

Combining Heterogeneous Information

Contributions to the Extraction and Analysis of Feature-Rich Complex Networks

Vincent Labatut

Laboratoire Informatique d'Avignon, Avignon Université – LIA EA 4128
vincent.labatut@univ-avignon.fr

16/06/2021

Committee :

Pr. Éric Gaussier, *Université Grenoble Alpes, LIG*

Pr. Jean-Loup Guillaume, *Université de La Rochelle, L3i*

Pr. Maguelonne Teisseire, *INRAE, UMR Tetis*

Pr. Roger Guimerà, *Universitat Rovira i Virgili, ICREA*

Pr. Céline Rouveirol, *Université Sorbonne Paris Nord, LIPN*

Reviewer

Reviewer

Reviewer

Examiner

Examiner



AVIGNON
UNIVERSITÉ

Outline

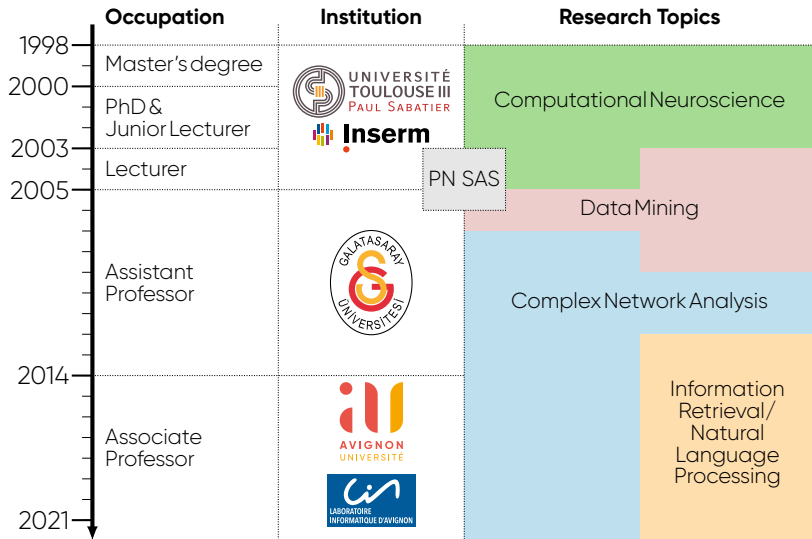
- 1 Academic Background
- 2 Notion of Feature-Rich Network
- 3 Vertex-Attributed Networks
- 4 Dynamic Networks
- 5 Spatial, Signed, and Multiplex Networks
- 6 Perspectives

Section 1

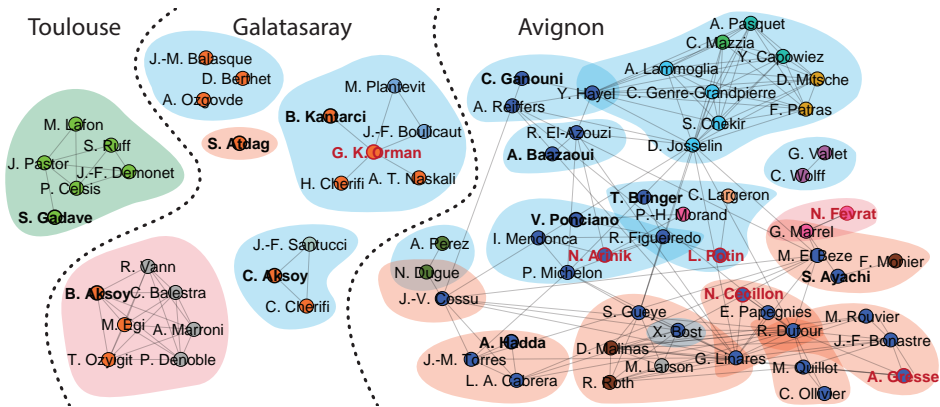
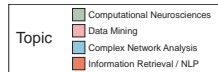
Academic Background

Academic Background

Career Milestones

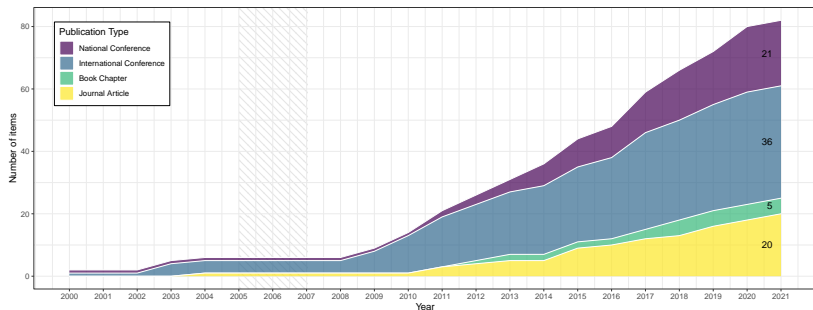


Academic Background Collaboration & Supervision



- 13 undergraduate theses
- 13 **graduate** theses
- 2 pre-doctoral internships
- 6 **doctoral** theses (4 ongoing)

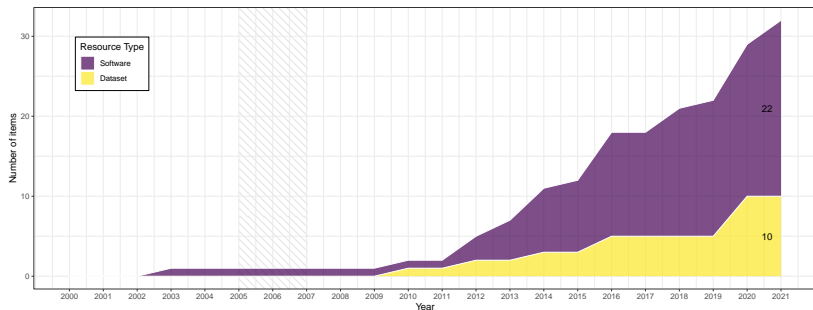
Academic Background Publications



- 2005–2007: change of **research topic**
- **Complex networks** outlets: J. Complex Networks, Soc. Netw., IEEE T. Comput. Soc. Syst., Physica A, Network Science, SNAM, J. Stat. Mech., ASONAM
- **IR/NLP** outlets: Multimedia Tools and Applications, LREC, ICASSP, SLSP, Interspeech, CICling
- **General** outlets: IEEE Access, ACM Computing Surveys, Springer Nature Computer Science

Academic Background

Other Resources



- **Software:** implementation of original methods from our papers, and of existing methods from other authors
- **Datasets:** annotated corpora, extracted networks, software outputs
- Produced resources publicly available online under an open license (whenever possible)
- All articles available on preprint servers (HAL, arXiv)

Section 2

Notion of Feature-Rich Network

Notion of Feature-Rich Network

Complex Systems and Complex Networks

Complex System

Relatively isolated set of objects interacting in a **heterogeneous** way. [Say15].

- **Heterogeneous**: non-linearity, self-organization, feedback loops...
- **Emerging** properties [Bar15]

Complex Network

Graph representing a **complex system** [AB02]: vertices and edges model the objects and their interactions.

- Emerging properties → **non-trivial** topological properties [AB02; New03]
- Analysis requires **specific** tools → **Network Science** [Bra+13]
- Network Science is a **Data Science** [Cos21]

Notion of Feature-Rich Network

Feature-Rich Complex Networks

- **Limitation** of graphs: encode only a purely **relational** information
- **Extensions** to include various types of additional information:
 - **Directed** edges: asymmetric interactions
 - **Weighted** edges: interaction cost or strength
 - **Signed** edges: antagonistic interactions
 - **Multiplex** graphs: different categories of interactions
 - **Multimodal** graphs: different types of objects
 - **Spatial** graphs: vertices have a position in a metric space
 - **Dynamic** graphs: time represented explicitly
- General description: add **attributes** to vertices or edges

Feature-Rich Network

Umbrella term [Int+19] referring to complex networks possessing edge and/or vertex **attributes**.

- **Interest**: more accurate models
- But work required to adapt/create analysis methods

Section 3

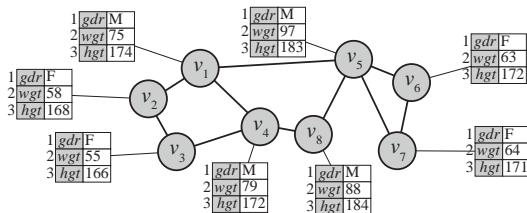
Vertex-Attributed Networks



AVIGNON
UNIVERSITÉ

Vertex-Attributed Networks

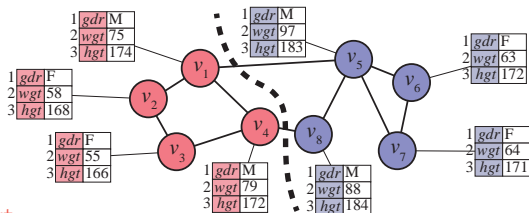
Context



- **Principle:** each vertex is described by a set of attributes
- **Interest:** combines two types of information (individual vs. relational)
- **Contributions:**
 - 1 Community structure **interpretation**, applied to Business Science
 - 2 **Vertex classification**, applied to influence prediction
 - 3 **Graph classification**, applied to abuse detection

Vertex-Attributed Networks

Contribution #1: Community Structure Interpretation 1/2



- Context

- Community structure: partition of the vertex set V , major task
 - Thousands of detection methods... but interpretation?
- Previous work on community detection in plain graphs with G. Orman (MSc) @GSÜ
- Collaboration^{1,2,3} with J.-M. Balasque (Bus. Sci.) @GSÜ, 2009–13

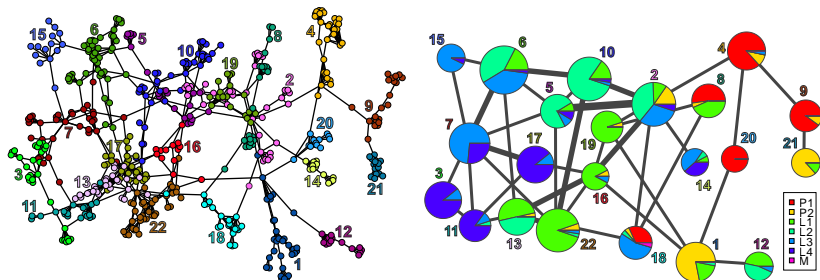
¹V. Labatut and J.-M. Balasque. "Business-oriented Analysis of a Social Network of University Students". In: *International Conference on Advances in Social Networks Analysis and Mining*. 2010, pp. 25–32. DOI: [10.1109/ASONAM.2010.115](https://doi.org/10.1109/ASONAM.2010.115).

²V. Labatut and J.-M. Balasque. "Detection and Interpretation of Communities in Complex Networks: Methods and Practical Application". In: *Computational Social Networks: Tools, Perspectives and Applications*. 2012, pp. 81–113. DOI: [10.1007/978-1-4471-4048-1_4](https://doi.org/10.1007/978-1-4471-4048-1_4).

³V. Labatut and J.-M. Balasque. "Informative Value of Individual and Relational Data Compared Through Business-Oriented Community Detection". In: *The Influence of Technology on Social Network Analysis and Mining*. 2013, pp. 303–330. DOI: [10.1007/978-3-7091-1346-2_13](https://doi.org/10.1007/978-3-7091-1346-2_13).

Vertex-Attributed Networks

Contribution #1: Community Structure Interpretation 2/2



• Highlights

- Data collection: field survey of Galatasaray students
- Compare relational/individual information through partitioning
 - Partial match between clusters vs. communities ($ARI \approx 0.45$)
- Attribute-based interpretation of communities: dominant attributes, LDA, regression
 - Business-actionable results, e.g. P_{15} most likely to use credit
 - Strong differences between communities, no general trend
- What about the homophily assumption?

Vertex-Attributed Networks

Contribution #2: Vertex Classification

- Context:

- Previous work on vertex (unsupervised) classification in plain graphs with N. Dugué & A. Perez (LIFO), 2012–2015
- CLEF 2014 RepLab challenge [Ami+14]:
 - Goal: predict *real-world* influence (rank) based on *online* activity
 - Dataset publicly available
 - All methods from CLEF 2014 focused on *textual* content
- Collaboration^{4,5} with N. Dugué (LIFO) & J.-V. Cossu (LIA), 2014–2016

- Highlights

- Thorough review of the features used in the literature
- Proposition of graph- and text-based custom features
- Main results:
 - Large improvement over state-of-the-art: +14–26%
 - Most important feature: unigrams UaD
 - Graph-based features perform *worse* than baseline

⁴J.-V. Cossu, N. Dugué, and V. Labatut. "Detecting Real-World Influence Through Twitter". In: *2nd European Network Intelligence Conference*. 2015, pp. 83–90. DOI: [10.1109/ENIC.2015.20](https://doi.org/10.1109/ENIC.2015.20).

⁵J.-V. Cossu, V. Labatut, and N. Dugué. "A Review of Features for the Discrimination of Twitter Users: Application to the Prediction of Offline Influence". In: *Social Network Analysis and Mining* 6 (2016), p. 25. DOI: [10.1007/s13278-016-0329-x](https://doi.org/10.1007/s13278-016-0329-x).

Vertex-Attributed Networks

Contribution #3: Graph Classification

- Context:

- Previous work on plain graph (unsupervised) classification with B. Kantarci (MSc) @GSÜ, 2012–2013
- Collaboration^{6,7} with R. Dufour, E. Papégnies, N. Cécillon (MSc, PhD) @LIA, 2015–present
- Industrial project with Castle Prod, providing the dataset
- Objective: automatic moderation of chat rooms

- Highlights:

- Traditional content-based methods sensitive to obfuscation
 - Add graph-based features representing the conversational graph
- Graph- better than text-based classification
- Fusion even better → complementarity
- Feature engineering is costly: use representation learning to generalize?

⁶É. Papégnies, V. Labatut, R. Dufour, and G. Linarès. "Conversational Networks for Automatic Online Moderation". In: *IEEE Transactions on Computational Social Systems* 6.1 (2019), pp. 38–55. DOI: [10.1109/TCSS.2018.2887240](https://doi.org/10.1109/TCSS.2018.2887240).

⁷N. Cécillon, V. Labatut, R. Dufour, and G. Linarès. "Graph embeddings for Abusive Language Detection". In: *Springer Nature Computer Science* 2 (2021), p. 37. DOI: [10.1007/s42979-020-00413-7](https://doi.org/10.1007/s42979-020-00413-7).

Section 4

Dynamic Networks

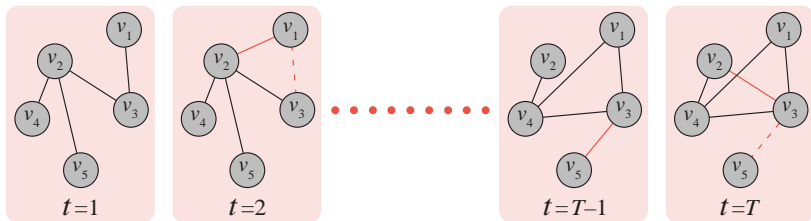


AVIGNON
UNIVERSITÉ

Dynamic Networks

Context

- **Principle:** sequence of graphs, each one modeling a time step
- **Interest:** explicitly represents the dynamics of the system



- **Contributions:**
 - 4 Characterization method based on the evolution of **node neighborhoods**, application to social media
 - 5 Characterization of dynamic **community structures**, application to social media
 - 6 **Segmentation** of dynamic networks, application to the summarization of TV series

Dynamic Networks

Sequential Pattern Mining

- **Context:**
 - Previous work on dynamic networks (WSN) with A. Ozgövde @GSÜ, 2011–2012
 - PhD of G. Orman (Galatasaray/INSA Lyon) co-supervised with J.-F. Boulicaut, M. Plantevit (INSA) 2012–2014⁸
 - Then collaboration⁹ with G. Orman and T. Naskali @GSÜ, 2015–2017
 - **Sequential Pattern Mining:**
 - Important Data Mining subdomain, targeting dynamic *tabular* data
 - In a dataset of sequential objects, look for longest/frequent/emergent/closed **subsequences**
- **Proposition:** adapt SPM to study dynamic networks
 - Propose a sequence-based representation of dynamic graphs
 - Leverage vertex attributes when available

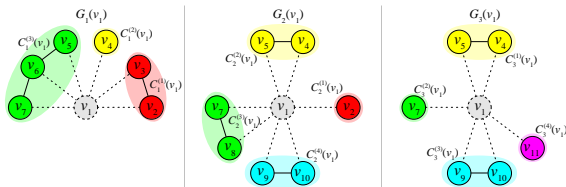
⁸G. K. Orman, V. Labatut, M. Plantevit, and J.-F. Boulicaut. "Interpreting communities based on the evolution of a dynamic attributed network". In: *Social Network Analysis and Mining* 5 (2015), p. 20. DOI: [10.1007/s13278-015-0262-4](https://doi.org/10.1007/s13278-015-0262-4).

⁹G. K. Orman, V. Labatut, and A. T. Naskali. "Exploring the Evolution of Node Neighborhoods in Dynamic Networks". In: *Physica A: Statistical Mechanics and its Applications* 482 (2017), pp. 375–391. DOI: [10.1016/j.physa.2017.04.084](https://doi.org/10.1016/j.physa.2017.04.084).

Dynamic Networks

Contributions #4 & 5: Description of Dynamic Networks

- **Contribution #4:** vertex characterization through sequences of component-related events in ego-networks
 - Detect LFS and CFS, cluster analysis over sequences
 - Application to DBLP, Enron, LastFm



- **Contribution #5:** community characterization through sequence of descriptors (topological measures and vertex attributes)
 - Use community as a descriptor and focus on community-related sequences
 - Detect LFS and emerging sequences
 - Application to DBLP, LastFm

Dynamic Networks

Contribution #6: Segmentation of Dynamic Networks 1/2

- Context:
 - Collaboration^{10,11} with X. Bost (LIA), 2015–present
 - Interdisciplinary work¹² with Communication Studies (CNE)
 - PhD funded by FR Agorantic
 - Objective: character-centered **extractive** video summarization
 - Work on 3 TV serials (HoC, GoT, BB)
 - Traditional methods focus only on multimedia features
- Proposition: model plot to leverage higher level information
 - Model = dynamic conversational network between characters
 - Literature: mainly **static** character networks extracted from **text**
 - Network extraction process:
 - Scene boundaries, speaker diarization, speech turns
 - Weight interactions using **narrative smoothing**

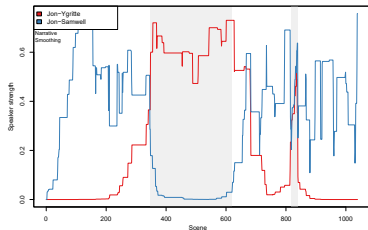
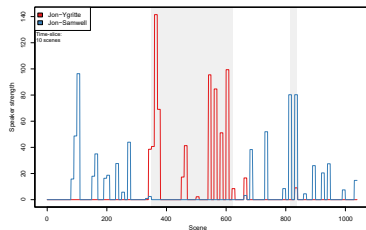
¹⁰X. Bost, V. Labatut, S. Gueye, and G. Linarès. "Extraction and analysis of dynamic conversational networks from TV series". In: *Social Network Analysis and Mining*. 2018, pp. 55–84. DOI: [10.1007/978-3-319-78196-9_3](https://doi.org/10.1007/978-3-319-78196-9_3).

¹¹V. Labatut and X. Bost. "Extraction and Analysis of Fictional Character Networks: A Survey". In: *ACM Computing Surveys* 52.5 (2019), p. 89. DOI: [10.1145/3344548](https://doi.org/10.1145/3344548).

¹²X. Bost, S. Gueye, V. Labatut, M. Larson, G. Linarès, D. Malinas, and R. Roth. "Remembering Winter Was Coming: Character-oriented Video Summaries of TV series". In: *Multimedia Tools and Applications* 78.24 (2019), pp. 35373–35399. DOI: [10.1007/s11042-019-07969-4](https://doi.org/10.1007/s11042-019-07969-4).

Dynamic Networks

Contribution #6: Segmentation of Dynamic Networks 2/2



- Summary extraction:
 - Compute temporal inter-neighborhood similarity, partition to get narrative episodes
 - Select most relevant shots in each episode
 - Musicality, shot size, social relevance
- Results:
 - Evaluation: survey of 187 students familiar with the shows
 - As good (trailer) or better (recap) with graph features
- Ongoing collaboration with X. Bost on plot modeling
 - Story decomposition, narrative classification, recommendation

Section 5

Spatial, Signed, and Multiplex Networks

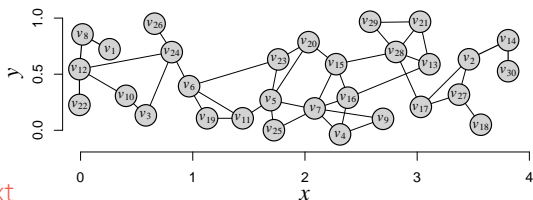


AVIGNON
UNIVERSITÉ

Spatial, Signed, and Multiplex Networks

Spatial Networks

- **Characteristics:** each vertex holds a position in a metric space, each edge has a length
- **Interest:** explicitly model spatial dependencies (eg. transportation system)



- **Context**
 - Previous work with A. Özgövde on WSN @GSÜ, 2011–2012
 - MoMIS project (CNRS PEPS), collaboration¹³ with D. Josselin (UMR Espace) & biologists @Avignon, 2014–2016 + solo work¹⁴
 - **Goal:** compare topology of artificial structures vs. orb-webs

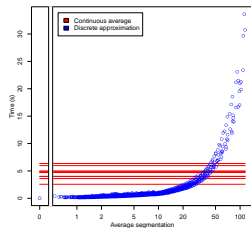
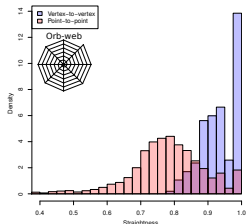
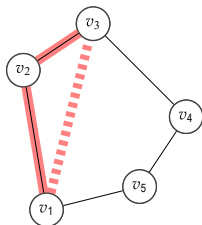
¹³D. Josselin, V. Labatut, and D. Mitsche. "Straightness of rectilinear vs. radio-concentric networks: modeling, simulation and comparison". In: *7th Annual Symposium on Simulation for Architecture and Urban Design*. 2016, pp. 95–102. URL: <http://www.simaud.org/proceedings/index.php%5C#2016>.

¹⁴V. Labatut. "Continuous Average Straightness in Spatial Graphs". In: *Journal of Complex Networks* 6.2 (2018), pp. 269–296. DOI: 10.1093/comnet/cnx033.

Spatial, Signed, and Multiplex Networks

Straightness of Spatial Networks

- Approach centered on the **Straightness** measure

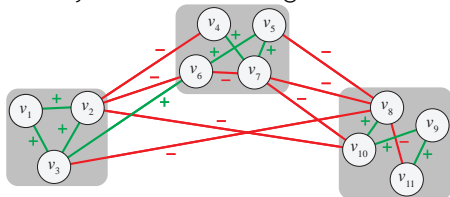


- **Contribution #7:** comparison between orb-webs and geometric networks
 - Center-to-periphery: orb-webs Straightness better if > 8 radii
 - All routes: also better, difference between V2v and P2p
- **Contribution #8:** proposition of measures based on continuous Straightness averaging
 - More accurate, faster, and less memory than approximation
 - Allows building additional measures (e.g. edge centrality)

Spatial, Signed, and Multiplex Networks

Signed Networks

- **Characteristics:** each edge is either positive or negative
- **Interest:** model systems with antagonistic relationships



- **Context**

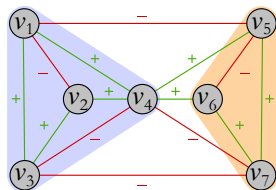
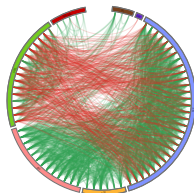
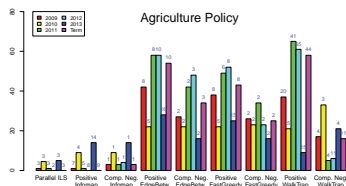
- Collaboration^{15,16} with R. Figueiredo (LIA) 2014–present, N. Arınık (PhD), *et al.*
- Several funded projects: PGM0, LIA, Agorantic, Avignon University
- **Goal:** explore partitioning problems related to **structural balance**, and their resolution methods

¹⁵I. Mendonça, R. Figueiredo, V. Labatut, and P. Michelon. "Relevance of Negative Links in Graph Partitioning: A Case Study Using Votes From the European Parliament". In: *2nd European Network Intelligence Conference*. 2015, pp. 122–129. DOI: 10.1109/ENIC.2015.25.

¹⁶N. Arınık, R. Figueiredo, and V. Labatut. "Multiplicity and Diversity: Analyzing the Optimal Solution Space of the Correlation Clustering Problem on Complete Signed Graphs". In: *Journal of Complex Networks* 8.6 (2020), cnaa025. DOI: 10.1093/comnet/cnaa025.

Spatial, Signed, and Multiplex Networks

Partitioning of Signed Networks

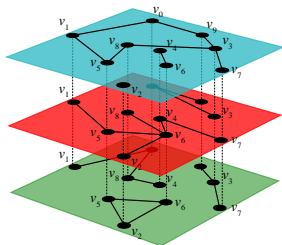


- **Contribution #9:** behavior of signed graph partitioning methods
 - Experiments on a dataset of vote similarity networks (EP)
 - Relevance of negative edges: should be taken into account
 - Effect of edge filtering: faster processing, similar partitions
 - Comparison of CC and RCC: left/right vs. left/right/center
- **Contribution #10:** space of optimal solutions of CC
 - Possibly very large number of optimal solutions
 - Randomly generated complete graphs, with planted partition
 - Typology of 4 possible cases

Spatial, Signed, and Multiplex Networks

Centrality in Multiplex Networks

- **Characteristics:** several layers, same vertices, possibly different edges
- **Interest:** model systems with several types of relations

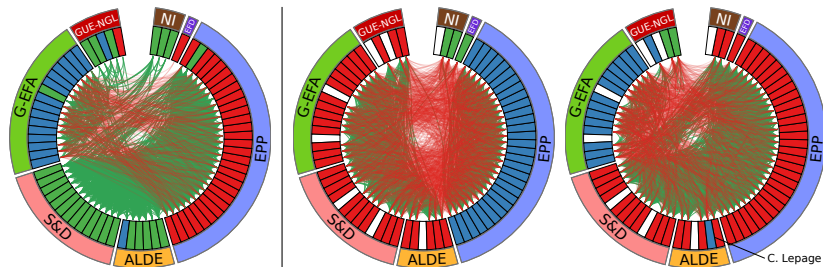


- **Contribution #11:** multiplex centrality measure based on an opinion diffusion model
 - Collaboration¹⁷ with A. Reiffers (LIA), 2015–2016
 - Analogy with users communicating through **several** social media
 - Centrality value = solution of a budget allocation optimization problem
 - Empirical assessment: behavior clearly different from other multiplex measures

¹⁷A. Reiffers and V. Labatut. "Opinion-based centrality in multiplex networks: A convex optimization approach". In: *Network Science* 5.2 (2017), pp. 213–234. DOI: [10.1017/nws.2017.7](https://doi.org/10.1017/nws.2017.7).

Spatial, Signed, and Multiplex Networks

Partitioning of Multiplex Networks



- **Contribution #12:** multiple partitioning of multiplex networks
 - Collaboration¹⁸ with N. Arnik (PhD) and R. Figueiredo (LIA) 2019–21
 - Traditional partitioning approaches look for a **single** partition
 - **Proposition:** partition each layer separately, then cluster analysis
 - Clusters characterized in terms of the concerned layers
 - Application to EP dataset (vote similarity)
 - Uncovering of previously overlooked voting patterns

¹⁸N. Arnik, R. Figueiredo, and V. Labatut. "Multiple partitioning of multiplex signed networks: Application to European parliament votes". In: *Social Networks* 60 (2020), pp. 83–102. DOI: [10.1016/j.socnet.2019.02.001](https://doi.org/10.1016/j.socnet.2019.02.001).

Section 6

Perspectives



AVIGNON
UNIVERSITÉ

Perspectives

Multimodality & Interdisciplinarity

- Multimodality

- Hot topic that connects with my work since the PhD
- Adapt existing methods to leverage additional information
- Solve issues specific to the inclusion of this information
- Is the additional information worth the effort?
- Explore graph embedding methods

- Interdisciplinarity

- Also an aspect of my work since the PhD
- Collaborating with other fields can be challenging
- Great provider of new applications and new problems to solve
- Better interpretation of the results
- Work with social scientists around textual documents

Perspectives

DeCoMaP Project

- ANR project, started in 2020
- **Objective**: Detecting corruption in public procurement markets
- Take advantage of French (BOAMP) and European (TED) open data
- **Interdisciplinary**: includes economists, political scientists, and jurists
- **Multimodal**:
 - Various types of entities: clients, suppliers, directors, contract
 - Described by a number of tabular attributes (id, name, activity domain, location...)
 - Various types of relationships between them: bidding, contracting, hierarchy
 - Some can be weighted, directed, signed, and attributed
 - Bids and contracts have a time stamp
 - Certain entities have a geographic location
- **Problems** of interest: graph partitioning, link prediction

Questions?

Section 7

Appendix

Galatasaray Survey

Properties of the Communities

Com.	#	Homophily		Class 1		Class 2		Dept 1		Hob Val	Phone Val	Plyr Val	Frd avg	Loan avg
		Gdr	Cls	#	Val	#	Val	#	Val					
P_1	32	0.24	0.30	25	P2	6	L1	7	BS	M	No	Ap	3.80	2.60
P_2	39	0.36	0.54	15	L2	12	L3	10	IE	C	No	-	3.00	1.78
P_3	28	0.12	0.00	25	L4	3	L3	12	BS	-	No	-	3.25	2.50
P_4	30	0.11	-0.04	26	P1	3	P2	6	So	M	No	Cr	2.78	1.78
P_5	23	0.19	-0.05	17	L2	-	-	15	RI	S	Sa	Ap	2.75	2.25
P_6	46	0.01	0.65	19	L2	18	L3	14	So	R	Sa	Ap	3.43	1.54
P_7	34	0.25	0.19	25	L3	9	L4	24	BS	C	Sa	Ap	3.78	2.67
P_8	23	0.17	0.74	12	P1	10	L1	9	IE	S	Sa	Ap	3.00	3.00
P_9	20	0.19	0.00	18	P1	2	P2	5	BS	-	No	Ap	2.17	1.67
P_{10}	39	0.51	0.55	31	L2	6	L1	17	BS	C	No	Ap	2.92	1.92
P_{11}	20	-0.15	0.16	17	L4	2	L3	11	IE	M	No	-	3.80	2.00
P_{12}	15	0.11	0.61	7	L1	7	L2	7	La	-	SE	Ap	4.00	1.50
P_{13}	28	0.46	0.60	14	L1	13	L2	12	Ma	M	No	So	3.64	1.64
P_{14}	13	0.00	0.56	6	L3	5	L4	7	CS	-	No	-	3.50	2.00
P_{15}	14	-0.12	0.00	13	L3	-	-	8	RI	-	-	-	2.67	3.50
P_{16}	13	-0.10	-0.03	11	L1	-	-	10	Ph	P	-	-	3.67	1.33
P_{17}	28	0.48	0.00	25	L4	3	L3	14	IE	C	No	Ap	3.11	2.22
P_{18}	22	-0.09	0.54	12	L3	-	-	-	-	-	No	-	3.71	1.17
P_{19}	20	-0.06	0.00	19	L1	-	-	9	Ma	-	No	-	3.71	1.71
P_{20}	12	-0.14	1.00	12	P1	-	-	2	-	R	No	Ap	2.14	1.71
P_{21}	15	-0.16	0.00	13	P2	2	L1	11	La	T	No	-	4.00	1.67
P_{22}	38	-0.05	-0.02	34	L1	2	P2	22	BS	C	Sa	Ap	4.00	2.00
Net.	552	0.25	0.78	124	L1	107	L2	96	BS	S	No	Ap	3.29	1.94

Influence Prediction Classification Performances

Feature and methods					Auto	Bank	Avg
Unigrams	User-as-Document	Separated	All		0.833	0.751	0.792
Unigrams	User-as-Document	Separated	Artex		0.829	0.745	0.787
Unigrams	Bag-of-Tweets	Separated	Artex	Sum	0.820	0.721	0.770
Unigrams	Bag-of-Tweets	Separated	All	Sum	0.817	0.709	0.763
Unigrams	Bag-of-Tweets	Separated	Artex	Count	0.796	0.719	0.757
Unigrams	Bag-of-Tweets	Separated	All	Count	0.786	0.702	0.744
Unigrams	User-as-Document	Joint	All		0.782	0.682	0.732
Unigrams	User-as-Document	Joint	Artex		0.773	0.672	0.722
Ramírez-de-la-Rosa et al. [Ram+14]					0.696	0.693	0.694
Unigrams	Bag-of-Tweets	Joint	All	Count	0.725	0.641	0.683
Unigrams	Bag-of-Tweets	Joint	All	Sum	0.725	0.641	0.683
<i>MF-Baseline</i>					0.500	0.500	0.500
Bigrams	Co-occurrence network				0.403	0.417	0.410

Classification performances ordered by macro-averaged F -Measure. The highest scores are represented in bold. Column *Separated* vs. *Joint* indicates whether tweets were filtered based on language during the classifier training. Column *All* vs. *Artex* indicates whether the Artex methods was used to select only the most relevant tweets during training. Column *Count* vs. *Sum* denotes the voting scheme selected when using the Bag-of-Tweets model.

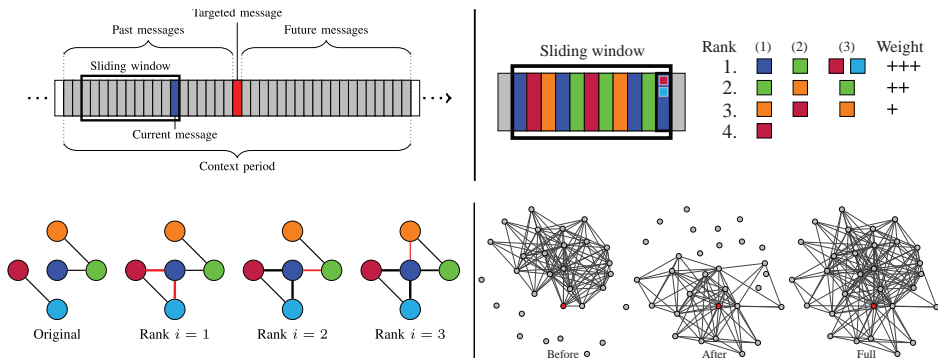
Influence Prediction Ranking Performances

Feature and methods					Auto	Bank	Avg
Unigrams	User-as-Document	Separated	All		0.803	0.626	0.714
Unigrams	Bag-of-Tweets	Separated	All	Sum	0.779	0.628	0.703
Unigrams	Bag-of-Tweets	Separated	Artex	Sum	0.774	0.633	0.703
Unigrams	User-as-Document	Separated	Artex		0.782	0.623	0.702
Unigrams	Bag-of-Tweets	Separated	Artex	Count	0.778	0.612	0.695
Unigrams	Bag-of-Tweets	Separated	All	Count	0.762	0.592	0.677
Unigrams	User-as-Document	Joint	All		0.735	0.538	0.636
Unigrams	User-as-Document	Joint	Artex		0.722	0.547	0.634
Unigrams	Bag-of-Tweets	Joint	All	Sum	0.699	0.526	0.612
Unigrams	Bag-of-Tweets	Joint	All	Count	0.626	0.504	0.565
UTDBRG	Aleahmad et al. [Ale+14]				0.721	0.410	0.565
LyS	Vilares et al. [Vil+14]				0.602	0.524	0.563
Total number of tweets					0.332	0.449	0.385
Best combination of scalar features					0.424	0.338	0.381
<i>RepLab Baseline</i>					0.370	0.385	0.378
Bigrams	Co-occurrence network				0.298	0.300	0.299
Klout score					0.304	0.275	0.289

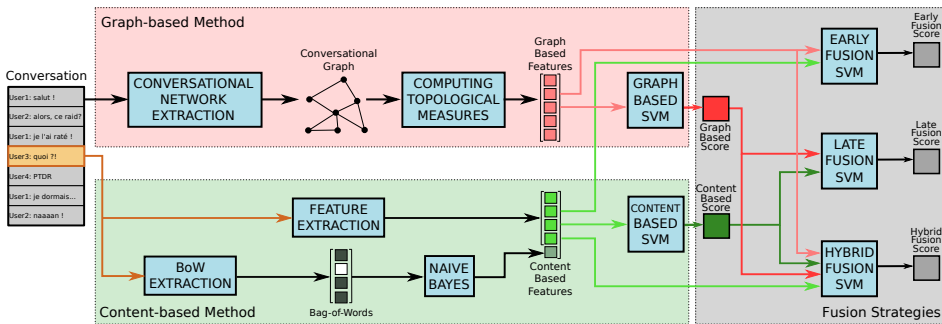
Ranking performances ordered by mean average precision (MAP). The highest scores are represented in bold. Column *Separated* vs. *Joint* indicates whether tweets were filtered based on language during the classifier training. Column *All* vs. *Artex* indicates whether the Artex methods was used to select only the most relevant tweets during training. Column *Count* vs. *Sum* denotes the voting scheme selected when using the Bag-of-Tweets model.

Abuse Detection

Abuse Detection – Conversational Graph



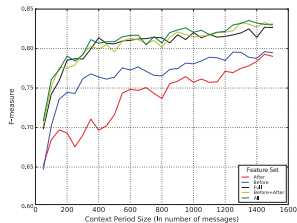
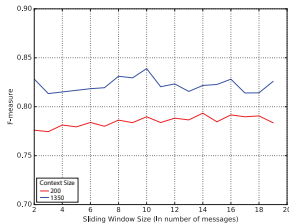
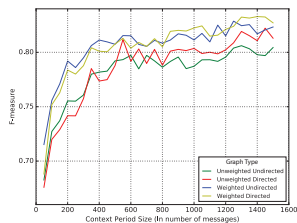
Abuse Detection Processing Pipeline



Representation of our processing pipeline. The green part corresponds to the processing of text-based feature, the red part to graph-related features, and the gray part to the fusion strategies.

Abuse Detection

Effect of the Parameters



Classification performance expressed in F -measure, depending on various extraction parameters. **(a)** as a function of the context period size, for the (un)directed and (un)weighted feature sets. **(b)** as a function of the sliding window size, for 2 context period sizes (200 and 1,350 messages). **(c)** as a function of the context period size, for the 4 considered feature sets (*Before*, *After*, *Full*, *All*) as well as a combination of the first two (*Before+After*).

Abuse Detection

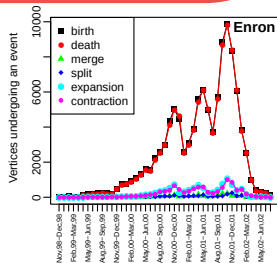
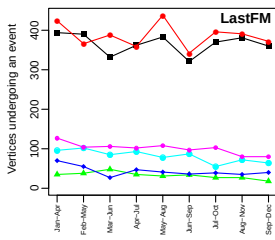
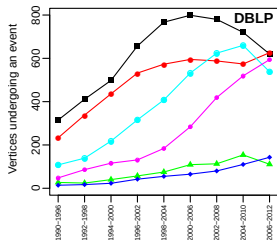
Best Performance

Method	Number of Features	Total Runtime	Average Runtime	Precision	Recall	F -meas
Content-Based	29	0:00:52	0.02s	0.79	0.84	0.81
Content-Based TF	3	0:00:21	0.01s	0.76	0.83	0.79
Graph-Based	459	8:19:10	7.56s	0.90	0.88	0.89
Graph-Based TF	10	0:14:22	0.03s	0.89	0.85	0.87
Early Fusion	488	8:26:41	7.68s	0.91	0.89	0.90
Early Fusion TF	4	0:11:29	0.17s	0.89	0.87	0.88
Late Fusion	488 (2)	8:23:57	7.64s	0.94	0.92	0.93
Late Fusion TF	13	0:15:42	0.24s	0.92	0.90	0.91
Hybrid Fusion	490	8:27:01	7.68s	0.92	0.90	0.91
Hybrid Fusion TF	4	0:16:57	0.26s	0.91	0.89	0.90

Comparison of the performances obtained with our 5 methods, by considering all features and only the Top Features (TF). The total runtime is expressed as *h:min:s*.

Description of Dynamic Networks

Neighborhood Events



Network **Most supported Closed Frequent Sequence s**

Network	Most supported Closed Frequent Sequence s	$ S(s) $	Sup(s)
DBLP	$\langle\langle(B)\rangle\rangle$	2,041	0.950
	$\langle\langle(D)\rangle\rangle$	2,032	0.940
LastFM	$\langle\langle(D)\rangle\rangle$	1,530	0.900
	$\langle\langle(B)\rangle\rangle$	1,510	0.880
Enron	$\langle\langle(B, D)\rangle\rangle$	28,649	0.990
	$\langle\langle(B, D), (B, D)\rangle\rangle$	28,634	0.990

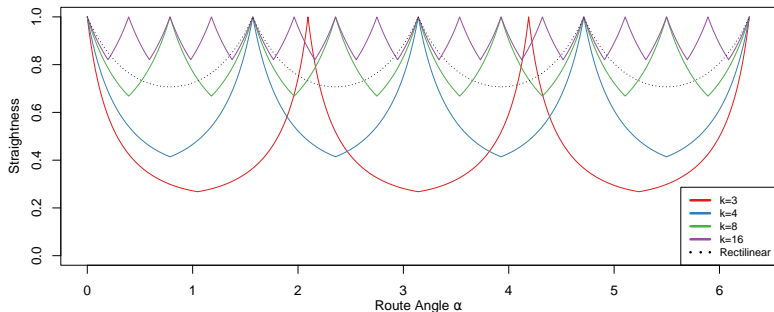
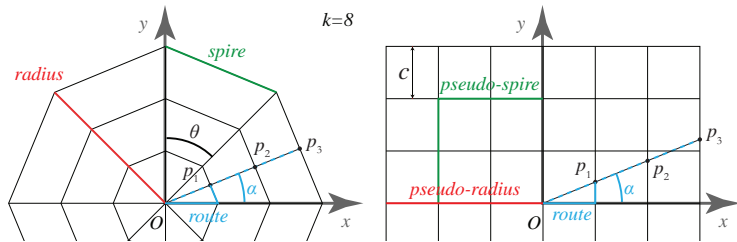
Network **Longest Frequent Sequence s**

Network	Longest Frequent Sequence s	$ S(s) $	Sup(s)
DBLP	$\langle\langle(B), (B), (E), (E), (E), (C), (C), (C)\rangle\rangle$	33	0.010
LastFM	$\langle\langle(B), (D, M, E), (D), (D, C), (C), (C), (C)\rangle\rangle$	14	0.008
Enron	$\langle\langle(B, D), (B), (D), (B), (D), (B), (D), (B), (D), (B), (B), (D), (B, D)\rangle\rangle$	1,266	0.004

Most supported closed frequent sequences (top three rows) and longest frequent sequences (bottom three rows) with their number of supporting vertices ($|S(s)|$) and support (Sup(s)). Events are represented by their initial.

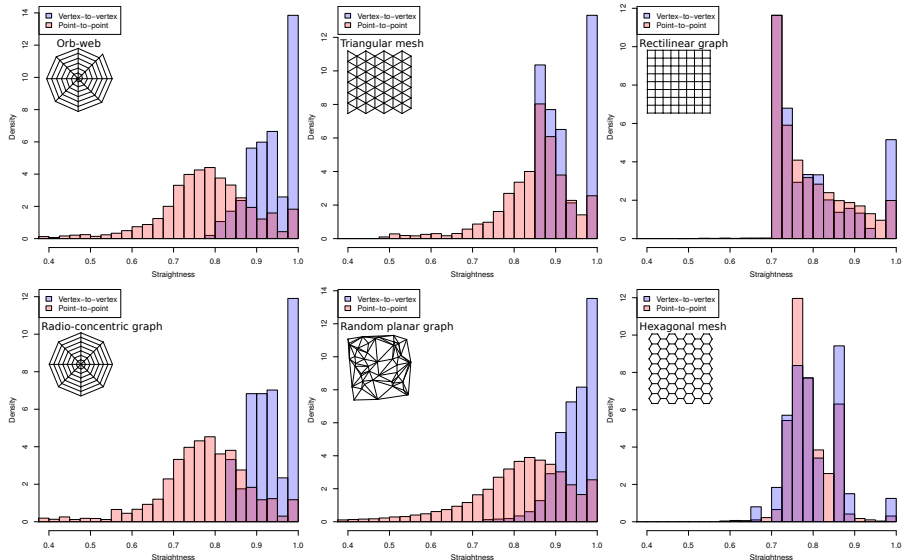
Straightness of Spatial Networks

Center-to-Periphery Routes



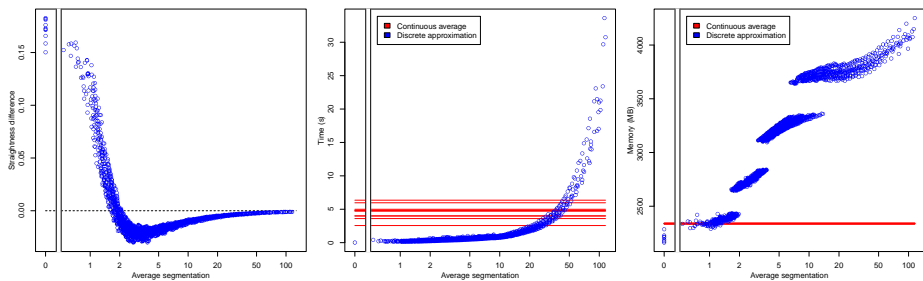
Straightness of Spatial Networks

All Routes



Straightness of Spatial Networks

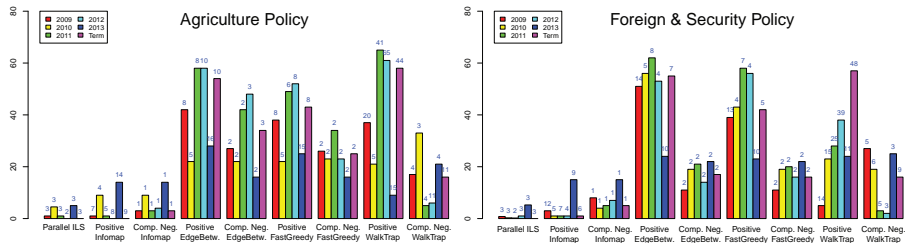
Continuous Average Straightness



Results obtained on random planar graphs, as functions of the average edge segmentation β . **(a)** Difference between the discrete approximation and the continuous Straightness ($\sigma_\beta(G) - S_G(G)$). **(b)** Processing time and **(c)** memory usage required by the continuous average Straightness $S_G(G)$ (red) and its discrete approximation $\sigma_\beta(G)$ (blue).

Partitioning of Signed Networks

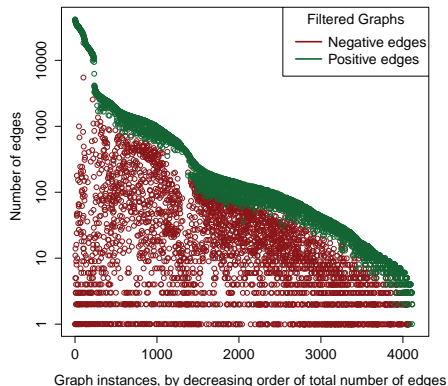
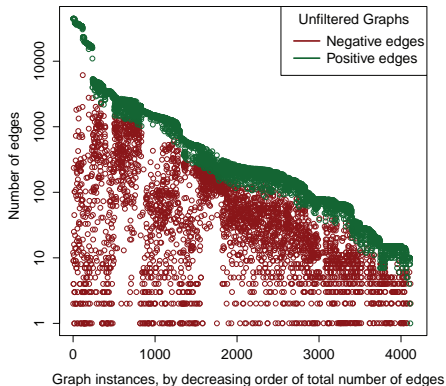
Relevance of Negative Edges



Imbalance of the partitions (bars) and numbers of detected communities (blue values), obtained through Parallel ILS (left bar group) and unsigned community detection methods (other bar groups), for each year and the whole term (see legend), processed for the *Agriculture* (a) and *Foreign & Security* (b) policy domains.

Partitioning of Signed Networks

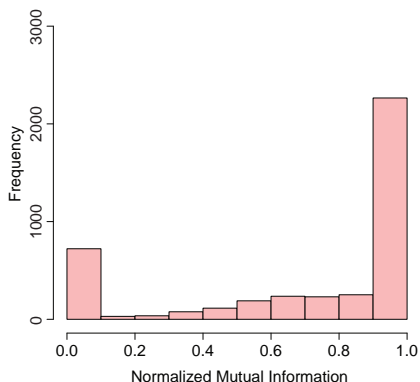
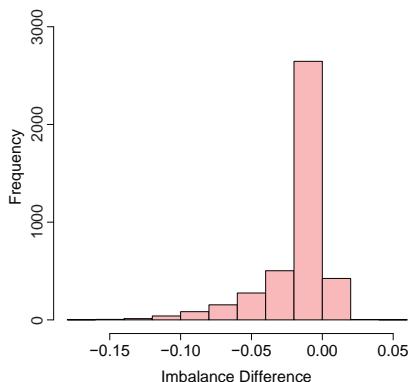
Edge Filtering – Effect on Edges



- (a) Numbers of positive (green) and negative (red) edges in all unfiltered networks.
- (b) Same thing for filtered networks.

Partitioning of Signed Networks

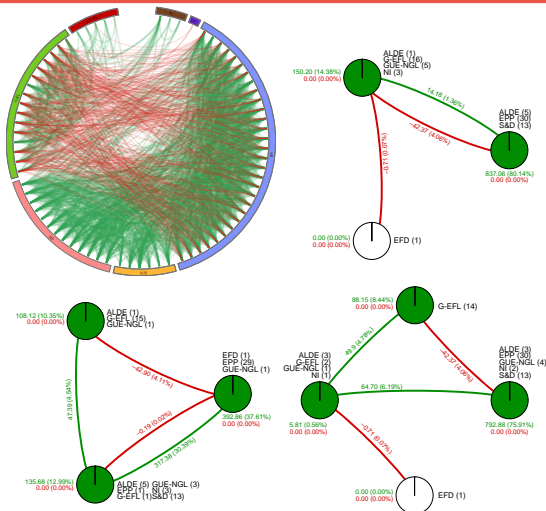
Edge Filtering – Effect on Performance



(a) Distribution of the Imbalance difference between partitions detected by Ex-CC *with vs. without* filtering. **(b)** Distribution of the NMI when comparing the partitions detected by Ex-CC *with vs. without* filtering.

Partitioning of Signed Networks

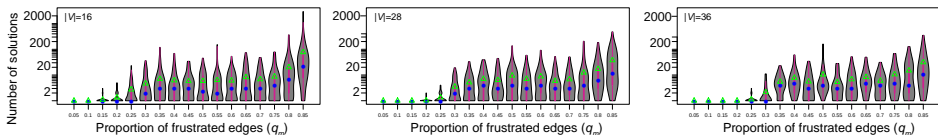
European Parliament – Comparison of Partitioning Methods



Voting similarity network between French MEPs for roll-calls related to agriculture during the year 2012-13 **(a)**, and communities detected using Ex-CC **(b)** and Ex-RCC with $K = 3$ **(c)** and $K = 4$ **(d)**. The outer ring colors in the similarity network match the EP political groups.

Partitioning of Signed Networks

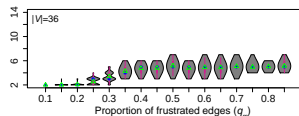
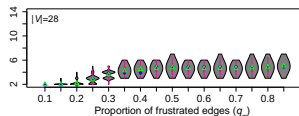
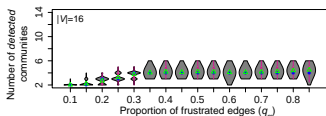
Solution Space – Number of Solutions



Number of solutions as a function of q_m , for $K_0 = 2$. In each violin plot, the interquartile range is shown as a purple thick line, the mean as a green triangle, and the median as a blue dot.

Partitioning of Signed Networks

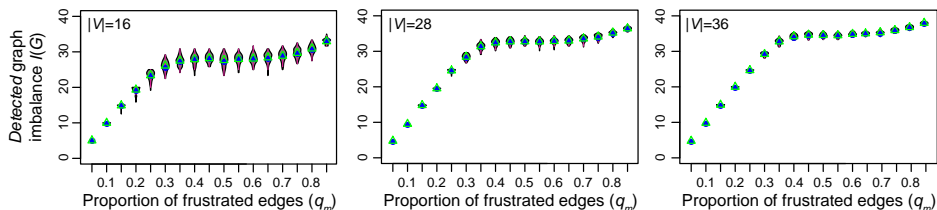
Solution Space – Number of Detected Communities



Number of detected communities as a function of q_m , for $K_0 = 2$.

Partitioning of Signed Networks

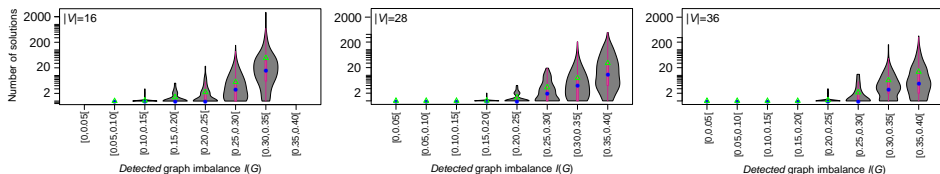
Solution Space – Detected Graph Imbalance



Detected graph imbalance as a function of q_m , for $K_0 = 2$.

Partitioning of Signed Networks

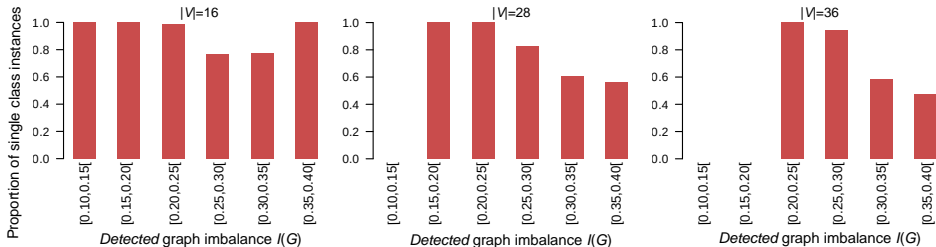
Solution Space – Solutions vs. Detected Imbalance



Number of solutions as a function of the detected imbalance, for $K_0 = 2$. Missing values correspond to undefined parameter values or unavailable data.

Partitioning of Signed Networks

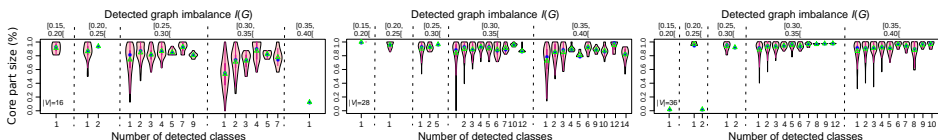
Solution Space – Proportion of Single-Class Instances



Proportion of single-class instances as a function of the detected imbalance, for $K_0 = 2$.
Missing values correspond to undefined parameter values or unavailable data.

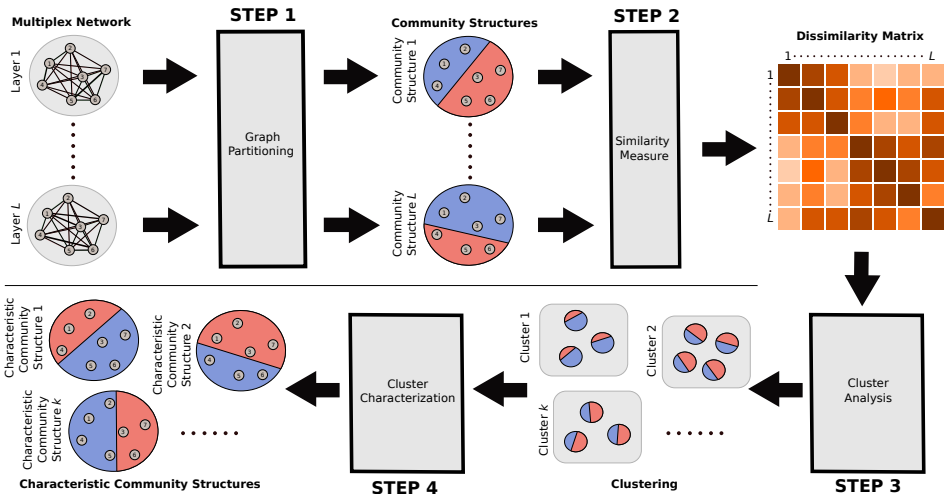
Partitioning of Signed Networks

Solution Space – Coverage of Core Parts



Proportion of the graph covered by the class core parts, as a function of the detected imbalance and number of classes, for $K_0 = 2$. Missing values correspond to undefined parameter values or unavailable data.

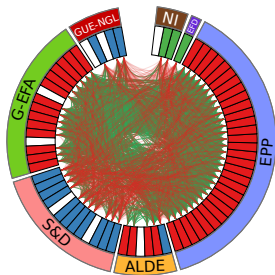
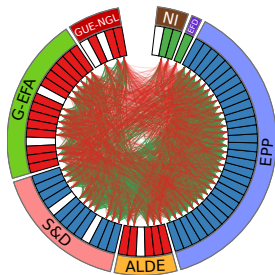
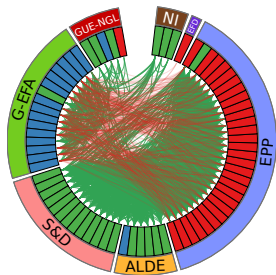
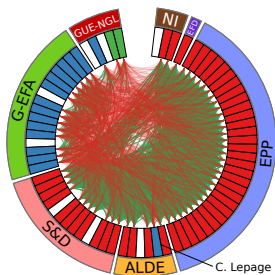
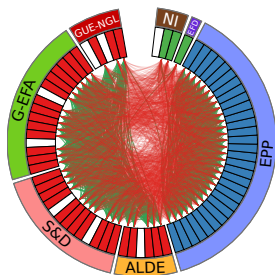
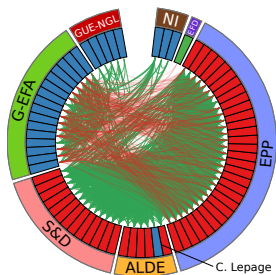
Partitioning of Multiplex Networks Processing Pipeline



General workflow of the proposed partitioning method.

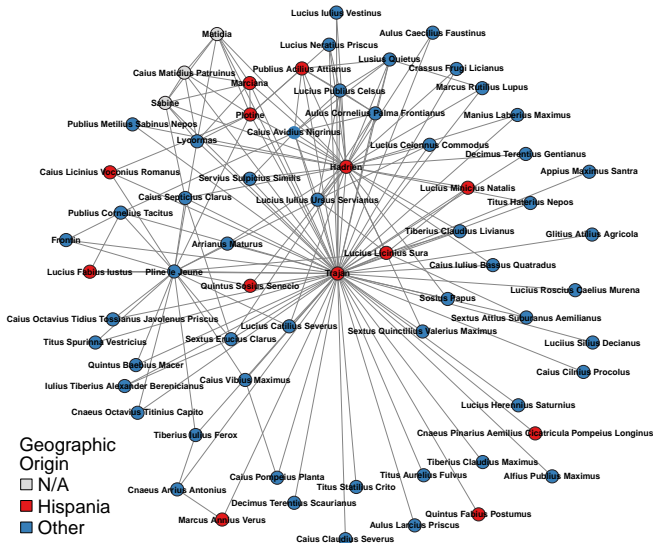
Partitioning of Multiplex Networks

Results on the EP Dataset



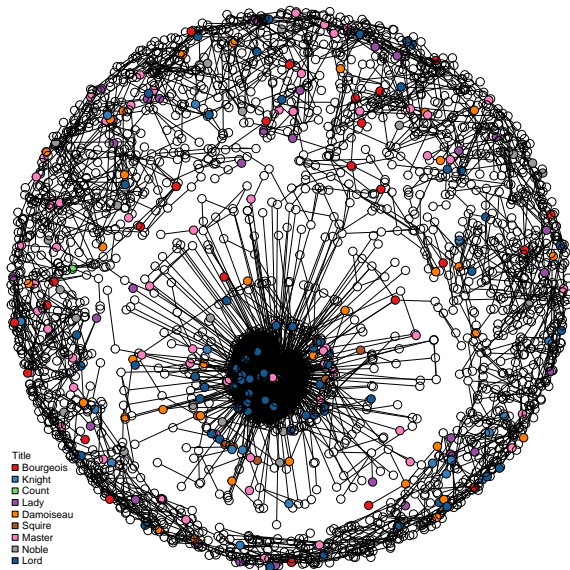
Ongoing Work

Trajan's Social Network



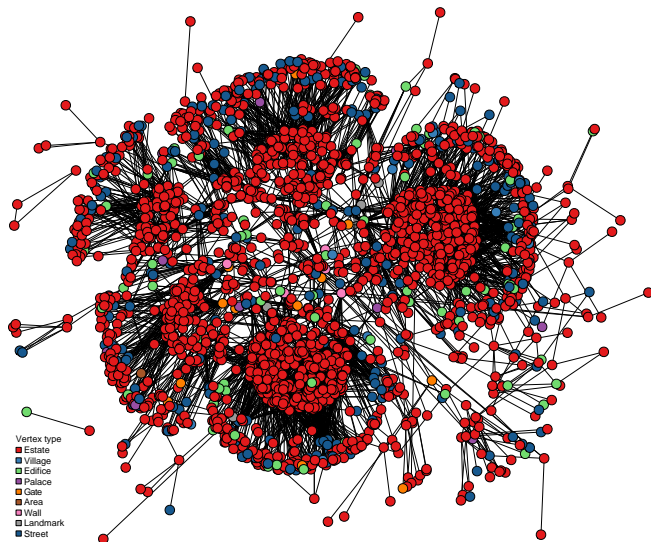
Ongoing Work

Medieval Avignon - Social Network



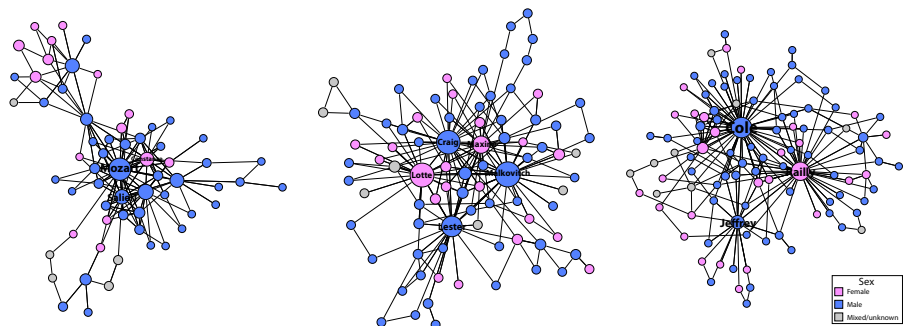
Ongoing Work

Medieval Avignon – Confront Network



Ongoing Work

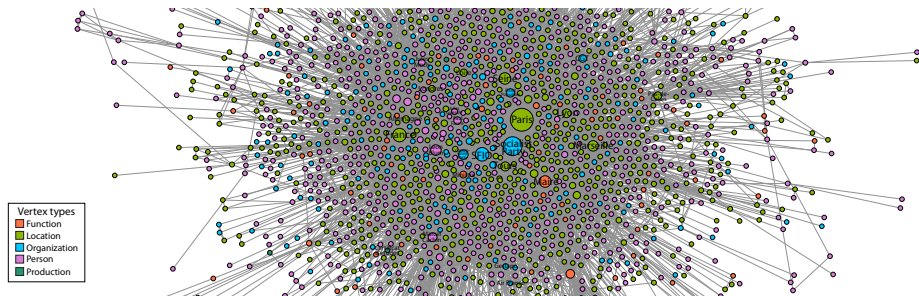
Character Networks Extracted from Scripts



Examples of character interaction networks extracted from movie scripts: **(a)** *Amadeus*, **(b)** *Being John Malkovich*, **(c)** *12 Monkeys*. The displayed vertex attribute corresponds to the estimated gender.

Ongoing Work

Entity Network Extracted from Biographies



Excerpt of the network extracted from the corpus of biographical texts describing socialist members of the French parliament during the inter-war period. Vertices represent named entities, and edges correspond to co-occurrences.

Section 8

References

References I

- [AB02] R. Albert and A.-L. Barabási. "Statistical mechanics of complex networks". In: *Reviews of Modern Physics* 74.1 (2002), pp. 47–96. DOI: [10.1103/RevModPhys.74.47](https://doi.org/10.1103/RevModPhys.74.47).
- [Ale+14] A. Aleahmad, P. Karisani, M. Rahgozar, and F. Oroumchian. "University of Tehran at RepLab 2014". In: *Working Notes for CLEF 2014 Conference*. Vol. 1180. CEUR Workshop Proceedings. 2014, pp. 1528–1536. URL: <http://ceur-ws.org/Vol-1180/CLEF2014wn-Rep-AleahmadEt2014.pdf>.
- [Ami+14] E. Amigó, J. Carrillo-de-Albornoz, I. Chugur, A. Corujo, J. Gonzalo, E. Meij, M. de Rijke, and D. Spina. "Overview of RepLab 2014: Author Profiling and Reputation Dimensions for Online Reputation Management". In: *International Conference of the Cross-Language Evaluation Forum for European Languages*. Vol. 8685. Lecture Notes in Computer Science. Springer, 2014, pp. 307–322. DOI: [10.1007/978-3-319-11382-1_24](https://doi.org/10.1007/978-3-319-11382-1_24).
- [AFL20a] N. Arinik, R. Figueiredo, and V. Labatut. "Multiple partitioning of multiplex signed networks: Application to European parliament votes". In: *Social Networks* 60 (2020), pp. 83–102. DOI: [10.1016/j.socnet.2019.02.001](https://doi.org/10.1016/j.socnet.2019.02.001).
- [AFL20b] N. Arinik, R. Figueiredo, and V. Labatut. "Multiplicity and Diversity: Analyzing the Optimal Solution Space of the Correlation Clustering Problem on Complete Signed Graphs". In: *Journal of Complex Networks* 8.6 (2020), cnaa025. DOI: [10.1093/comnet/cnaa025](https://doi.org/10.1093/comnet/cnaa025).

References II

- [Bar15] A.-L. Barabási. *Network Science*. Cambridge University Press, 2015. URL: <http://barabasi.com/networksciencebook/>.
- [Bos+19] X. Bost, S. Gueye, V. Labatut, M. Larson, G. Linarès, D. Malinas, and R. Roth. "Remembering Winter Was Coming: Character-oriented Video Summaries of TV series". In: *Multimedia Tools and Applications* 78.24 (2019), pp. 35373–35399. DOI: [10.1007/s11042-019-07969-4](https://doi.org/10.1007/s11042-019-07969-4).
- [Bos+18] X. Bost, V. Labatut, S. Gueye, and G. Linarès. "Extraction and analysis of dynamic conversational networks from TV series". In: *Social Network Analysis and Mining*. 2018, pp. 55–84. DOI: [10.1007/978-3-319-78196-9_3](https://doi.org/10.1007/978-3-319-78196-9_3).
- [Bra+13] U. Brandes, G. Robins, A. McRanie, and S. Wasserman. "What is network science?" In: *Network Science* 1.1 (2013), pp. 1–15. DOI: [10.1017/nws.2013.2](https://doi.org/10.1017/nws.2013.2).
- [Céc+21] N. Cécillon, V. Labatut, R. Dufour, and G. Linarès. "Graph embeddings for Abusive Language Detection". In: *Springer Nature Computer Science* 2 (2021), p. 37. DOI: [10.1007/s42979-020-00413-7](https://doi.org/10.1007/s42979-020-00413-7).
- [Cos21] M. Coscia. *The Atlas for the Aspiring Network Scientist*. Michele Coscia, 2021. URL: <https://www.networkatlas.eu/>.

References III

- [CDL15] J.-V. Cossu, N. Dugué, and V. Labatut. "Detecting Real-World Influence Through Twitter". In: *2nd European Network Intelligence Conference*. 2015, pp. 83–90. DOI: [10.1109/ENIC.2015.20](https://doi.org/10.1109/ENIC.2015.20).
- [CLD16] J.-V. Cossu, V. Labatut, and N. Dugué. "A Review of Features for the Discrimination of Twitter Users: Application to the Prediction of Offline Influence". In: *Social Network Analysis and Mining 6* (2016), p. 25. DOI: [10.1007/s13278-016-0329-x](https://doi.org/10.1007/s13278-016-0329-x).
- [Int+19] R. Interdonato, M. Atzmueller, S. Gaito, R. Kanawati, C. Largeron, and A. Sala. "Feature-rich networks: going beyond complex network topologies". In: *Applied Network Science 4* (2019), p. 4. DOI: [10.1007/s41109-019-0111-x](https://doi.org/10.1007/s41109-019-0111-x).
- [JLM16] D. Josselin, V. Labatut, and D. Mitsche. "Straightness of rectilinear vs. radio-concentric networks: modeling, simulation and comparison". In: *7th Annual Symposium on Simulation for Architecture and Urban Design*. 2016, pp. 95–102. URL: <http://www.simaud.org/proceedings/index.php%5C#2016>.
- [Lab18] V. Labatut. "Continuous Average Straightness in Spatial Graphs". In: *Journal of Complex Networks 6.2* (2018), pp. 269–296. DOI: [10.1093/comnet/cnx033](https://doi.org/10.1093/comnet/cnx033).

References IV

- [LB10] V. Labatut and J.-M. Balasque. "Business-oriented Analysis of a Social Network of University Students". In: *International Conference on Advances in Social Networks Analysis and Mining*. 2010, pp. 25–32. DOI: [10.1109/ASONAM.2010.15](https://doi.org/10.1109/ASONAM.2010.15).
- [LB12] V. Labatut and J.-M. Balasque. "Detection and Interpretation of Communities in Complex Networks: Methods and Practical Application". In: *Computational Social Networks: Tools, Perspectives and Applications*. 2012, pp. 81–113. DOI: [10.1007/978-1-4471-4048-1_4](https://doi.org/10.1007/978-1-4471-4048-1_4).
- [LB13] V. Labatut and J.-M. Balasque. "Informative Value of Individual and Relational Data Compared Through Business-Oriented Community Detection". In: *The Influence of Technology on Social Network Analysis and Mining*. 2013, pp. 303–330. DOI: [10.1007/978-3-7091-1346-2_13](https://doi.org/10.1007/978-3-7091-1346-2_13).
- [LB19] V. Labatut and X. Bost. "Extraction and Analysis of Fictional Character Networks: A Survey". In: *ACM Computing Surveys* 52.5 (2019), p. 89. DOI: [10.1145/3344548](https://doi.org/10.1145/3344548).
- [Men+15] I. Mendonça, R. Figueiredo, V. Labatut, and P. Michelon. "Relevance of Negative Links in Graph Partitioning: A Case Study Using Votes From the European Parliament". In: *2nd European Network Intelligence Conference*. 2015, pp. 122–129. DOI: [10.1109/ENIC.2015.25](https://doi.org/10.1109/ENIC.2015.25).

References V

- [New03] M. E. J. Newman. "The structure and function of complex networks". In: *SIAM Review* 45 (2003), pp. 167–256. DOI: [10.1137/S003614450342480](https://doi.org/10.1137/S003614450342480).
- [OLN17] G. K. Orman, V. Labatut, and A. T. Naskali. "Exploring the Evolution of Node Neighborhoods in Dynamic Networks". In: *Physica A: Statistical Mechanics and its Applications* 482 (2017), pp. 375–391. DOI: [10.1016/j.physa.2017.04.084](https://doi.org/10.1016/j.physa.2017.04.084).
- [Orm+15] G. K. Orman, V. Labatut, M. Plantevit, and J.-F. Boulicaut. "Interpreting communities based on the evolution of a dynamic attributed network". In: *Social Network Analysis and Mining* 5 (2015), p. 20. DOI: [10.1007/s13278-015-0262-4](https://doi.org/10.1007/s13278-015-0262-4).
- [Pap+19] É. Papégnies, V. Labatut, R. Dufour, and G. Linarès. "Conversational Networks for Automatic Online Moderation". In: *IEEE Transactions on Computational Social Systems* 6.1 (2019), pp. 38–55. DOI: [10.1109/TCSS.2018.2887240](https://doi.org/10.1109/TCSS.2018.2887240).
- [Ram+14] G. Ramírez-de-la-Rosa, E. Villatoro-Tello, H. Jiménez-Salazar, and C. Sánchez-Sánchez. "Towards Automatic Detection of User Influence in Twitter by Means of Stylistic and Behavioral Features". In: *Mexican International Conference on Artificial Intelligence*. Vol. 8856. Lecture Notes in Computer Science. Springer, 2014, pp. 245–256. DOI: [10.1007/978-3-319-13647-9_23](https://doi.org/10.1007/978-3-319-13647-9_23).

References VI

- [RL17] A. Reiffers and V. Labatut. "Opinion-based centrality in multiplex networks: A convex optimization approach". In: *Network Science* 5.2 (2017), pp. 213–234. DOI: [10.1017/nws.2017.7](https://doi.org/10.1017/nws.2017.7).
- [Say15] H. Sayama. *Introduction to the Modeling and Analysis of Complex Systems*. OpenSUNY, 2015. URL: <http://textbooks.opensuny.org/introduction-to-the-modeling-and-analysis-of-complex-systems/>.
- [Vil+14] D. Vilares, M. Hermo, M. A. Alonso, C. Gómez-Rodríguez, and J. Vilares. "LyS at CLEF RepLab 2014: Creating the State of the Art in Author Influence Ranking and Reputation Classification on Twitter". In: *Working Notes for CLEF 2014 Conference*. Vol. 1180. CEUR Workshop Proceedings. 2014, pp. 1468–1478. URL: <http://ceur-ws.org/Vol-1180/CLEF2014wn-Rep-VilaresEt2014.pdf>.