Tie formation mechanisms in social networks

Examples from bibliometric data

(MPIDR Summer Incubator Program tutorial)

Aliakbar Akbaritabar[1]

[1]Max Planck Institute for Demographic Research, Rostock, Germany;
akbaritabar@demogr.mpg.de

CSS course outline and syllabus (today's topic is one part of it)

- Week 1 – Introduction ... What is "Computational Social Science"? Inductive and deductive research; Big data revolution
- Week 2 – Digital Trace Data ... Observational data; Available vs. designed data; APIs and web scrapping; Representativeness
- Week 3 – Mobility and Migration ... Computational approaches to migration research
- Week 4 – Science of Science ... Robert K. Merton; Sociology of scientific knowledge vs Bibliometrics/ Scientometrics
- Week 5 – Network Analysis ... Tie formation mechanisms in social networks; Violence of independence of observations
- Week 6 – Ethics in Computational Social Science ... Informed consent; Personal data; GDPR
- Week 7 – Mid-semester evaluation ... Short essay and multiple choice questions
- Week 8 – Social Simulation ... Agent-based modelling for social scientists; Micro-Macro link and Coleman's boat
- Week 9 – Text as Data ... Natural Language Processing; Topic modelling; Structural Topic Models
- Week 10 – Open Science and Reproducibility ... Reproducibility crisis; Pre-registration; Version control and Git
- Week 11 – Machine Learning ... Supervised and unsupervised use of observational data; Feature learning
- Week 12 – Other Computational Social Science Skills ... Parallelization; Functional vs Object-Oriented Programming; Graph databases
- Week 13 – Limitations of Computational Social Science ... Pitfalls of digital trace data and computational approaches; Representativeness
- Week 14 – Conclusions ... Thick vs Big data, Survey experiments; Linked data; Future of CSS
- Week 15 – Final semester evaluation ... Short essay and multiple choice questions
- Week 16 – Student presentations ... For those opting for giving a talk instead of writing an essay

Outline

1
2
3
4
5
6
7
8
9
10

Outline

Outline

Outline

Outline

1 Introduction: What is "Tie formation mechanisms in social networks"? Observations by variables vs. relational; Violation of independence
2 Network theory and tie formation mechanisms: Foci of activity; Homophily
3 Network theory and tie formation mechanisms: Reciprocity; Activity; Transitivity
4 Network theory and tie formation mechanisms: Small-World; Popularity and Preferential attachment
5 Empirical analysis and methods: Conditional Uniform Graphs (CUGs); Quadratic Assignment Procedure (QAP) Regression
6 Empirical analysis and methods: Exponential Random Graph Models (ERGMs)
7 Empirical analysis and methods: Stochastic Actor-Oriented Models (SAOMs)
8 Empirical analysis and methods: Stochastic Block-Models (SBMs); Network Simulation with Agent-Based Models (ABMs)
9 Ethics: Ethical considerations in network analysis and modelling tie formation
10 Concluding discussions: Limitations and pitfalls of network data; Future directions

Outline ⇒ What will be covered today

1. Introduction: What is "Tie formation mechanisms in social networks"? Observations by variables vs. relational; Violation of independence
2. Network theory and tie formation mechanisms: Foci of activity; Homophily
3. Network theory and tie formation mechanisms: Reciprocity; Activity; Transitivity
4. Network theory and tie formation mechanisms: Small-World; Popularity and Preferential attachment
5. Empirical analysis and methods: Conditional Uniform Graphs (CUGs); Quadratic Assignment Procedure (QAP) Regression
6. Empirical analysis and methods: Exponential Random Graph Models (ERGMs)
7. Empirical analysis and methods: Stochastic Actor-Oriented Models (SAOMs)
8. Empirical analysis and methods: Stochastic Block-Models (SBMs); Network Simulation with Agent-Based Models (ABMs)
9. Ethics: Ethical considerations in network analysis and modelling tie formation
10. Concluding discussions: Limitations and pitfalls of network data; Future directions

Introduction: A closer look at the course title

1
2
3
4
5
6
7
8
9
10
11
12

Tie formation mechanisms in social networks

Introduction: A closer look at the course title

1
2
3
4
5
6
7
8
9
10
11
12

Tie formation mechanisms in social networks

1
2
3
4
5
6
7
8
9
10
11
12

Tie formation mechanisms in social networks

ID	Name	Age	Political view	Education
1	Tom	24	left	NA
2	Sara	22	right	BA
3	Bill	30	neutral	MA
4	Margaret	31	NA	PhD

Tie formation mechanisms in social networks

$\overbrace{\hspace{20em}}^{\text{Variables/attributes}}$

ID	Name	Age	Political view	Education
1	Tom	24	left	NA
2	Sara	22	right	BA
3	Bill	30	neutral	MA
4	Margaret	31	NA	PhD

Tie formation mechanisms in social networks

A variable by observation table

Variables/attributes

Respondents/ Observations

ID	Name	Age	Political view	Education
1	Tom	24	left	NA
2	Sara	22	right	BA
3	Bill	30	neutral	MA
4	Margaret	31	NA	PhD

Tie formation mechanisms in social networks

What if respondents know each other?

Variables/attributes

Respondents/ Observations

ID	Name	Age	Political view	Education
1	Tom	24	left	NA
2	Sara	22	right	BA
3	Bill	30	neutral	MA
4	Margaret	31	NA	PhD

Introduction: A closer look at the course title

1 2 3 4 5 6 7 8 9 10 11 12

Tie formation mechanisms in social networks

Different contexts of familiarity

Variables/attributes

Respondents/ Observations

ID	Name	Age	Political view	Education
1	Tom	24	left	NA
2	Sara	22	right	BA
3	Bill	30	neutral	MA
4	Margaret	31	NA	PhD

Tie formation mechanisms in social networks

Violation of independence of observations

Respondents/ Observations (columns)

Respondents/ Observations (rows)

	Tom	Sara	Bill	Margaret
Tom	-	Instagram celebrity	Is not sure remembers	NA
Sara	NA	-	Brother	Brother's crush
Bill	(Former) same gym member	Sister	-	MA classmate / friend / smartest batchmate
Margaret	NA	NA	MA classmate	-

1 2 3 4 5 6 7 8 9 10 11 12

Tie formation mechanisms in social networks

Adjacency (familiarity) matrix

Respondents/ Observations

	Tom	Sara	Bill	Margaret
Tom	-	1	0	0
Sara	0	-	1	1
Bill	1	1	-	1
Margaret	0	0	1	-

Respondents/ Observations

Tie formation mechanisms in social networks

Network mechanisms are local/global configurations that allow us to explain the presence of a tie in a network (the output)[1,2].

[1] Stadtfeld, C., & Amati, V. (2021). Network mechanisms and network models. Research Handbook on Analytical Sociology. https://doi.org/10.4337/9781789906851.00032

[2] Fuhse, J. A., & Gondal, N. (2022). Networks from culture: Mechanisms of tie-formation follow institutionalized rules in social fields. Social Networks. https://doi.org/10.1016/j.socnet.2021.12.005

Tie formation mechanisms in social networks

Network mechanisms are local/global configurations that allow us to explain the presence of a tie in a network (the output)[1,2].

$\Downarrow \Downarrow$

Foci of activity; Homophily; Reciprocity; Activity; Transitivity; Popularity and Preferential attachment

[1] Stadtfeld, C., & Amati, V. (2021). Network mechanisms and network models. Research Handbook on Analytical Sociology. https://doi.org/10.4337/9781789906851.00032

[2] Fuhse, J. A., & Gondal, N. (2022). Networks from culture: Mechanisms of tie-formation follow institutionalized rules in social fields. Social Networks. https://doi.org/10.1016/j.socnet.2021.12.005

Introduction: A closer look at the course title

Tie formation mechanisms in social networks

The research focused on social network tie formation is carried out with different:

Theoretical concepts[1]

- Social influence and social selection in tie formation
- (endogenous) Network self-organization
- (endogenous) Actor attributes
- (exogenous) Contextual factors

Methods

- Quantitative/computational
- Qualitative
- Mixed methods, etc.

Units of analysis/approaches

- Sociocentric
- Egocentric
- Multi-level

Data sources

- Secondary/Re-purposed
- Survey & Interview
- Social media
- Other digital traces
- Spatial/sensors/wearables

In various disciplines

- Social Sciences
- Natural Sciences
- etc.

[1] Lusher et al. (2013) Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications: p24

1 2 3 4 5 6 7 8 9 10 11 12

Tie formation mechanisms in social networks

The research focused on social network tie formation is carried out with different:

Theoretical concepts[1]

- Social influence and social selection in tie formation
- (endogenous) Network self-organization
- (endogenous) Actor attributes
- (exogenous) Contextual factors

Methods

Quantitative/computational
Qualitative
Mixed methods, etc.

Units of analysis/approaches

Sociocentric
Egocentric
Multi-level

Data sources

Secondary/Re-purposed
Survey & Interview
Social media
Other digital traces
Spatial/sensors/wearables

In various **disciplines**

Social Sciences
Natural Sciences
etc.

[1] Lusher et al. (2013) Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications: p24

Introduction: What to expect in the actual course

1 2 3 4 5 6 7 8 9 10 11 12

Tie formation mechanisms in social networks

Further topics to cover:

- Observations by variables vs. relational approach
- Violation of independence of observations
- Context of interactions; Sociocentric vs. egocentric network analysis
- Centrality and importance of nodes
- Cohesive subgroups and community detection
- Social capital; resources embedded in social ties
- Visualizing; Simulating networks; Statistical analysis of networks

1
2
3
4
5
6
7
8
9
10
11
12

Tie formation mechanisms in social networks

Further topics to cover:

- Observations by variables vs. relational approach
- Violation of independence of observations
- Context of interactions; Sociocentric vs. egocentric network analysis
- Centrality and importance of nodes
- Cohesive subgroups and community detection
- Social capital; resources embedded in social ties
- Visualizing; Simulating networks; Statistical analysis of networks

Key reading

Lusher, et al. (2013). Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications.

Tie formation mechanisms in social networks

Further topics to cover:

- Observations by variables vs. relational approach
- Violation of independence of observations
- Context of interactions; Sociocentric vs. egocentric network analysis
- Centrality and importance of nodes
- Cohesive subgroups and community detection
- Social capital; resources embedded in social ties
- Visualizing; Simulating networks; Statistical analysis of networks

Key reading

Lusher, et al. (2013). Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications.

Additional readings

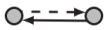
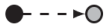
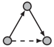
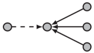
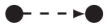
- Nooy, W. D., Mrvar, A., & Batagelj, V. (2018). Exploratory Social Network Analysis with Pajek.
- Stadtfeld, C., & Amati, V. (2021). Network mechanisms and network models. https://doi.org/10.4337/9781789906851.00032
- Kolaczyk, E. D., & Csárdi, G. (2020). Statistical Analysis of Network Data with R.
- Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. https://doi.org/10.1126/science.1165821
- Cioffi-Revilla, C. (2017). Introduction to Computational Social Science. https://doi.org/10.1007/978-3-319-50131-4
- Borgatti, S. P., & Halgin, D. S. (2011). On Network Theory. https://doi.org/10.1287/orsc.1100.0641
- Palla, G., Barabási, A.-L., & Vicsek, T. (2007). Quantifying social group evolution. https://doi.org/10.1038/nature05670
- Akbaritabar, A., Traag, V. A., Caimo, A., & Squazzoni, F. (2020). Italian Sociologists: A Community of Disconnected Groups. https://doi.org/10.1007/s11192-020-03555-w
- Stadtfeld, C. (2018). The Micro-Macro Link in Social Networks. https://doi.org/10.1002/9781118900772.etrds0463

Recap of topics 2-4: Network theory; Tie formation mechanisms

1
2
3
4
5
6
7
8
9
10
11
12

Recap of topics 2-4: Network theory; Tie formation mechanisms

Table 23.1 Most common network mechanisms and their representation. Dashed ties are the explained ties. Continuous ties and black nodes represent elements defining the structural position of the network mechanism

Network mechanism	Representation	Network mechanism	Representation
Reciprocity		Activity	
Transitivity		Attraction	
Popularity		Homophily	

(Tie formation mechanisms Stadtfeld & Amati 2021)

Table 23.1 Most common network mechanisms and their representation. Dashed ties are the explained ties. Continuous ties and black nodes represent elements defining the structural position of the network mechanism

Network mechanism	Representation	Network mechanism	Representation
Reciprocity		Activity	
Transitivity		Attraction	
Popularity		Homophily	

(Tie formation mechanisms
Stadtfeld & Amati 2021)

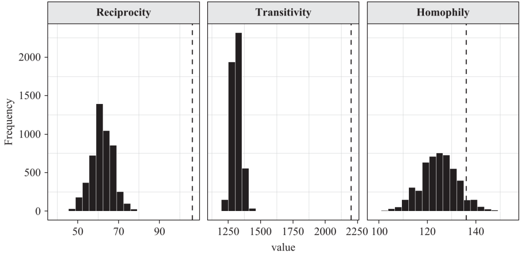


Figure 23.2 Distributions of mutual and homophilous dyads, and transitive triads obtained by sampling 5 000 networks from a CUG conditional on the observed number of advice ties of Lazega's data. The dashed line represents the observed values. The empirical p-value is the proportion of simulated values greater than or equal to the corresponding observed values

(Conditional Uniform Graphs (CUGs)
Stadtfeld & Amati 2021)

Topic 7: Learning objectives

Goal: To learn about a new empirical method and its application for network analysis to study tie formation.

The specific case of Exponential Random Graph Models (ERGMs).

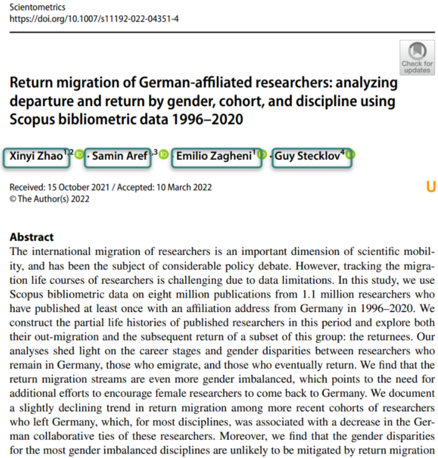
You will learn what the bibliometric data looks like. How to use it to construct co-authorship networks.

In the lab session, we will look at some exercises in Python and R on how to handle the data and model it using ERGMs.

A simple and scalable idea:

Co-presence in the list of publication's authors can be used to infer collaboration ties for individual scholars and for populations

Constructing co-authorship edge-list and network (1)


Scientometrics
https://doi.org/10.1007/s11192-022-04351-4

Check for updates

Return migration of German-affiliated researchers: analyzing departure and return by gender, cohort, and discipline using Scopus bibliometric data 1996–2020

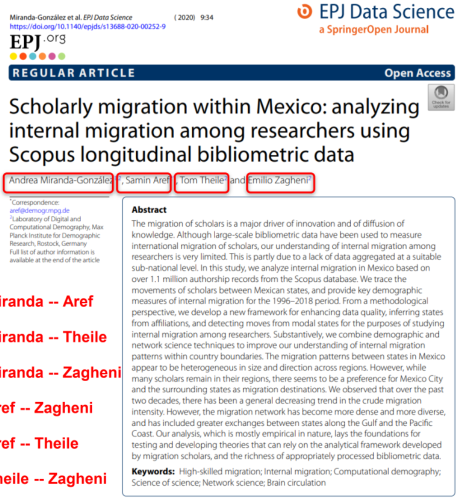
Xinyi Zhao[1,2] · Samin Aref[1,3] · Emilio Zagheni[1] · Guy Stecklov[4]

Received: 15 October 2021 / Accepted: 10 March 2022

Abstract
The international migration of researchers is an important dimension of scientific mobility, and has been the subject of considerable policy debate. However, tracking the migration life courses of researchers is challenging due to data limitations. In this study, we use Scopus bibliometric data on eight million publications from 1.1 million researchers who have published at least once with an affiliation address from Germany in 1996–2020. We construct the partial life histories of published researchers in this period and explore both their out-migration and the subsequent return of a subset of this group: the returnees. Our analyses shed light on the career stages and gender disparities between researchers who remain in Germany, those who emigrate, and those who eventually return. We find that the return migration streams are even more gender imbalanced, which points to the need for additional efforts to encourage female researchers to come back to Germany. We document a slightly declining trend in return migration among more recent cohorts of researchers who left Germany, which, for most disciplines, was associated with a decrease in the German collaborative ties of these researchers. Moreover, we find that the gender disparities for the most gender imbalanced disciplines are unlikely to be mitigated by return migration

Undirected Edge-lists

Zhao -- Aref	Miranda -- Aref
Zhao -- Zagheni	Miranda -- Theile
Zhao -- Stecklov	Miranda -- Zagheni
Aref -- Zagheni	Aref -- Zagheni
Aref -- Stecklov	Aref -- Theile
Zagheni -- Stecklov	Theile -- Zagheni


Miranda-González et al. EPJ Data Science (2020) 9:34
https://doi.org/10.1140/epjds/s13688-020-00252-9
EPJ.org
EPJ Data Science
a SpringerOpen Journal
REGULAR ARTICLE Open Access

Check for updates

Scholarly migration within Mexico: analyzing internal migration among researchers using Scopus longitudinal bibliometric data

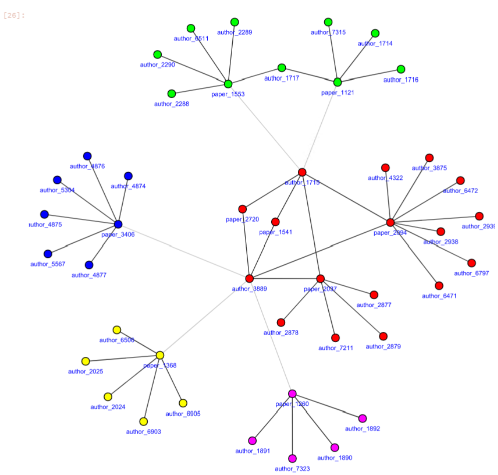
Andrea Miranda-González[2], Samin Aref[*], Tom Theile[2] and Emilio Zagheni[2]

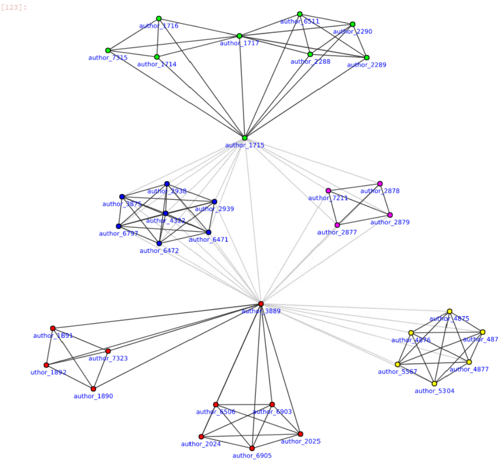
*Correspondence: aref@demogr.mpg.de
[2]Laboratory of Digital and Computational Demography, Max Planck Institute for Demographic Research, Rostock, Germany
Full list of author information is available at the end of the article

Abstract
The migration of scholars is a major driver of innovation and of diffusion of knowledge. Although large-scale bibliometric data have been used to measure international migration of scholars, our understanding of internal migration among researchers is very limited. This is partly due to a lack of data aggregated at a suitable sub-national level. In this study, we analyze internal migration in Mexico based on over 1.1 million authorship records from the Scopus database. We trace the movements of scholars between Mexican states, and provide key demographic measures of internal migration for the 1996–2018 period. From a methodological perspective, we develop a new framework for enhancing data quality, inferring states from affiliations, and detecting moves from modal states for the purposes of studying internal migration among researchers. Substantively, we combine demographic and network science techniques to improve our understanding of internal migration patterns within country boundaries. The migration patterns between states in Mexico appear to be heterogeneous in size and direction across regions. However, while many scholars remain in their regions, there seems to be a preference for Mexico City and the surrounding states as migration destinations. We observed that over the past two decades, there has been a general decreasing trend in the crude migration intensity. However, the migration network has become more dense and more diverse, and has included greater exchanges between states along the Gulf and the Pacific Coast. Our analysis, which is mostly empirical in nature, lays the foundations for testing and developing theories that can rely on the analytical framework developed by migration scholars, and the richness of appropriately processed bibliometric data.

Keywords: High-skilled migration; Internal migration; Computational demography; Science of science; Network science; Brain circulation

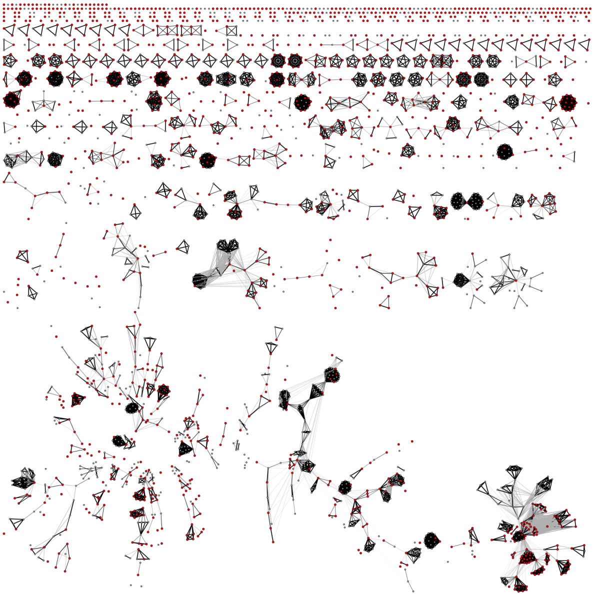
Constructing co-authorship edge-list and network (2)






```
[27]: ig.plot(g1)
```

Constructing co-authorship edge-list and network (3)



Source of this data:
Bibliometric data is accumulated by some companies such as Clarivate (Web of Science), Elsevier (Scopus), Digital-Science (Dimensions), etc.

Here we used Scopus data.

Breaknig the network down to its elements (1)

1 2 3 4 5 6 7 8 9 10 11 12

Table 2 Gender and sectors composition and internationality of members of the communities detected from the giant component (Percentages are calculated by rows for each community separately for gender, country and sectors)

		Gender			Country				Scientific disciplinary sectors (SPS)						
ID	# member	F (%)	M (%)	Missing G (%)	EU (%)	IT (%)	Other (%)	Missing C (%)	07 (%)	08 (%)	09 (%)	10 (%)	11 (%)	Postdoc (%)	Missing S (%)
0	254	43	54	3	54	29	11	5	1	5	0	0	0	2	91
1	142	50	49	1	36	55	6	3	6	3	8	1	1	2	78
2	122	38	61	1	37	56	3	4	10	1	7	0	1	5	76
3	103	45	54	1	41	44	5	11	4	2	12	1	0	2	80
4	91	47	49	3	32	57	9	2	7	7	0	1	2	1	82

Breaknig the network down to its elements (1)

1 2 3 4 5 6 7 8 9 10 11 12

Table 2 Gender and sectors composition and internationality of members of the communities detected from the giant component (Percentages are calculated by rows for each community separately for gender, country and sectors)

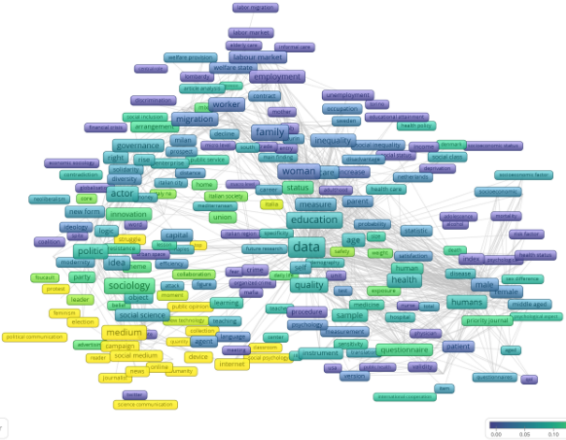
		Gender			Country				Scientific disciplinary sectors (SPS)						
ID	# member	F (%)	M (%)	Missing G (%)	EU (%)	IT (%)	Other (%)	Missing C (%)	07 (%)	08 (%)	09 (%)	10 (%)	11 (%)	Postdoc (%)	Missing S (%)
0	254	43	54	3	54	29	11	5	1	5	0	0	0	2	91
1	142	50	49	1	36	55	6	3	6	3	8	1	1	2	78
2	122	38	61	1	37	56	3	4	10	1	7	0	1	5	76
3	103	45	54	1	41	44	5	11	4	2	12	1	0	2	80
4	91	47	49	3	32	57	9	2	7	7	0	1	2	1	82

Breaknig the network down to its elements (1)

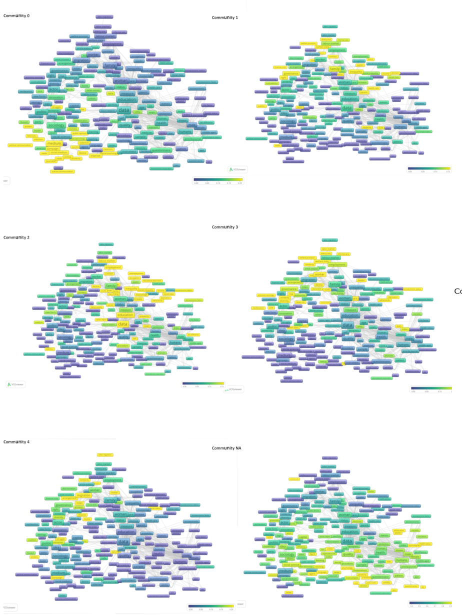
Table 2 Gender and sectors community separately for ge

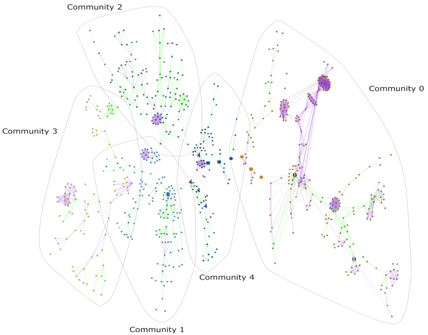
		Gender
ID	# member	F (%)
0	254	43
1	142	50
2	122	38
3	103	45
4	91	47

- 65% foreigners
- Medium, science communication, social medium, internet, political communication & public opinion



Breaknig the network down to its elements (2)





Wait a moment!

1 2 3 4 5 6 7 8 9 10 11 12

Which one of these reasons are driving the collaboration ties to be formed (or not) and to what extent?

ERGMs to check all these reasons for tie formation (1)

1 2 3 4 5 6 7 8 9 10 11 12

The use of an ERGM is consistent with some basic theoretical assumptions about social networks[2]:

- Social networks are locally emergent.
- Network ties not only self-organize (i.e., there are dependencies between ties), but they are also influenced by actor attributes and other exogenous factors.
- The patterns within networks can be seen as evidence for ongoing structural processes.
- Multiple processes can operate simultaneously.
- Social networks are structured, yet stochastic.

---

[2] Lusher et al. (2013) Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications: p10; p43

ERGMs to check all these reasons for tie formation (1)

The use of an ERGM is consistent with some basic theoretical assumptions about social networks[2]:

- Social networks are locally emergent.
- Network ties not only self-organize (i.e., there are dependencies between ties), but they are also influenced by actor attributes and other exogenous factors.
- The patterns within networks can be seen as evidence for ongoing structural processes.
- Multiple processes can operate simultaneously.
- Social networks are structured, yet stochastic.

The ERGM probability is formally defined as

$$P(X=x)=\frac{\exp\left(\sum_k \beta_k s_k(x)\right)}{\sum_{x'\in\mathcal{X}}\exp\left(\sum_k \beta_k s_k(x')\right)}.$$

The probability is given for an empirically observed network x. Functions s_k are typically count statistics that evaluate the number of, for example, ties, reciprocal or homophilous ties, transitive structures, or in-degrees. They are thus similar to structural positions of network mechanisms after the inclusion of the explained tie. Coefficients β are associated with these structures and are subject to statistical estimation. If a graph has, for instance, a higher number

(ERGMs formula
Stadtfeld & Amati 2021)

[2] Lusher et al. (2013) Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications: p10; p43

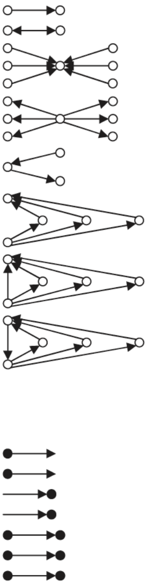
ERGMs to check all these reasons for tie formation (1)

1
2
3
4
5
6
7
8
9
10
11
12

The use of an ERGM is consistent with some basic theoretical assumptions about social networks[2]:

- Social networks are locally emergent.
- Network ties not only self-organize (i.e., there are dependencies between ties), but they are also influenced by actor attributes and other exogenous factors.
- The patterns within networks can be seen as evidence for ongoing structural processes.
- Multiple processes can operate simultaneously.
- Social networks are structured, yet stochastic.

Network effect
Purely structural effects (endogenous)
Arc
Reciprocity
Popularity (in-degree)
Activity (out-degree)
Simple 2-path[3]
Multiple 2-paths
Transitivity (transitive path closure of multiple 2-paths)
Cyclic closure (cyclic closure of multiple 2-paths)
Actor relation effects (exogenous) (black nodes indicates actor with attribute)
Sender (seniority)
Sender (projects)
Receiver (seniority)
Receiver (projects)
[4] Homophily (seniority)
Heterophily (projects)
Homophily (office)



[2] Lusher et al. (2013) Exponential Random Graph Models for Social Networks: Theory, Methods, and Applica

ERGMs to check all these reasons for tie formation (2)

	The giant component of Italian sociologists and their coauthors			
	ERGM models			
	(1)	(2)	(3)	(4)
Ties	−4.551*** (0.023)	−3.654*** (0.187)	−13.140*** (2.058)	−13.367*** (2.522)
Preferential attachment	15.224*** (4.697)			4.069*** (1.152)
Within females ties		−0.250 (0.169)	−0.042 (0.176)	−0.097 (0.259)
Within males ties		0.547*** (0.177)	0.357* (0.185)	0.385 (0.286)
Males main effect		−0.257 (0.164)	−0.086 (0.171)	−0.125 (0.267)
Within community 0			10.472*** (2.049)	10.771*** (2.487)
Within community 1			6.448*** (1.295)	6.513*** (1.470)
Within community 2			6.413*** (1.296)	6.064*** (1.567)
Within community 3			7.228*** (1.502)	7.484*** (1.747)
Within community 4			6.001*** (1.297)	5.986*** (1.430)
Community 1 main effect			2.034 (1.255)	2.153 (1.488)
Community 2 main effect			2.126* (1.254)	2.451 (1.582)
Community 3 main effect			1.809 (1.304)	1.837 (1.611)
Community 4 main effect			2.509** (1.252)	2.675* (1.407)
Within Europe		0.770*** (0.102)	0.840*** (0.105)	0.849*** (0.148)
Within Italy		1.054*** (0.125)	0.877*** (0.129)	0.872*** (0.178)
Within other countries		1.956*** (0.233)	1.842*** (0.241)	1.857*** (0.356)
Italy main effect		−0.456*** (0.099)	−0.194* (0.103)	−0.186 (0.138)
Other countries main effect		−0.438*** (0.102)	−0.465*** (0.105)	−0.482*** (0.151)
Difference in total pubs		0.060*** (0.002)	0.064*** (0.002)	0.065*** (0.004)
Difference in first pub		−0.090*** (0.006)	−0.095*** (0.007)	−0.096*** (0.009)
Difference in last pub		−0.372*** (0.014)	−0.367*** (0.014)	−0.378*** (0.020)
Akaike Inf. Crit.	25,238.680	22,710.210	16,466.530	16,344.340
Bayesian Inf. Crit.	25,270.010	22,835.510	16,685.800	16,584.500

Discussion and Hands-on lab session

1 2 3 4 5 6 7 8 9 10 11 12

Question for discussion:

I have shown you how to construct two-mode paper-author networks and project them to one-mode co-authorship network. What are the problems with this projection?

Discussion and Hands-on lab session

Question for discussion:

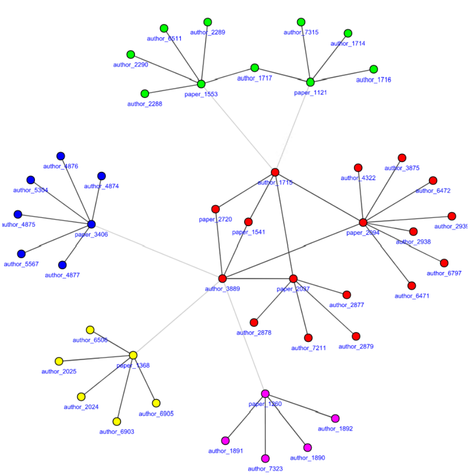
I have shown you how to construct two-mode paper-author networks and project them to one-mode co-authorship network. What are the problems with this projection?

Example problem:

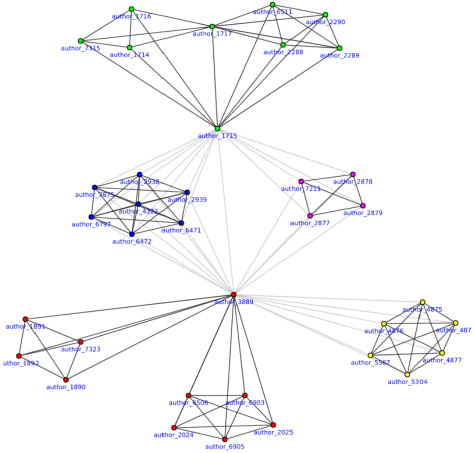
1) Different structures in two-mode networks are projected to the same one-mode structure causing an information loss about the underlying structure.
2) The one-mode projection can present an artificially higher density and connectivity due to publications with a high number of authors which project to maximally connected cliques.

Discussion and Hands-on lab session

[26]:



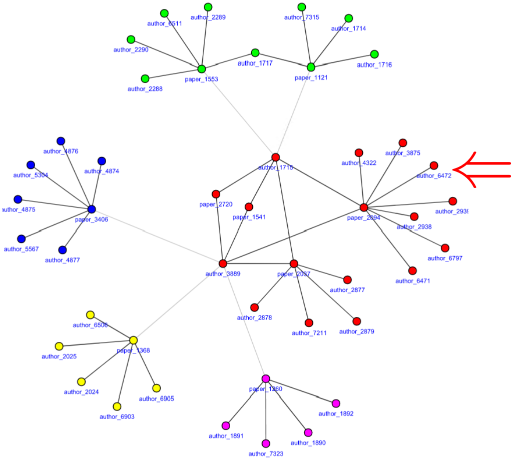
[123]:



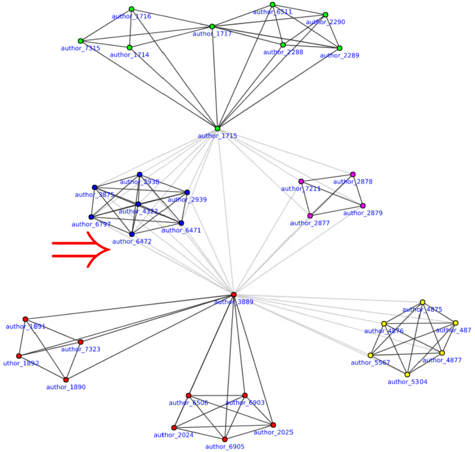

```
[27]:  ig.plot(g1)
```

Discussion and Hands-on lab session

[26]:



[123]:



[27]: `ig.plot(g1)`

Discussion and Hands-on lab session

Question for discussion:

I have shown you how to construct two-mode paper-author networks and project them to one-mode co-authorship network. What are the problems with this projection?

Example problem:

1) Different structures in two-mode networks are projected to the same one-mode structure causing an information loss about the underlying structure.
2) The one-mode projection can present an artificially higher density and connectivity due to publications with a high number of authors which project to maximally connected cliques.

In the hands-on lab session:

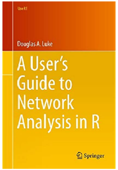
We will learn how to use R and Python scripts to retrieve data from Scopus API (as one example) and construct networks. In addition, we will discuss the limitations and pitfalls of re-purposing digital trace data for network modelling e.g., author name disambiguation.

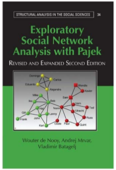
Outline (refresher) + recommended readings/conferences

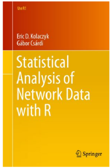
1 Introduction: What is "Tie formation mechanisms in social networks"? Observations by variables vs. relational; Violation of independence
2 Network theory and tie formation mechanisms: Foci of activity; Homophily
3 Network theory and tie formation mechanisms: Reciprocity; Activity; Transitivity
4 Network theory and tie formation mechanisms: Small-World; Popularity and Preferential attachment
5 Empirical analysis and methods: Conditional Uniform Graphs (CUGs); Quadratic Assignment Procedure (QAP) Regression
6 Empirical analysis and methods: Exponential Random Graph Models (ERGMs)
7 Empirical analysis and methods: Stochastic Actor-Oriented Models (SAOMs)
8 Empirical analysis and methods: Stochastic Block-Models (SBMs); Network Simulation with Agent-Based Models (ABMs)
9 Ethics: Ethical considerations in network analysis and modelling tie formation
10 Concluding discussions: Limitations and pitfalls of network data; Future directions

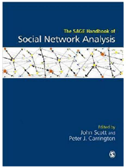
Check these books, conferences, and videos:

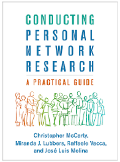
- INSNA-Sunbelt: https://www.insna.org
- EUSN: https://www.insna.org/european-conference-of-social-networks-eusn
- NetSci: https://netscisociety.net
- Videos: https://www.youtube.com/@Akbaritabar/playlists

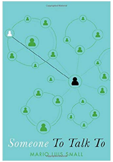


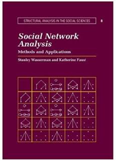


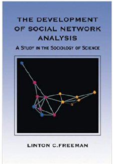












Prepared Bibliometric data for incubator project (1/3)

- One table structure of the edges data that includes edges, authors', organizations', and papers' attributes
- Using "organization" identification number and "author" identification number will allow constructing "co-affiliation" networks per year.
- Using "publication" identification number and "author" identification number will allow constructing "co-authorship" networks per year.
- Please check the reading materials and data codebook for examples and ask your questions.
    1. 10k sample: covers a large-enough sample that allows experimenting with the data (for the first team presentation as proof of concept).
    2. All Scopus authors: covers 8+ million authors, 28+ million article/review publications (to scale up the proof of concept to whole Scopus)
    3. Author/paper level calculated attributes (e.g. gender, academic age, and elaborated list on next slide)

Prepared Bibliometric data for incubator project (2/3)

- Example of author/paper level attributes include gender, academic age
- Other attributes regarding productivity, collaboration/internationalization, mobility, impact/visibility are listed below, more info: `https://dx.doi.org/10.4054/MPIDR-WP-2023-029`
  - The average number of coauthors per paper (as a measure for collaboration/internationalization)
  - The number of internationally coauthored publications (collaboration/internationalization)
  - The number of nationally coauthored publications (collaboration/internationalization)
  - The number of coauthored papers (collaboration/internationalization)
  - The number of international changes in academic affiliation (mobility)
  - The number of national changes in academic affiliation (mobility)
  - The number of affiliated organizations (mobility)
  - The average number of citations per paper (impact/visibility)
  - The total number of citations (impact/visibility)
  - The fractional count of publications (productivity)
  - The number of publications (productivity)
  - The number of first-author publications (productivity)

Prepared Bibliometric data for incubator project (3/3)

- This data enables two ways to construct networks and look at gender homophily (using authors' gender) over time
- Two approaches to "co-presence" network
  - Affiliation to the same institute (can be divided over past/current organisation versus future organizations)
  - Collaboration with others (can be divided over past/current collaborators versus future collaborators)

Thank you for your attention!

Questions and comments are welcome!

Aliakbar Akbaritabar

Research scientist

Deputy Head for Training

Max Planck Institute for Demographic Research (MPIDR), Rostock, Germany.

Email: akbaritabar@demogr.mpg.de

Twitter/BlueSky/Mastodon: @akbaritabar