the non-backtracking matrix works all the way down to the detectability threshold in the sparse regime, showing the robustness of the algorithm. This follows after a precise characterization of the non-backtracking spectrum of sparse DC-SBM's. We further perform tests on well-known real networks.

II) Online two-sided matching markets such as Q&A forums and online labour platforms critically rely on the ability to propose adequate matches based on imperfect knowledge of the two parties to be matched. We develop a model of a task / server matching system describing platform operation in the presence of such uncertainty. For this model, we give a necessary and sufficient condition for an incoming stream of tasks to be manageable by the system. We further identify a so-called back-pressure policy under which the throughput that the system can handle is maximal. We show that this policy achieves strictly larger throughput than a natural greedy policy. Finally, we validate our model and confirm our theoretical findings with experiments based on user-contributed content on an online platform.

Keywords

Machine learning, community detection, social networks, degree-corrected stochastic block models, spectral methods, random matrices, non-backtracking matrix, random graphs, recommendation systems, reinforcement learning, queueing theory.