ADVANCES IN NETWORK CLUSTERING AND BLOCKMODELING

Patrick Doreian
University of Pittsburgh

Vladimir Batagelj
University of Ljubljana

Anuška Ferligoj
University of Ljubljana

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Title, etc

To be set by Wiley
Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

CONTRIBUTORS

VLADIMIR BATAGELJ, Professor Emeritus, Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia

NEJC BODLAJ, Ljubljana, Slovenia

STEVE BORGATTI, LINKS Center for Social Network Analysis, Gatton College of Business and Economics, University of Kentucky, Lexington, KY, USA

MARJAN CUGMAS, Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia

ANUŠKA FERLIGOJ, Faculty of Social Sciences , University of Ljubljana, Ljubljana, Slovenia

PATRICK DOREIAN, Professor Emeritus, Department of Sociology, University of Pittsburgh, Pittsburgh, USA

MARTIN EVERETT, School of Social Sciences, Institution University of Manchester, Manchester, United Kingdom

LUKA KRONEGGER, Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia

RENAUD LAMBIOTTE, Department of Mathematics, University of Namur, Namur, Belgium

ANDREJ MRVAR, Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia

CARL NORDLUND, Department of Political Science, Center for Network Science (CNS), Budapest, Hungary

TIAGO PIEXOTO, Department of Mathematical Sciences, University of Bath, Bath, United Kingdom

MARTIN ROSVAL, Department of Physics, Umeå University, Umeå, Sweden

MICHAEL SCHAUB, ICTEAM, Louvain-la-Neuve, Belgium

LOVRO ŠUBELJ, Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia

VINCENT TRAAG, Centre for Science and Technology Studies,Leiden University, Leiden, The Netherlands

ALEŠ ŽIBERNA, Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia

ANJA ŽNIDARŠIČ, Faculty of Organizational Sciences, University of Maribor, Maribor, Slovenia

CONTENTS IN BRIEF

CONTENTS

ACRONYMS

SNA	Social Network Analysis
WoS	Web of Science

SYMBOLS

$\mathbb{N}$ set of natural numbers; $0 \in \mathbb{N}$; `\NN`

$\mathbb{Z}$ set of integers; `\ZZ`

$\mathbb{R}$ set of reals; `\RR`

$\mathcal{N}$ network; $\mathcal{N} = (\mathcal{V}, \mathcal{L}, w)$; `\Network`

$\mathcal{V}$ set of nodes (vertices); `\Nodes`

$\mathcal{L}$ set of links; $\mathcal{L} = \mathcal{E} \cup \mathcal{A}$; `\Links`

$\mathcal{E}$ set of edges (undirected links); `\Edges`

$\mathcal{A}$ set of arcs (directed links); `\Arcs`

n number of nodes; $n = |\mathcal{V}|$

m number of links; $m = |\mathcal{L}|$

$(u : v)$ edge with end-nodes u and v; $(u : v) = (v : u)$; `\edge{u}{v}`

(u, v) arc leding from node u to node v; `\arc{u}{v}`

$\mathbf{C}$ partition, clustering; `\cling`

indeg input degree; `\indeg`

outdeg output degree; `\outdeg`

argmin argument at which minimum is attained; `\argmin`

CHAPTER 1

INTRODUCTION

PATRICK DOREIAN[1], VLADIMIR BATAGELJ[2], AND ANUŠKA FERLIGOJ[1]

[1]Pittsburgh
[2]FMF, University of Ljubljana
[3]FDV, University of Ljubljana

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,* **ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 2

BIBLIOMETRIC ANALYSES OF THE NETWORK PARTITIONING LITERATURE

Vladimir Batagelj[1], Anuška Ferligoj[2], and Patrick Doreian[3]

[1]FMF, University of Ljubljana
[2]FDV, University of Ljubljana
[3]Pittsburgh

2.1 Introduction

REFERENCES

1. Batagelj, V.: WoS2Pajek. 2007. Manual for version 1.4, July 2016. `http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek:wos2pajek`
2. Batagelj, V.: Efficient Algorithms for Citation Network Analysis, 2003. `http://arxiv.org/abs/cs/0309023`
3. Batagelj, V., Doreian, P., Ferligoj, A. and Kejžar, N.: Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution. Wiley Series in Computational and Quantitative Social Science. Wiley, October 2014.
4. Batagelj, V., Cerinšek, M. (2013). On bibliographic networks. Scientometrics, 96(3), 845–864. doi: 10.1007/s11192-012-0940-1
5. Batagelj, V., Zaveršnik, M.: *Fast algorithms for determining (generalized) core groups in social networks.* Advances in Data Analysis and Classification 5 (2) (2011), 129-145.
6. Cerinšek, M., Batagelj, V. (2015). Network analysis of Zentralblatt MATH data. Scientometrics, 102(2015)1, 977-1001.

7. Cole, S., and Cole, J.R.: Visibility and the structural bases of awareness of scientific research. American Sociological Review 33(1973): 397-413.
8. De Nooy, W., Mrvar, A., and Batagelj, V.: Exploratory Social Network Analysis with Pajek; Revised and Expanded Second Edition. Structural Analysis in the Social Sciences, Cambridge University Press, September 2011.
9. Gauffriau, M., Larsen, P.O., Maye, I., Roulin-Perriard, A., von Ins, M.: Publication, cooperation and productivity measures in scientific research, Scientometrics, 73(2007)2, 175-214.
10. Harary, F., Norman, R.Z., and Cartwright, D.: *Structural models - an introduction to the theory of directed graphs*. Wiley, New York, 1965.
11. Liu, J.S. and Lu, L.Y.Y.: An integrated approach for main path analysis: Development of the Hirsch index as an example. Journal of the American Society for Information Science and Technology, 63 (2012),528-542
12. Garfield, E., Pudovkin, A.I., Istomin, V.S.: Why do we need algorithmic historiography? Journal of the American Society for Information Science and Technology, 54 (2003) 5, 400–412.
13. Garfield E., Pudovkin A.I., Istomin, V.S.: ***Histcomp*** – (*comp*iled *Hist*oriography program) `http://garfield.library.upenn.edu/histcomp/guide.html` `http://www.garfield.library.upenn.edu/histcomp/index.html`
14. Wikipedia: Peer review. `http://en.wikipedia.org/wiki/Peer_review`
15. Wikipedia: Histcite. `https://en.wikipedia.org/wiki/Histcite`
16. Wikipedia: Chinese surname. `https://en.wikipedia.org/wiki/Chinese_surname`
17. Zaveršnik, M., Batagelj, V.: Islands. XXIV. International Sunbelt Social Network Conference, Portorož, May 12–16, 2004. `http://vlado.fmf.uni-lj.si/pub/networks/doc/sunbelt/islands.pdf`

CHAPTER 3

CLUSTERING APPROACHES TO NETWORKS

Vladimir Batagelj[1]

[1]FMF, University of Ljubljana

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,* **ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 4

COMMUNITY DETECTION

Martin Rosvall[1] and Renaud Lambiotte[2]

[1] Umeå University
[2] University of Namur

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,***ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 5

LABEL PROPAGATION FOR PARTITIONING

Lovro Šubelj

FRI, University of Ljubljana

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,* **ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 6

PARTITIONING VALUED NETWORK DATA

Aleš Žiberna[1] and Carl Nordlund[2]

[1] FDV, University of Ljubljana
[2] Center for Network Science

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,* **ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 7

TREATING MISSING NETWORK DATA

Anja Žnidaršič[1], and Anuška Ferligoj[2]

[1]University of Maribor
[2]University of Ljubljana

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,***ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 8

PARTITIONING SIGNED NETWORKS

VINCENT TRAAG[1], PATRICK DOREIAN[2], AND ANDREJ MRVAR[3]

[1] University of XYZ
[2] University of Pittsburgh
[3] University of Ljubljana

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,***ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 9

PARTITIONING MULTI-WAY NETWORKS

MARTIN EVERETT[1] AND STEVE BORGATTI[2]

[1] University of Manchester
[2] Gatton College of Business and Economics

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,* **ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 10

PARTITIONING MULTI-LEVEL NETWORKS

Aleš Žiberna

FDV, University of Ljubljana

REFERENCES

1. J. S. Kilby, “Invention of the Integrated Circuit,” *IEEE Trans. Electron Devices,***ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, “A Theoretical Study of Gate/Drain Offset in LDD MOSFETs” *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O’Connor, “A Pipelined 32b Microprocessor with 13 kb of Cache Memory,” it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 11

STOCHASTIC BLOCKMODELING

Tiago Piexoto

University of Bath

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,* **ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 12

ESTABLISHING OVERLAPPING PARTITIONS

VLADIMIR BATAGELJ[1] AND NEJC BODLAJ[2]

[1] University of Ljubljanar
[2] XYZ Ljubljana

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,* **ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 13

DYNAMICS OF PARTITIONED NETWORK STRUCTURES

Michael Schaub

Université catholique de Louvain

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,* **ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 14

SCIENTIFIC CO-AUTHORSHIP NETWORKS

MARJAN CUGMAS, ANUŠKA FERLIGOJ, AND LUKA KRONEGGER

FDV, University of Ljubljana

14.1 Introduction

Network studies of science offer researchers a great insight into the dynamics of knowledge creation and the social structure of scientific society. The flow of ideas and overall cognitive structure of the scientific community is observed through citations between scientific contributions, usually manifested as patents or papers published in scientific journals. The social structure of this society consists of relationships among scientists. [De Haan, 1997] suggests six operationalized indicators of collaborative relations between scientists: co-authorship; shared editorship of publications; shared supervision of PhD projects; writing a research proposal together; participation in formal research programs; and shared organization of scientific conferences.

Due to accessibility and the ease of acquiring data through bibliographic databases, most scientific collaboration analyses are performed on co-authorship data, which play a particularly important role in research into the collaborative social structure of science. Co-authorship networks are personal networks in which the vertices represent authors, and two authors are connected by a tie if they co-authored one or more publications. These ties are necessarily symmetric. The study of community structures through scientific co-authorship is especially important because scientific (sub)disciplines can often display local properties that differ greatly from the properties of the scientific network as a whole. Co-authorship

Advances in Network Clustering and Blockmodeling,
February 5, 2017 17 : 15.

data have some flaws. The wide pallet of relationships among scientists do not result in common publications [Katz and Martin, 1997a, Melin and Persson, 1996, Laudel, 2002]. [Laudel, 2002] reports that about half of scientific collaborations are invisible in formal communication channels because they do not lead to either co-authored publications or formal acknowledgments in scientific texts. On the other hand, we also know that co-authorship sometimes represents false positive relations arising from resource-related issues [Ponomariov and Boardman, 2016]. Despite this, co-authorship data exist and at the moment provide the best available proxy for scientific research interaction.

The study of co-authorship networks has been influenced by the development of quantitative methodological approaches [Mali et al., 2010]. The choice of relatively simple descriptive statistics, deterministic modeling, stochastic agent-based modeling of network dynamics, or any other method is based on a particular study's objective. In the current article, we focus on blockmodeling co-authorship networks as a deterministic approach to network analysis.

There are relatively few applications of blockmodeling to co-authorship networks. This may be due to the method's limitations regarding the size of analyzed networks. One of the earliest applications can be found in [Ferligoj and Kronegger, 2009], who compared the results of blockmodeling (clustering of relational data) of a co-authorship network of Slovenian sociologists and the results of clustering with a relational constraint (clustering of attribute and relational data) on the same network according to researchers' publication performance. As expected, the methods produced different results, indicating their use should depend on the research problem under study. The unexpected result of their presented analysis was a core–periphery structure, with seven cores and a periphery, obtained when blockmodeling the co-authorship network.

Further investigation of the multicore–periphery structure was presented in [Kronegger et al., 2011] where the authors analyzed the development of a network structure over time. In their analysis of the co-authorship networks of four scientific disciplines (Physics, Mathematics, Biotechnology and Sociology) measured in four consecutive 5-year time spans, they observed a multicore–periphery structure was present from early on in the development of each scientific discipline. They also found that, although the number of cores increases with the growth of a discipline, the cores' sizes did not change. The structure's description as constituting multiple cores and a periphery was extended with two elements: a weakly connected semi-periphery that complements a completely empty periphery; and bridging cores, describing clusters of authors connecting two or more cores from the central part of the network. The authors described four levels of network complexity in the network structure's evolution through time:

1. Simple core–periphery form: Simple cores, semi-periphery, periphery
2. Weakly consolidated core–periphery form: Simple cores, bridging individuals, semi-periphery, periphery
3. Consolidated core–periphery form: Simple cores, bridging cores, semi-periphery, periphery
4. Strongly consolidated core–periphery form: Simple cores, bridging cores, bridging individuals, semi-periphery, periphery

Besides describing the overall structure, [Kronegger et al., 2011] attempted the first (visual) attempts to follow individual units in blockmodels' transition between timespans in order to pinpoint differences in the network dynamics between analyzed disciplines.

Which of the many approaches to studying co-authorship networks is chosen depends on the objective of the study under consideration. The most fundamental approaches to studying co-authorship networks relate to co-authorship networks' basic descriptive statistics, such as measuring the number and size of components in the network along with the degree and different measures of closeness and centrality (Liv et al. 2005). Some other authors proposed studying transformed co-authorship networks where the nodes are articles (instead of researchers) and links between two articles exist if they have one or several of the same authors (Gasko et al. 2006). An Exponential Random Graph Modeling was also applied to co-authorship networks in order to test a small world structure (Kronegger et al. 2012).

The multi-core–semi-periphery–periphery structure was also confirmed in a relatively small co-authorship network constructed from the curricula vitae (CVs) and bibliographies of teaching staff at the Faculty of Humanities and Education Science's Department of Library Science (DHUBI) at the National University of La Plata, Argentina [Chinchilla-Rodríguez et al., 2012] and might be present in the co-authorship networks of researchers from the Biomedical Research Networking Centers (CIBER) studied by Amat and Perruchas (2016).

14.2 Methods

A lot of attention has been paid to studying the relationship between collaboration on one side and the quality of research and speed of diffusion of scientific knowledge on the other [Hollis, 2001, Frenken et al., 2005, Abbasi et al., 2011, Lee and Bozeman, 2005]. While much research has considered the structure of co-authorship blockmodels [Ferligoj et al., 2015, Moody, 2004, Abbasi et al., 2012], not so much has examined the stability of long-term collaborations.

Here, it will be illustrated how blockmodeling can be used to reveal the global structure of co-authorship networks and how the stability of the blockmodels obtained can be operationalized and measured. This is especially important when seeking to explain the stability of research teams using common statistical methods such as linear regression.

14.2.1 Blockmodeling

The goal of blockmodeling is to reduce a large, potentially incoherent network to a smaller, comprehensible, and interpretable structure (Doreian et al. 2005). Compared to community detection methods (see [Lancichinetti and Fortunato, 2009] for some examples), blockmodeling can not only be used to find groups of highly linked units in a network, but also the relationships between the groups (deNoy et al. 2011). From this perspective, blockmodeling can reveal much more information about the global co-authorship structure than the community detection methods often used in bibliometric. A disadvantage of blockmodeling compared to community detection methods is that obtaining the solution (especially in the case of direct blockmodeling) can be very computationally expensive where networks with a higher number of units are involved.

The blockmodeling can be either direct or indirect. Indirect blockmodeling is based on a dissimilarity matrix among units. The calculated dissimilarity measure has to be consistent with a chosen equivalence between units. In the studies by [Kronegger et al., 2011] and [Cugmas et al., 2016], the corrected Euclidean distance, which is consistent with structural equivalence [Batagelj et al., 1992], was used. The process of hierarchical clustering of units can be visualized in a dendrogram in which the units (or groups) and the dissimilarity between the units (or groups) are represented. [Kronegger et al., 2011] and [Cugmas et al., 2016] defined the number of positions based on such visualization, although they reported that a slightly higher number of positions was chosen than would occur in a classical clustering procedure. This method for detecting the optimal number of positions by visual inspection is to some extent subjective.

On the other hand, unlike indirect blockmodeling direct blockmodeling can be achieved through a local optimization procedure (Batagelj et al. 1992), e.g. using an iterative method where for each displacement of a unit from one group into another, the value of the criterion function is calculated, defined as the difference between the ideal and empirical clustering where the ideal clustering has to express a blockmodel's assumed structure. It turns out that this procedure can be very time-consuming if a higher number of units in the network is analyzed. [Cugmas et al., 2016] also report that the algorithm implemented in Pajek has some difficulties detecting very small, structurally equivalent cores, particularly in the case of scientific disciplines with a very large number of researchers. To mitigate these characteristics, they removed the periphery and the structurally equivalent cliques from the network before applying the procedure. They later merged them to obtain the final result.

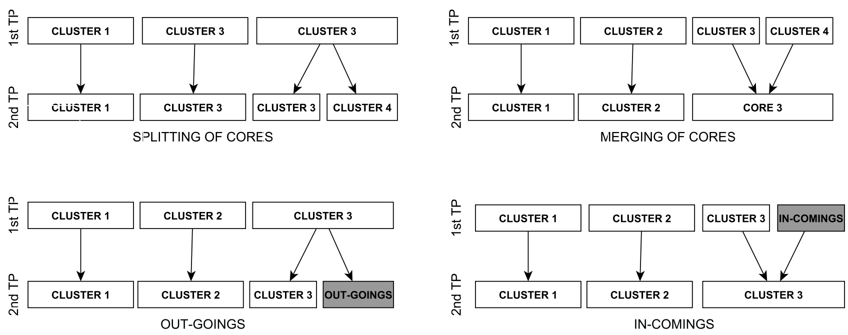
14.2.2 Measuring the stability of the obtained blockmodels

The main result of blockmodeling is a partition which assigns a researcher to a certain core, semi-periphery, or periphery. In the case of temporal co-authorship networks (where time is seen as a discrete variable), blockmodeling can be applied for each time period separately such that one partition for each time period is obtained[1]. A very important characteristic of temporal co-authorship networks is that some researchers (called in-comers) join the network at a later time period and others (called out-goings) leave the network at the later time period. Besides the presence of in-comers and out-goings, also the splitting of cores and merging of cores can be seen as separate factors that indicate the lower stability of the obtained blockmodels or cores.

Nevertheless, a combination of different factors usually appears simultaneously; a visualization of each factor is presented in Figure 14.1. Each visualization is divided into two parts: the white rectangles at the top visualize the clusters (which are cores obtained by blockmodeling in the case of co-authorship blockmodels) from the partition $U = \{u_1, ..., u_r\}$ obtained on the set of units from the first time period while the black rectangles on the bottom visualize the clusters from the partition $V = \{v_1, ..., v_c\}$ obtained on the set of units from the second time period. Gray rectangles are added to the clusters and visualize the out-comers and in-comers. The links between the rectangles visualize the clusters' stability.

[1]Along with the methods for generalised blockmodeling of multilevel networks [Žiberna, 2014], which can also be used for blockmodeling of temporal networks , different versions of stochastic blockmodeling exist for temporal networks [Matias and Miele, 2015, Xu and Hero III, 2013, Xing et al., 2010, Airoldi et al., 2007].

Figure 14.1: The factors that can be used as indicators of less stable clusterings



Adjusted Rand Index On the two assumptions that the merging and splitting of clusters are indicators of a lower level of cluster stability in time and that there are no in-comers or no out-goings present (or, at least, that they are neglected), one of the most widely and popular indexes for comparing partitions, the Adjusted Rand Index ([Hubert and Arabie, 1985, Saporta and Youness, 2002]), can be used. Here, the adjective "adjusted" refers to the necessary correction for chance since the expected value is usually not 0 in the case of two random and independent partitions. This correction allows the values of the index obtained from different partitions to be compared. Let us focus on the Rand Index [Rand, 1971], which is defined as

$$RI = \frac{a+d}{a+b+c+d}$$

where a stands for the number of pairs of researchers classified in the same cluster in both time periods, b stands for the number of pairs of researchers classified in the same clusters in the first period but in different clusters in the second period, c stands for the number of researchers classified in different clusters in the first, but in the same cluster in the second period and, finally, d stands for the number of pairs of researchers classified in different clusters in both the first and second time periods. Following this definition, the Rand Index can be interpreted in the co-authorship network context as the proportion of all possible pairs of researchers classified in the same or in different clusters in both time periods out of all possible pairs of researchers.

Wallace indices There are situations when the merging and splitting of clusters has to be considered differently. Therefore, one of two Wallace Indices can be used: in the case of the Wallace Index' (WI'), only the splitting of clusters is considered a factor indicating lower cluster stability while with the Wallace Index" (WI") only the merging of cores is considered a factor indicating the lower stability of clusters. Formally, WI' is defined as

$$WI' = \frac{a}{a+b}$$

where a and b are defined the same as in the case of RI. WI' can be interpreted as the proportion of all researcher pairs placed in the same core in the first period out of the number of all possible researcher pairs placed in the same core in both time periods. Similarly, WI" is defined as

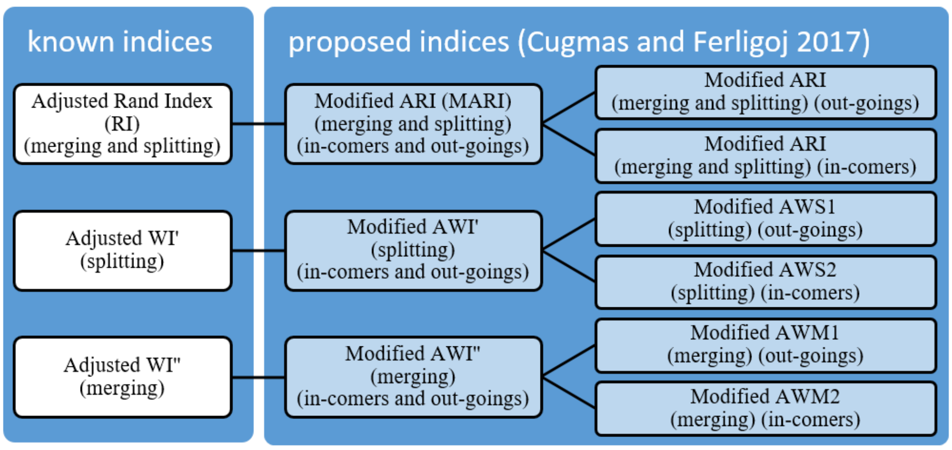
$$WI'' = \frac{a}{a+c}$$

and interpreted as the ratio between the number of all possible researcher pairs classified in the same cluster in both periods and the number of all possible researcher pairs classified in the same cluster in the second period (the probability that a pair of researchers will be placed in the same cluster in the second period if they were placed in the same cluster in the first period).

Modified Rand Index and Wallace indices As mentioned, it is common in temporal co-authorship networks that some researchers join the network and some leave the network in later time periods. When this happens, one can either simply ignore those researchers when calculating the Rand or Wallace indices, or treat the in-comers and out-goings as factors indicating a lower level of stability of the cores. When the latter is assumed, one has to form new partitions $U' = \{u_1, u_2, ..., u_{r+1}\}$ and $V' = \{v_1, v_2, ..., v_{c+1}\}$ with the new clusters of in-comers u_{r+1} and out-goings v_{c+1} added to the partitions U and V. Then, the Modified Adjusted Rand Index (MARI), the Modified WI', and the Modified WI" are calculated in the same way as RI, WI', and WI" where the values in the numerator consider the partitions U' and V'. The modified Rand Index and the modified Wallace indices can be further modified in such a way that only in-comers or only out-goings are considered as factors indicating lower core stability (for more details, see Cugmas and Ferligoj (2017)) (see Figure 14.2).

Along with the modified Rand Index and the modified Wallace indices, Cugmas and Ferligoj (2017) proposed a correction for chance (based on Monte Carlo simulations) that allows one to compare the values of indices obtained in different scientific disciplines. With non-adjusted indices, the number of clusters (cores, in-comers, and out-goings) and the number of researchers also influence the expected value of an index in the case of two random and independent partitions. The expected value of two random and independent partitions in the case of adjusted indices equals 0, and the maximum value of an index is 1. It should be highlighted that higher values of the presented indices indicate a higher level of cluster stability, while lower values indicate a lower level of stability. Negative values are also possible.

Figure 14.2: The indices for measuring the stability of cores in time (in brackets the factors that lower the stability are given)



14.3 The data

The data for this research were obtained from the Co-operative Online Bibliographic System and Services (COBISS) and the Slovenian Current Research Information System (SICRIS) maintained by the Institute of Information Science (IZUM) and the Slovenian Research Agency (SRA).

SICRIS provides data about all researchers which have an ID assigned by the SRA, including their educational background and field of research according to the SRA's classification scheme. There are 7[2] scientific fields and 72 scientific disciplines defined in this classification scheme. There are some differences in the SRA's classification scheme compared to other classification schemes, e.g., the Common European Research Classification Scheme.

The analyzed data are based on complete personal bibliographies of each researcher (constructed based on SICRIS and COBISS). The network boundaries are therefore defined only by those researchers registered as a researcher at the SRA. Among such researchers, those who published at least one scientific bibliographic unit between 1990 and 2010 are analyzed. The bibliographic units considered as a scientific publication by the SRA are listed in Table 14.1.

Table 14.1: The number of published scientific bibliographic units by type for two time periods

Type of scientific bibliographic unit	1991 - 2000	2001 - 2010
Original scientific article	26531	47905
Review article	4895	5738
Short scientific article	969	2530
Published scientific conference contribution (invited lecture)	3427	5279
Published scientific conference contribution	28670	41138
Independent scientific component part in monograph	6417	14759
Scientific monograph	1725	2912
Scientific or documentary films, sound or video recording	44	133
Complete scientific database or corpus	73	182
Patent	381	710
Total	73132	121286

There were 73,132 scientific bibliographic units published between 1991 and 2000 and 121,286 scientific units between 2001 and 2010. The most common are published scientific conference contributions and original scientific articles. Also very common are monographs or parts of monographs, and review articles. The distribution of different types of bibliographic units varies among scientific disciplines. For example, published scientific conference contributions are very common to scientific disciplines from the technical sciences while original scientific articles are frequent among scientific disciplines within the Social Sciences and Humanities. There are differences at the level of scientific disciplines according to the distribution of types of scientific bibliographic units which can be published by one or several researchers. [Kronegger et al., 2015] who studied the differences between scientific disciplines according to collaboration patterns in time confirmed

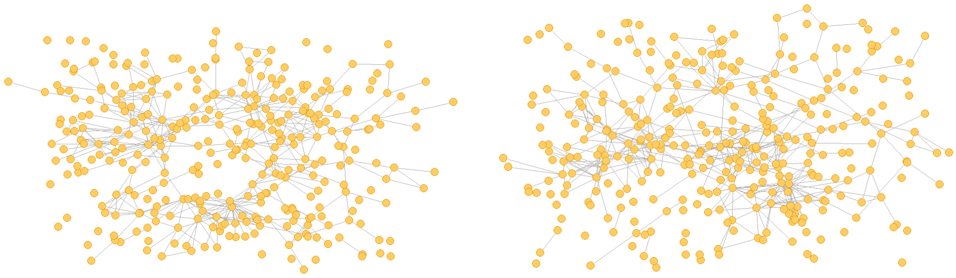
[2] The 7th scientific discipline is Interdisciplinary Studies and is not included in the analysis since it has never gained full recognition as a separate research field in Slovenia (Ferligoj et al. 2015).

the scientific discipline Geography is more similar to scientific disciplines in the scientific fields Natural Sciences and Mathematics than the scientific field of the Humanities where it belongs according to the SRA's classification scheme. Even within a number of scientific disciplines one can expect some differences in types of co-authorships. In the case of Sociology, [Moody, 2004] concluded that quantitative work is more likely to be co-authored than non-quantitative work.

Compared to the analysis conducted by [Kronegger et al., 2011] who studied four selected scientific disciplines in four time periods, the current analysis is performed on data for two consecutive 10-year periods between 1991 and 2010. The difference in the length of the periods mainly affects the size and density of the generated co-authorship networks and, in terms of the stability of research teams, result in a lower level of stability. However, the two periods selected reflect a time of major changes to scientific research and development policies in Slovenia. The first period (1990–2000) is marked by the independence of Slovenia, meaning that Slovenia had started adopting and implementing its own science policies, while the second period (2001–2010) is marked by the country joining the European Union and adopting European Union standards. By the end of this period, Slovenia had already partly integrated its national science system into the European one.

Although many co-authorship networks are analyzed in this study, we present Sociology co-authorship networks in Figure 14.3 as an example. The units represent researchers and a link between two researchers exists if they published at least one scientific bibliographic unit in co-authorship. Therefore, only symmetric links are possible in the case of co-authorship networks. Looking at the Sociology co-authorship networks reveals there are also some researchers without any link. These researchers are later classified in the so-called periphery, explained in greater detail in the next section. However, it should be pointed out that the absence of links is not necessarily the consequence of only single-authored scientific bibliographic units by a certain researcher, but can also be the outcome of co-authoring only with researchers who do not have a researcher ID, for example with researchers from abroad. Researchers who are without any link are present in both time periods. The next important network characteristic seen in this figure and common to almost all scientific disciplines is that the co-authorship networks grow in time. There are more researchers in Figure 14.3a compared to Figure 14.3b, although the number of researchers is relatively small in both periods. Without using any specific clustering method, several internally highly connected groups of researchers can be observed. To study this, the matrix visualization is often seen as more appropriate, as demonstrated in the next section, along with a detailed description of the disciplines' characteristics.

Figure 14.3: Visualization of the co-authorship network for Sociology for two periods



(a) 1991 to 2000

(b) 2001 to 2010

14.4 The structure of obtained blockmodels

Based on four scientific disciplines, [Kronegger et al., 2011] showed that the structure of co-authorship networks consists of the multi-core, semi-periphery, and periphery. To confirm that this structure is also present in other scientific disciplines, [Cugmas et al., 2016] used indirect blockmodeling to detect the approximate number of cores and direct blockmodeling to obtain the final solution as described in section 14.2.1. The assumed blockmodel structure was confirmed in all scientific disciplines included in the analysis. Most disciplines that were excluded (in Figure 14.4 indicated by asterisks) were removed due to a small number of researchers in the first or second time period or absence of co-authorship in the current period. One such discipline is Theology that did not have a single co-authored scientific bibliographic item published in the first period. It can also be seen in Figure 14.4 that the number of researchers who published at least one scientific bibliographic item is increasing over time in almost all scientific disciplines. The average growth in the number of researchers publishing at least one scientific bibliographic item in the second period is 34 %. Only in the disciplines Veterinary Medicine, Stomatology and Mining and Geotechnology is a decrease in the number of researchers from the first to the second period observed.

Figure 14.12 visualizes two empirical blockmodels of the scientific discipline Sociology (the corresponding co-authorship networks are visualized in Figure 14.3). The first blockmodel corresponds to the first period (from 1991 to 2000) while the second blockmodel corresponds to the second period (from 2001 to 2010). The rows and columns of each blockmodel contain the IDs of the researchers assigned by the Slovenian Research Agency, where the black dots in the cells denote co-authorships between two given researchers. A clear multi-core–semi-periphery–periphery structure can be seen in the case of Sociology. Along with the already described multi-core, semi-periphery, and periphery, in the blockmodel in the first period a so-called bridging core is seen (as a full off-diagonal block) (Figure 14.5a). The bridging core is a group of researchers who collaborate between each other very systematically and also with researchers from at least two other cores. They are called "bridging" since they connect two or more cores. They are relatively common in Slovenian scientific disciplines. There was a minimum of one bridging core in at least one time period in 20 of all analyzed scientific disciplines. However, one can notice the structures obtained for the first and second periods are similar. It can be observed that the periphery has decreased in time for the case of Sociology.

To gauge the application of blockmodeling vs. community detection methods, the Louvain Method [Blondel et al., 2008] for identifying communities in large networks was also applied to the Sociology co-authorship network. The results are visualized in Figure 14.6. The method is able to detect the periphery and structurally equivalent clusters very well without any error. Let us remind that the blockmodeling procedure implemented in Pajek has some difficulties detecting these units if insufficient iterations are used. On the other hand, the Louvain Method is unable to identify the semi-periphery. Instead, each quasi core cluster obtained by the Louvain Method seems to have its own semi-periphery.

By neglecting the bridging cores, another blockmodel visualization for the two periods can be made. The visualization in Figure 14.7a emphasizes the transitions of researchers between the obtained cores (including the semi-periphery and periphery) for the two periods. The visualization has two parts: the upper part visualizes the classification of researchers for the first period and the bottom part visualizes the classification of researchers for the second period. Figure 14.12 shows that the share of researchers classified in the periphery is decreasing in Sociology, which cannot be seen in the visualization of researchers'

Figure 14.4: List of scientific disciplines with number of researchers in the first and second periods (an asterisk indicates scientific disciplines which were not considered in the analysis)

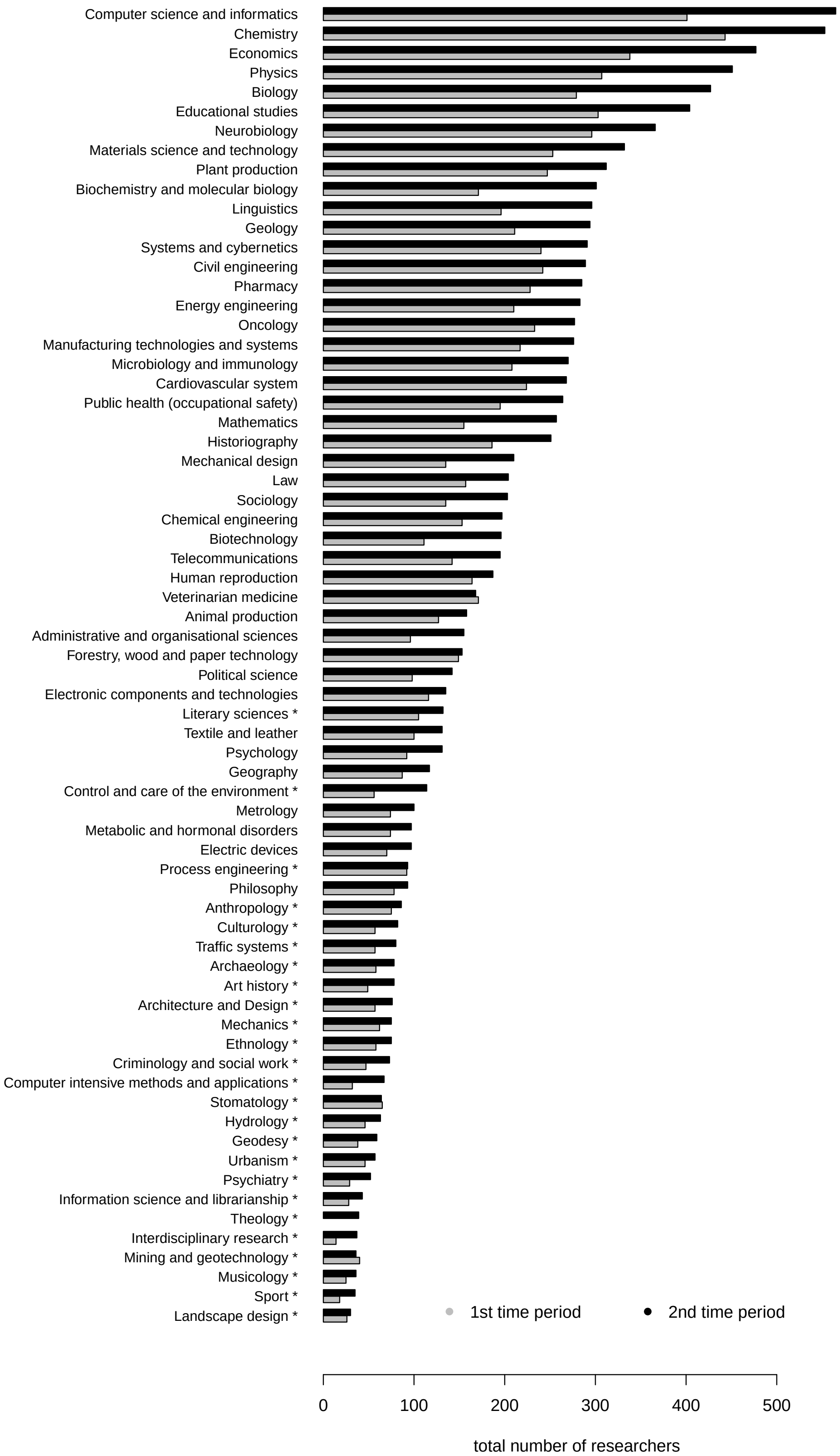
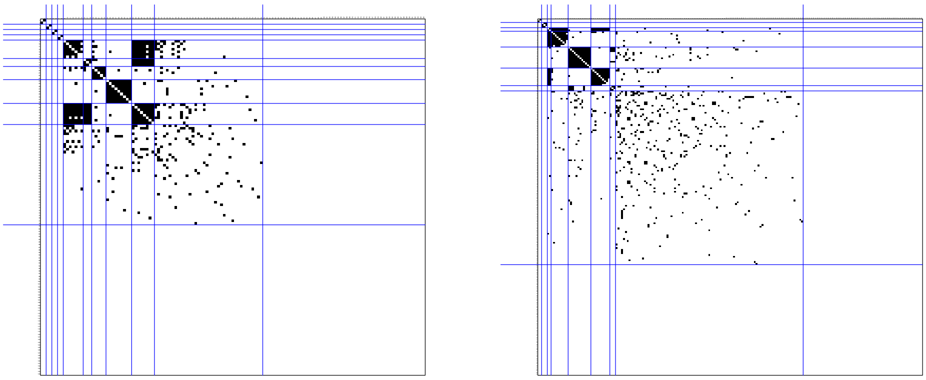


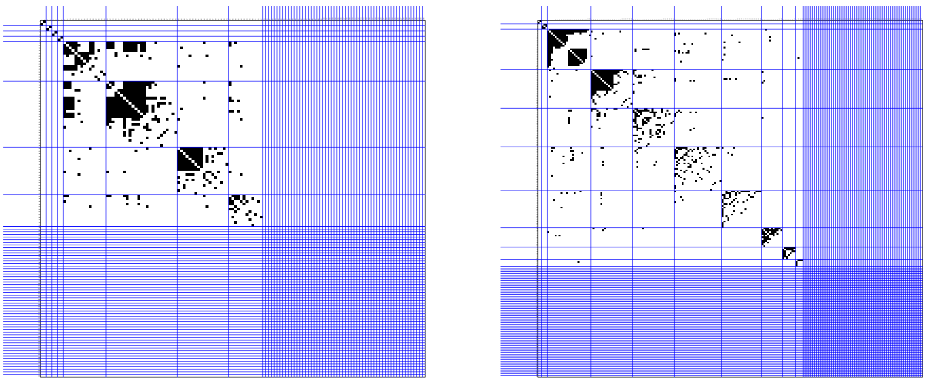
Figure 14.5: Structure of Sociology co-authorship blockmodel for the first and second time periods



(a) 1991 to 2000

(b) 2001 to 2010

Figure 14.6: Structure of Sociology co-authorship network where the order of units by the rows and the columns is obtained using the Louvain community detection method for the first and second periods



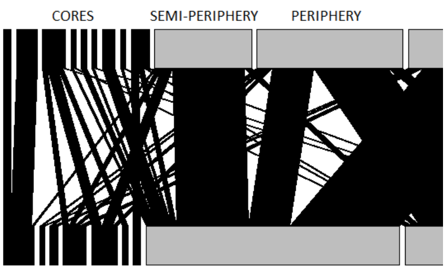
(a) 1991 to 2000

(b) 2001 to 2010

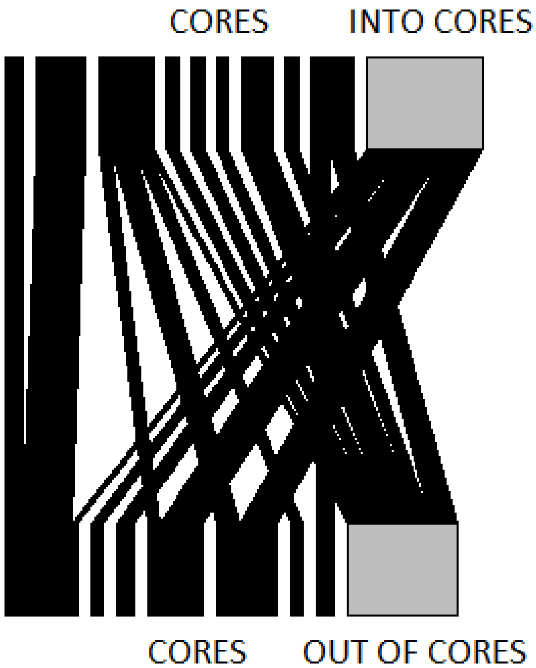
transitions in time in Figure 14.7b. This is caused by the in-comers and out-goings. Figure 14.7a reveals a high share of researchers who were not classified in the cores in both time periods (e.g. many researchers were classified in the periphery in the first and second periods). Further, many in-comers were classified in the semi-periphery or periphery in the second period. Some evidence that many new researchers were not connected to any previously existing authors in the field of steel structures is reported by [Abbasi et al., 2012].

Since the main interest of study is the stability of the cores of the obtained blockmodels, researchers not classified in the cores in at least one period can be removed from the visualization. Therefore, a new visualization can be presented in Figure 14.7b consisting of two parts (one for each period) without the semi-periphery, periphery, in-comers, and out-goings. Instead, researchers classified in the cores in the first but not in the second period are now called "out-of-cores" researchers and, similarly, researchers not classified in the cores in the first period but were classified in the core in the second period are now called "into-cores" researchers. Focusing on the core part of the Sociology example, it can be observed that cores 1 and 2 merged in the second period, while core 3 splits into three cores in the second period. There are also many cores which disappear in the second period (out-of-cores researchers) and a lot of researchers not classified in the cores in the first but are classified in the cores in the second period. These into cores researchers usually join already the existing cores in the second period. The dynamics of these transitions differ by disciplines: in some, researchers collaborate with the same colleagues in large research groups for a long time, while in others researchers work in small groups for a shorter period.

Figure 14.7: Visualization of researchers' transitions between the cores, semi-periphery and periphery (a) and visualization researchers' transitionss between the cores, into cores and out of cores (b) in two time periods for



(a)



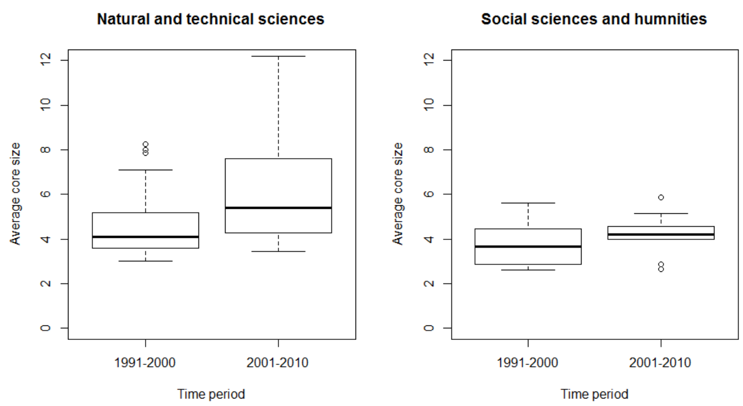
(b)

Visualizations of researchers' transitions between the core, into cores and out of cores in the two periods are made for all analyzed scientific disciplines (Figure 14.8). A relatively high share of into-cores and out-of-cores researchers in all analyzed scientific disciplines and some merging and splitting of cores in the core part of the visualized transitions can be seen. Here, the into-cores and out-of-cores researchers are seen as the primary source of instability of the core part of scientific disciplines. Although the share of into-cores researchers is higher than the share of out-of-cores researchers in almost all analyzed scientific disciplines, some scientific disciplines reveal the share of out-of-cores prevails over the share of into-cores researchers.

The number and the size of the cores, the size of the semi-periphery and the size of the periphery vary across scientific disciplines. For example, the discipline Administrative and Organizational Sciences consists of 6 cores in the first period and 16 cores in the second (see Figure 14.9 and Table 14.2). Here, the average core size is lower in the second period. On the other hand, there are 15 cores in the discipline Physics in both the first and second period, but the average core size is higher in the second period than in the first. Across the disciplines, the highest average core size in the first period is observed in Oncology (8.3 researchers) and Human Reproduction (8.0 researchers), while the lowest average core size in the first period is observed in Linguistics (2.6 researchers) and Psychology (2.9 researchers). In general, the overall average number of cores is similar in both periods (around 11 cores), while the overall average size of the cores is increasing in time (from an average core size of 4.4 to an average of 5.6 researchers), as confirmed by [Amat and Perruchas, 2015].

Following the distinction between the natural and technical sciences on one side, and the social sciences and humanities on the other, it can be concluded that the average size of the cores is increasing in both the natural and technical sciences and the social sciences and humanities, although the average core size is lower in the social sciences and humanities in both periods (Figure 14.9).

Figure 14.9: The average core size by field and time period



Solo authors or authors who published only in co-authorship with authors from outside the discipline are classified in the periphery. The share of these authors is decreasing in time (from a 39 % average share of the periphery in the first period to a 30 % average share of the periphery in the second period). The biggest reduction in the percentage of the periphery in the second period is observed in Criminology and Social Work (a 65 % decrease). In some scientific disciplines, the percentage of the periphery increased in the second period. These are mainly disciplines from the natural and technical fields. However, the size of the periphery is greater in fields of the social sciences and humanities than in scientific disciplines classified in the natural and technical sciences. In addition, the

Figure 14.8: Visualization of researchers' transitions between the cores in the two periods for all analyzed scientific disciplines

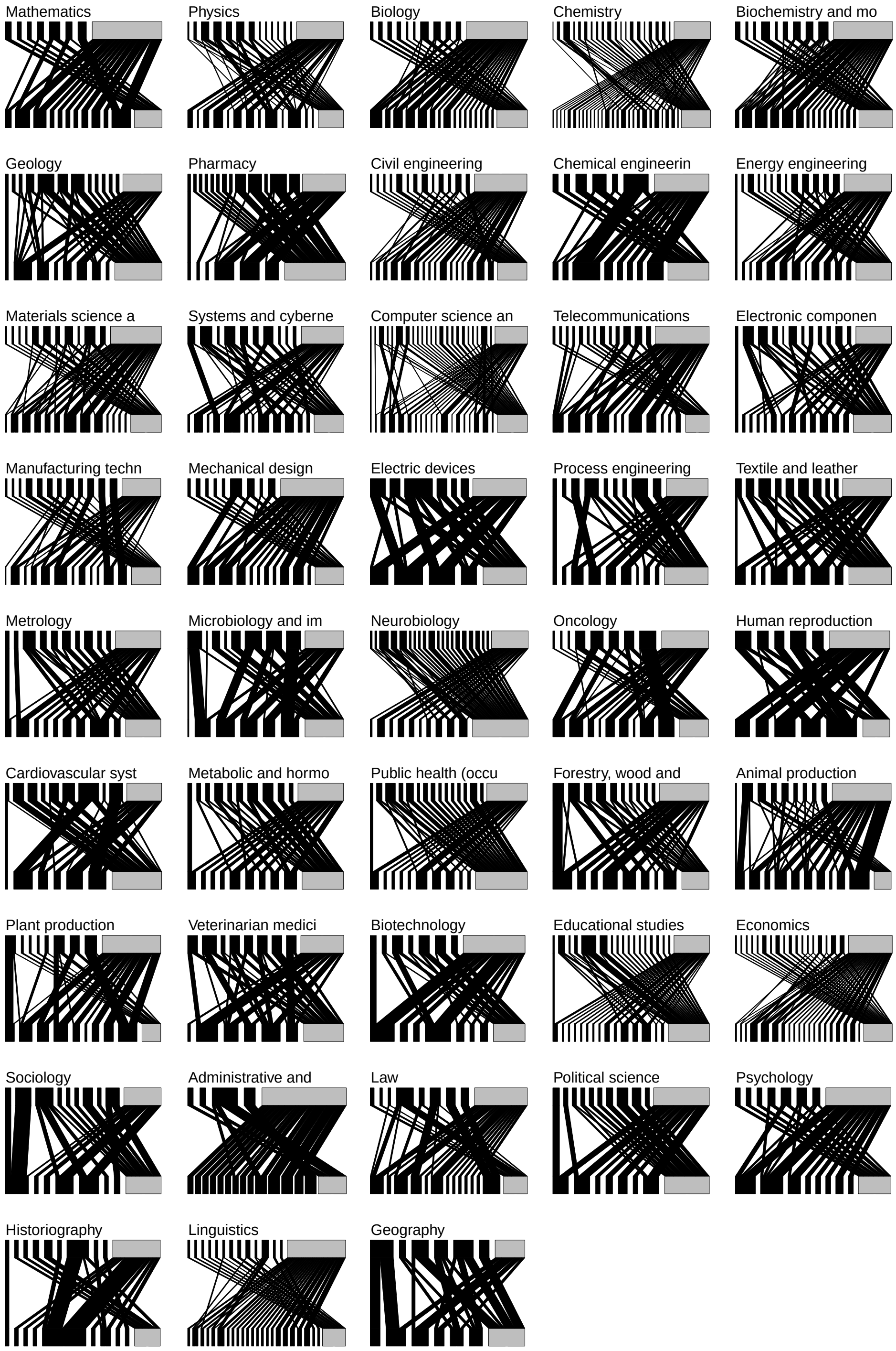
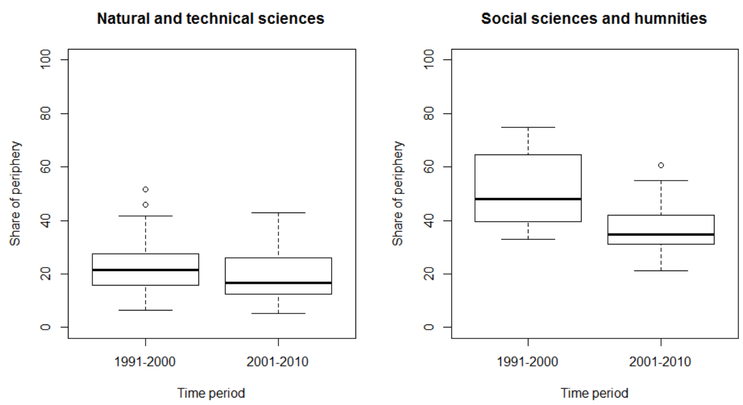


Table 14.2: The considered scientific disciplines and their characteristics in the two periods

	1991 - 2000					2001 - 2010					
Scientific discipline and scientific field	N	number of cores	semi-per. (%)	per. (%)	average core size	N	number of cores	semi-per. (%)	per. (%)	average core size	% of researchers that left the cores
1. Natural sciences and mathematics											
Biochemistry and molecular biology	171	11	45.61	30.41	4.56	301	17	45.85	33.55	4.13	68.29
Biology	279	12	46.59	37.63	4.40	427	18	55.04	27.40	4.69	70.45
Chemistry	443	22	64.79	12.87	4.95	553	29	67.09	11.93	4.30	68.69
Geology	211	14	44.55	34.12	3.75	294	10	43.54	42.86	5.00	64.44
Mathematics	155	9	34.19	51.61	3.14	257	13	51.36	32.30	3.82	59.09
Pharmacy	228	15	51.75	25.00	4.08	285	8	71.58	14.04	6.83	77.36
Physics	307	15	61.89	15.96	5.23	451	15	65.63	12.42	7.62	52.94
2. Engineering sciences and technologies											
Chemical engineering	153	8	60.13	24.18	4.00	197	10	66.50	18.78	3.62	62.50
Civil engineering	242	15	65.29	15.70	3.54	289	18	61.94	13.49	4.44	69.57
Computer science and informatics	401	25	51.87	27.43	3.61	565	21	63.36	21.24	4.58	62.65
Electric devices	70	8	31.43	25.71	5.00	97	6	38.14	26.80	8.50	56.67
Electronic components and technologies	116	14	36.21	23.28	3.92	135	15	40.74	25.93	3.46	61.70
Energy engineering	210	15	57.14	16.67	4.23	283	14	60.07	15.90	5.67	70.91
Manufacturing technologies and systems	217	14	45.62	28.57	4.67	276	14	52.90	23.55	5.42	50.00
Materials science and technology	253	13	60.08	17.39	5.18	332	15	62.05	13.86	6.15	61.40
Mechanical design	135	11	55.56	21.48	3.44	210	13	56.19	17.62	5.00	64.52
Metrology	74	12	39.19	20.27	3.00	100	11	35.00	30.00	3.89	56.67
Process engineering	92	11	52.17	11.96	3.67	93	11	48.39	17.20	3.56	66.67
Systems and cybernetics	240	12	54.17	21.67	5.80	291	14	58.42	17.53	5.83	48.28
Telecommunications	142	15	39.44	32.39	3.08	195	13	45.13	21.03	6.00	50.00
Textile and leather	100	11	55.00	12.00	3.67	131	9	61.07	11.45	5.14	63.64
3. Medical sciences											
Cardiovascular system	224	11	54.91	16.52	7.11	268	8	67.54	12.69	8.83	57.81
Human reproduction	164	7	53.66	21.95	8.00	187	7	55.61	11.76	12.20	42.50
Metabolic and hormonal disorders	74	12	13.51	45.95	3.00	97	11	38.14	28.87	3.56	66.67
Microbiology and immunology	208	10	54.81	14.90	7.88	270	8	67.78	8.89	10.50	47.62
Neurobiology	296	22	51.01	25.68	3.45	366	12	69.67	16.67	5.00	81.16
Oncology	233	10	54.08	17.60	8.25	277	11	56.32	13.36	9.33	45.45
Public health (occupational safety)	195	17	34.87	41.54	3.07	264	12	56.06	29.55	3.80	80.43
4. Biotechnical sciences											
Animal production	127	11	64.57	6.30	4.11	158	12	48.73	7.59	6.90	35.14
Biotechnology	111	9	49.55	18.92	5.00	196	9	58.16	13.78	7.86	60.00
Forestry, wood and paper technology	149	12	62.42	11.41	3.90	153	10	53.59	11.76	6.62	43.59
Plant production	247	10	68.83	12.96	5.62	312	11	66.35	8.33	8.78	35.56
Veterinarian medicine	171	10	60.23	8.19	6.75	168	8	61.31	5.36	9.33	51.85
5. Social sciences											
Administrative and organisational sciences	96	6	17.71	64.58	4.25	155	16	34.19	41.94	2.64	58.82
Economics	338	20	49.11	32.84	3.39	477	22	61.01	21.17	4.25	73.77
Educational studies	303	19	38.28	39.60	3.94	404	17	51.24	33.91	4.00	76.12

average share of the periphery decreases from the first to the second periods in both the natural and technical sciences and the social sciences and humanities (Figure 14.10).

Figure 14.10: The average size of the periphery by field and time period



14.5 Stability of the obtained blockmodel structures

In this section, the stability of cores is studied according to different operationalisations of core stability. Although the presented visualizations of researchers' transitions between two time periods (Figure 14.8) are a very efficient tool for studying the stability of the cores obtained but whose interpretation is complex, the values of the indices proposed in Chapter 14.2 are calculated. These indices are more objective operationalizations of core stability and allow us to compare the values calculated for different scientific disciplines. The scientific disciplines are then clustered according to the calculated indices. The groups of scientific disciplines thus obtained are further analyzed.

In the second part, the operationalization of the stability of cores is restricted to one of the described indices for measuring core stability, namely, as applied in [Cugmas et al., 2016] only the splitting of cores and the out-of-cores researchers are seen as factors indicating lower stability of the cores. The hypothesis about differences in the mean stability of cores among different scientific fields is studied using linear regression. Some further controlling explanatory variables are also included in the model.

First, the values of each presented index for each analyzed scientific discipline are shown in Table 14.3 and provide the basis for all further analyses. In this table, one sees that the values of the Adjusted Rand Index and the adjusted Wallace indices are relatively large, while the others are relatively small. This is due to the high share of into-cores and out-of-cores researchers which are not considered when calculating the values of the Adjusted Rand Index and the adjusted Wallace indices for each scientific discipline. The high values of the first three indices and the low values of the others confirm that the into-cores researchers and out-of-cores researchers are the biggest source of the obtained cores' instability.

Table 14.3: The values of different indices for measuring the stability of cores for all analyzed scientific disciplines

Discipline	ARI	AW'	AW"	MARI1	MAWIS1	MAWIM1	MARI2	MAWIS2	MAWIM2
Mathema	1.00	1.00	1.00	0.05	0.27	0.08	0.01	0.01	0.09
Civil e	0.86	0.76	1.00	0.01	0.14	0.02	0.01	0.01	0.06
Chemica	1.00	1.00	1.00	0.08	0.33	0.14	0.09	0.08	0.30
Energy	0.81	0.75	0.88	0.01	0.15	0.02	0.05	0.01	0.09
Materia	0.33	0.26	0.45	0.01	0.06	0.02	0.02	0.01	0.04
Systems	0.70	0.57	0.92	0.04	0.19	0.09	0.12	0.04	0.15
Compute	0.63	0.57	0.71	0.01	0.19	0.03	0.04	0.02	0.14
Telecom	0.69	0.89	0.56	0.04	0.35	0.08	0.06	0.02	0.09
Electro	0.62	0.47	0.91	0.01	0.13	0.03	0.11	0.03	0.19
Manufac	0.95	0.90	1.00	0.06	0.43	0.12	0.11	0.07	0.25
Physics	0.40	0.67	0.28	0.03	0.15	0.04	0.03	0.02	0.08
Mechani	1.00	1.00	1.00	0.04	0.23	0.06	0.04	0.01	0.09
Electri	0.50	0.46	0.56	0.03	0.11	0.06	0.04	0.03	0.07
Process	0.84	0.72	1.00	0.02	0.14	0.04	0.1	0.04	0.17
Textile	0.80	0.80	0.80	0.03	0.18	0.04	0.00	0.03	0.10
Metrolo	0.42	0.44	0.40	0.01	0.16	0.04	0.00	0.02	0.09
Biology	0.38	0.48	0.31	0.01	0.09	0.02	-0.02	0.00	0.04
Microbi	0.88	0.86	0.91	0.09	0.32	0.16	0.25	0.16	0.26
Neurobi	0.43	0.68	0.32	0.00	0.08	0.01	0.03	0.01	0.09
Oncolog	0.89	0.85	0.93	0.11	0.42	0.23	0.12	0.11	0.35
Human r	0.40	0.93	0.26	0.08	0.28	0.14	0.12	0.06	0.14
Cardiov	1.00	1.00	1.00	0.06	0.30	0.10	0.32	0.18	0.30
Metabol	0.73	1.00	0.57	0.02	0.07	0.01	-0.02	0.01	0.06
Public	0.37	0.32	0.42	0.00	0.03	0.00	0.00	0.00	0.02
Chemist	0.60	0.46	0.89	0.01	0.17	0.02	0.04	0.01	0.17
Forestr	0.64	0.69	0.60	0.06	0.34	0.14	0.10	0.05	0.15
Animal	0.49	0.51	0.47	0.06	0.19	0.13	0.06	0.01	0.06
Plant p	0.90	0.84	0.97	0.15	0.45	0.34	0.11	0.05	0.19
Veterin	0.52	0.68	0.43	0.04	0.15	0.05	0.13	0.05	0.09
Biotech	0.73	1.00	0.57	0.04	0.14	0.05	-0.01	0.01	0.04
Educati	0.32	0.34	0.31	0.00	0.04	0.01	-0.02	0.01	0.07
Economi	0.71	0.64	0.80	0.01	0.14	0.01	0.01	0.01	0.07
Sociolo	0.52	0.55	0.50	0.06	0.36	0.16	0.25	0.14	0.23
Biochem	-0.16	-0.11	-0.27	-0.02	0.00	0.00	-0.03	0.00	0.00
Adminis	0.80	0.67	1.00	0.06	0.11	0.09	0.05	0.01	0.19
Law 5	0.58	0.80	0.45	0.09	0.29	0.17	-0.14	0.06	0.22
Politic	0.86	1.00	0.75	0.01	0.13	0.02	-0.03	0.01	0.05
Psychol	0.04	0.07	0.03	0.00	0.02	0.00	-0.09	0.00	0.01
Geology	0.09	0.11	0.07	0.00	0.02	0.00	0.00	0.01	0.03
Histori	0.57	0.68	0.49	0.08	0.39	0.20	0.05	0.07	0.09
Linguis	0.43	0.29	0.82	0.00	0.16	0.03	0.03	0.00	0.08
Geograp	0.36	0.29	0.49	0.03	0.15	0.12	0.22	0.12	0.21
Pharmac	0.50	0.69	0.39	0.01	0.03	0.01	0.09	0.01	0.03

14.5.1 Clustering of scientific disciplines according to different operationalisations of the stability of cores

Based on the standardized calculated indices presented in Table 14.3, the analyzed scientific disciplines are clustered using Ward's agglomerative clustering method and squared Euclidean distance. Using the GAP Statistics [Tibshirani et al., 2001] and the obtained dendrogram three clusters are chosen. By observing the means of the calculated standardized indices for each cluster (Table 14.4), the obtained clusters can be ordered from the least stable (Cluster 1) where all values are below zero to the most stable cluster (Cluster 3) where all values are above zero. Cluster 2 is seen as stable/unstable since the values of the standardized means are around zero.

Table 14.4: The standardised mean values of the calculated indices for each cluster

Included	core part only			core part with into cores			core part with out of cores		
Factors	MS	S	M	MS	S	M	MS	S	M
Index	ARI	AW'	AW"	MARI1	MAWS1	MAWM1	MARI2	MAWS2	MAWM2
Cluster 1	-1.1	-1.00	-0.99	-0.84	-0.94	-0.74	-0.58	-0.63	-0.84
Cluster 2	0.37	0.40	0.33	0.06	0.09	-0.05	-0.18	-0.25	-0.09
Cluster 3	0.76	0.52	0.71	1.2	1.29	1.34	1.44	1.71	1.61

MS = merging and splitting; M = merging; S = splitting

Table 14.5 summarizes some descriptive statistics of other blockmodels' characteristics:

- *The percentage of the into-cores (% into-cores) and out-of-cores (% out-of-cores) researchers.* The percentage of into-cores researchers is defined as the ratio between the number of researchers not in the cores in the first period and the number of researchers classified in the cores in the first period. On the other hand, the percentage of out-of-cores researchers is defined as the ratio between the number of researchers who joined the cores in the second period and the number of researchers classified in cores in the second period. Since Slovenian scientific disciplines are generally growing, the average share of into-cores researchers is lower than the share of out-of-cores researchers. However, a higher percentage of into-cores than out-of-cores researchers is typical for the unstable cluster of scientific disciplines.

- *The overall average core size (core size) and the overall number of researchers across clusters of scientific disciplines (# of res.).* The average size of the cores is relatively small, the smallest is in the case of an unstable cluster (3.9 researchers) and the highest in the case of the most stable cluster (5.8 researchers). While a higher average core size is typical for more stable scientific disciplines, a higher number of researchers per discipline is related to less stable scientific disciplines.

- *The number of scientific disciplines.* The stable/unstable cluster has the highest number of scientific disciplines, followed by the unstable and the stable cluster.

In the Slovenian Research Agency's classification scheme, the scientific fields are further divided into several scientific disciplines and then into scientific sub-disciplines. Based on this, most scientific disciplines from the fields of Engineering Sciences and Technologies (9 out of 14), Biotechnological Sciences (4 out of 5) and Social Sciences (4 out of

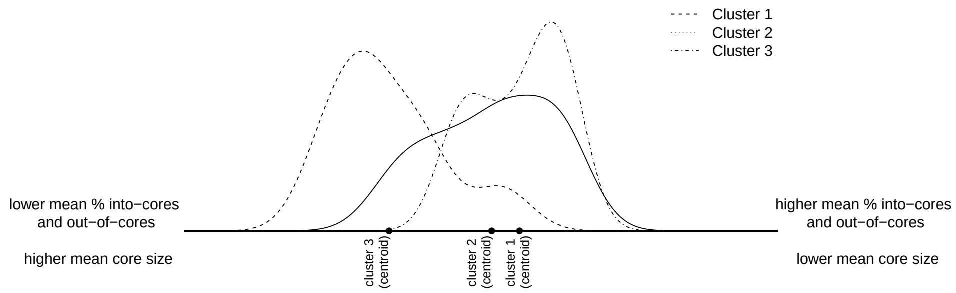
7) were classified in the unstable cluster. Most (5 out of 7) scientific disciplines from the Natural Sciences and Mathematics were classified in the stable/unstable cluster and three out of seven scientific disciplines from the field of Medical Sciences were classified in the most stable cluster. We can say the most stable scientific disciplines are from Medical Sciences and the most unstable from the technical field and Social Sciences. Similarly, [Melin, 2000] concluded that researchers from the Medical Sciences field almost always work in teams and from time to time collaborate with other teams. [Kyvik, 2003] reports that the greatest number of multi-authored papers in Norway is in Medicine.

Table 14.5: Basic descriptive statistics of the obtained clusters (averages on the level of clusters are reported)

Cluster	% into cores	% out of cores	core size	# of res.
Cluster 1 (N = 13) (unstable)	72	67	3.9	322
Cluster 2 (N = 22) (stable/unstable)	69	58	4.2	274
Cluster 3 (N = 8) (stable)	53	48	5.8	272

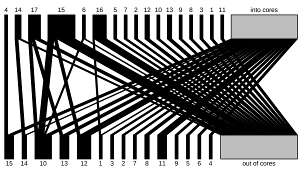
Since a scientific discipline's affiliation with a certain cluster is a categorical variable, one can check if the basic characteristics presented in Table 14.5 can be used to predict the cluster in which a given scientific discipline belongs. To do this, discriminant analysis can be used. Since there are three clusters of scientific disciplines, two discriminant functions can be calculated based on the four explanatory variables presented in Table 14.5. Only the first discriminant function is statistically significant ($p < 0.01$), meaning that based on the four explanatory variables one can separate well between the stable cluster (cluster 3) on one side and stable/unstable and unstable clusters (clusters 1 and 2) on the other. The discriminant functions are defined as linear combinations of explanatory variables. In Figure 14.11, the first discriminant function is visualized. Here, the highest values of the first discriminant function are characterized by higher mean percentage of into-cores (0.74) and percentage out-of-cores researchers (0.20) and a lower average number of researchers in the cores (-0.31). The value of the standardized canonical coefficient of the explanatory variable 'number of researchers' is relatively low (-0.09) and is therefore not shown in Figure 14.11. The centroids for each cluster are also marked along with the distribution of the standardized discriminant function for the disciplines by clusters.

Figure 14.11: The distribution of standardized values of the first canonical discriminant function by clusters

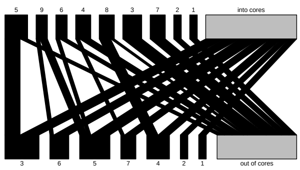


From each cluster of the scientific disciplines one was chosen to represent the cluster (the closest one to the centroid). The representative of the unstable cluster is the scientific discipline Educational Studies. Here, many into-cores and out-of-cores researchers can be seen. Most pairs of researchers classified in the same core at the first time point were not classified in the same core in the second period. The representative of the stable/unstable cluster is the scientific discipline of Textile and Leather. Here, the share of out-of-cores and into-cores researchers is lower. Some relatively large cores which remain relatively stable in the second period can also be observed. This is more typical for the representative of the stable cluster, namely Microbiology and Immunology.

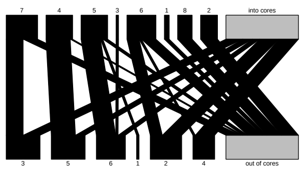
Figure 14.12: Visualizations of researchers' transitions between the cores into cores and out of cores for the two periods for the representative scientific discipline in each cluster



(a) Educational studies (N=379)



(b) Textile and leather (N=123)



(c) Microbiology and immunology (N=226)

14.5.2 Explaining the stability of cores

Version 1: To analyse the differences in the stability of cores among scientific fields, [Cugmas et al., 2016] classified the fields into two categories: fields Natural sciences and mathematics, Engineering sciences and technologies, Medical sciences, Biotechnical sciences into the category the *natural and technical sciences* and Social sciences and Humanities into the category the *social sciences and humanitites*. The selected factors lowering the stability of cores were the splitting of clusters and out-of-cores researchers and therefore, the stability of clusters were measured by the MAWS2. They shows that there is no statistically significant difference in average core stability.

Version 2: The study by [Cugmas et al., 2016] shows there is no statistically significant difference in average core stability when the core stability is measured by the MAWS2 and when scientific disciplines are analyzed at the level of scientific fields (the natural and technical sciences vs. the social sciences and the humanities[3]).

Given the high level of variability in the characteristics of the co-authorship networks and the blockmodel structures across scientific disciplines, the stability of the cores must be controlled by some additional network and blockmodel characteristics. Therefore, to explain the differences in the stability of cores across scientific fields, as controlling explanatory variables [Cugmas et al., 2016] also included in the linear model the characteristics of the networks (number of researchers, growth from the first period to the second period in the number of researchers and the growth of the density) and the obtained blockmodels (average core size, percentage of cores, presence of a bridging core in the first time point, percentage of out-goings).

[3]The scientific fields are classified into two categories: fields Natural sciences and mathematics, Engineering sciences and technologies, Medical sciences, Biotechnical sciences into the category the natural and technical sciences and Social sciences and Humanities into the category the social sciences and humanitites.

Table 14.6: The impact of the characteristics of the network, blockmodel and disciplines on the stability of the cores

	b	$SE(b)$	p	b	$SE(b)$	p
intercept	0.0906	0.2027	0.66	0.8349	0.1840	0.00
number of researchers (first time period)	-0.0002	0.0003	0.58	0.0001	0.0002	0.77
growth of number of researchers (1st and 2nd time period)	0.0010	0.0015	0.53	0.0004	0.0010	0.72
growth of density (1st and 2nd time period)	0.0015	0.0010	0.04	0.0091	0.0005	0.07
average core size (1st time period)	0.0625	0.0177	0.00	0.0053	0.0152	0.73
percentage of cores (1st and 2nd time period)	-0.0054	0.0049	0.28	-0.0069	0.0033	0.05
presence of the bridge (1st time period)	0.0404	0.0450	0.38	-0.0005	0.0313	0.99
percentage of out-of-cores		*not included*		-1.0160	0.1667	0.00
Humanities (reference category)						
Natural science and math.	-0.1511	0.0892	0.10	0.0378	0.0680	0.58
Engineering sciences and tech.	-0.0120	0.0834	0.89	0.1339	0.0615	0.04
Medical sciences	-0.0850	0.0954	0.38	0.1421	0.0748	0.07
Biotechnical sciences	-0.0353	0.1008	0.72	0.0338	0.0694	0.63
Social sciences	-0.0707	0.0844	0.41	0.0847	0.0626	0.19
Number of obs. (disciplines):			*43*			*43*
Adjusted R^2:			*0.23*			*0.65*
F Statistics:			*2.151 (11; 31) ($p < 0.05$)*			*7.375 (12; 30) ($p < 0.01$)*
Method of estimation:			*Least Squares Method*			*Least Squares Method*

The main results are presented in Table 14.6. Here the Humanities is used as the reference field since many studies suggest the social sciences are becoming more similar to the natural and technical sciences regarding publishing behavior [Kyvik, 2003, Kronegger et al., 2015]. In Table 14.6, one can see there are no statistically significant differences between the Humanities and other scientific fields when the percentage of out-goings is not included in the model. However, when the percentage of out-goings is included in the model, the differences in the mean stability of cores between the Humanities and the Engineering Sciences and Technologies and the Humanities and the Medical Sciences become statistically significant (at $p < 0.10$). Here, the scientific disciplines of both fields are seen as more stable than the Humanities. Since the percentage of out-of-cores researchers forms part of the core stability index, the statistically significant differences between the mentioned scientific fields are mainly the consequence of the splitting of cores.

The effects of some controlling explanatory variables are statistically significant at ($p < 0.10$) as well. When the variable percentage of out-of-cores researchers is included in the model, the growth of the density and the average core size in the first time period is statistically significant. The density is defined as the share of all realized ties from all possible ties. The value is typically greater in the case of smaller networks with a low percentage of researchers in the periphery and many cores with a lot of researchers included. Therefore, together with the variable average core size, it can be argued that in the case of greater density there are more researchers who co-authored only occasionally (semi-periphery) and more complete cores with a higher number of researchers. The probability of creating ties with new researchers is therefore lower and the stability of the cores is higher. Similarly, [De Haan et al., 1994] mentioned that the size of a research group affects the persistence of collaboration.

When the percentage of out-of-cores researchers is included in the model, the growth of density and the percentage of core are statistically significant (at $p < 0.10$) along with the controlling explanatory variable percentage of out-of-cores researchers, which is highly statistically significant ($p < 0.01$). Since the latter is part of the definition of response variable, the percentage of explained variance of stability of cores is much higher in the model that includes percentage of out-of-cores researchers ($Adjusted R^2 = 0.65$) compared to the model where this variable is not included ($Adjusted R^2 = 0.23$).

14.6 Conclusions

It is crucial to understand how modern science works to ensure appropriate research and development policies are adopted that lead to improved scientific output. Modern information databases containing information about scientific bibliographic units can help in understanding the formation and maintenance of co-authorships among researchers. Although the borderline of scientific collaboration is unclear and there is no accurate way to measure it [Katz and Martin, 1997b], co-authorships can be seen as a rough operationalization of scientific collaboration, which is one of the primary results of scientific collaboration and represents one of the most formal manifestations of scientific communication [Groboljsek et al., 2014]. The co-authorship patterns were studied through co-authorship networks. These are networks where the vertices present authors (or researchers) and a link between them exists if they co-authored at least one scientific bibliographic unit. [Kronegger et al., 2011] analyzed the co-authorship networks of four Slovenian scientific disciplines (Physics, Mathematics, Biotechnology and Sociology) in four periods (from 1990 to 2010). Only by observing the number of links among different scientific disciplines could they confirm that different co-authorship cultures exist between ”lab” and ”office” scientific disciplines. Publishing in co-authorship is more common in ”lab” sciences while solo-authored scientific units are more common in ”office” scientific disciplines where teamwork is not so crucial for the research. [Hu et al., 2014] classified four scientific disciplines in two groups: theoretical disciplines and experimental disciplines. They observed a stronger correlation between collaboration and productivity in experimental disciplines compared to theoretical ones.

However, one of the chief interests of the study by [Kronegger et al., 2011] was on the global network structure. To analyze this, they used generalized blockmodeling on network slices in four 5-year consecutive periods. They confirmed the network structure of multi-cores, semi-periphery, and periphery being present in all scientific disciplines. It can happen that the mentioned structure is not so outstanding at the earliest time points in some scientific disciplines. They defined the core as a group of researchers who very systematically co-author with each other, but who usually do not collaborate with researchers from the other cores. The semi-periphery consists of authors who collaborate with others inside the network, but in a less systematic way. It is not possible to cluster researchers from the semi-periphery into several well-separated clusters. The last part, the periphery, is the biggest part of the analyzed networks. These are authors who publish at least one bibliographic unit but as a single author or with researchers from abroad (with researchers not registered at the Slovenian Research Agency). Besides the main three types of mentioned positions, they observed so-called bridging cores. These are groups of researchers who collaborate with at least two other cores, which are not connected.

[Cugmas et al., 2016] extended the analysis at the level of all Slovenian scientific disciplines. Like [Kronegger et al., 2011], they analyzed data for the period between 1991 and 2010, but only analyzed the data in two 10-year long periods. The wider time span has an effect on the network density. Despite this, there are some scientific disciplines without any links in the first or second period, e.g. Theology. These kinds of scientific disciplines were removed from the analysis, leaving 43 out of 72 scientific disciplines for further analysis. The assumed multi-core–semi-periphery–periphery structure was confirmed as being present in all analyzed scientific disciplines. In many of them, bridging cores are also found. On average, the number of researchers is increasing in time, also reflected in the higher average core size which is higher in the second period in both scientific disciplines from the fields of the natural and technical sciences and scientific disciplines from the social sciences and humanities. Here, the average size of cores is smaller in the social sciences and humanities in both time periods. The differences may be affected by the fact that authors from abroad are not included in the analysis since the rate of co-authored publications with researchers from abroad is higher in fields of the natural and technical sciences than in the social sciences and humanities. As reported by [Kronegger et al., 2011], the main part of co-authorship networks is represented by authors from the periphery, which is generally decreasing over time.

Another important property of co-authorship networks is that the cores can emerge in time, disappear, split, or merge. To measure the stability of cores, operationalized with these four rules in different ways, several indices were proposed. The value of each was calculated for each scientific discipline and, based on this, the scientific disciplines were clustered in three clusters. The observation of these clusters reveals that, according to the values of the proposed indices, they are mainly characterized by different levels of stability of the clusters and can therefore be ordered from least to most stable. The majority of scientific disciplines were classified in the stable-unstable cluster (22 scientific disciplines) while only a few were classified in the most stable cluster (8 scientific disciplines). It turns out that the average percentage of researchers classified in the cores in both periods is increasing along with the stability of the clusters. On the other hand, the percentage of researchers leaving the cores in the first time period and the percentage of researchers joining the cores in the second period is decreasing with the average stability of cores by the obtained clusters. The average core size is higher in the most stable cluster of scientific disciplines, indicating the existence of well-established scientific research teams in these scientific disciplines. [De Haan et al., 1994] mentioned that the size of a research group affects the persistence of collaboration.

A higher average number of researchers is associated with a lower level of stability of the cores. There are several explanations for this phenomenon, including the fact there are many opportunities to collaborate with different researchers in bigger scientific disciplines. The others are chiefly related to national research and development policies (e.g., the Young Researchers Program) and the nature of the work in such scientific disciplines (e.g., lab vs. office scientific disciplines or natural and technical sciences vs. social sciences and humanities).

To explain the differences between the natural and technical sciences and the social sciences and humanities, [Cugmas et al., 2016] performed a linear regression in which several network- and blockmodel-related variables (number of researchers in the scientific discipline, growth in number of researchers, growth in density, average core size, average percentage of cores, presence of a bridge) were included in the model as explanatory variables, while the stability of cores (response variable) was operationalized by the MAWS2, where the splitting of cores and out-of-cores researchers reduces the value of an index and thus indicates lower core stability. There were no statistically significant differences in the mean stability of cores between the natural and technical sciences on one hand and the social sciences and humanities on the other. This could be caused by many differences in the publication culture within these two groups of scientific disciplines (which is also a consequence of the characteristics of the particular national classification scheme of scientific fields, disciplines and sub-disciplines). In fact, even within some scientific disciplines the publication cultures vary widely. [Moody, 2004] found that quantitative work is more likely to be co-authored than non-quantitative work in Sociology.

However, when the analysis is performed on the level of scientific disciplines, the scientific discipline Natural Sciences and Mathematics is statistically significantly (at $p < 0.10$) less stable than the field of the Humanities. The growth of density and the average size of cores are also statistically significant (at $p < 0.05$) and positively correlated with the stability of the cores. When the additional variable percentage of out-of-cores researchers is included in the model, the difference in the average stability of cores between the Humanities and Medical Sciences becomes statistically significant (at $p < 0.10$). Here, it must be highlighted that when the variable percentage out-of-cores researchers is included in the model, only the splitting of cores is seen as a factor indicating lower core stability.

REFERENCES

[Abbasi et al., 2011] Abbasi, A., Altmann, J., and Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics*, 5(4):594–607.

[Abbasi et al., 2012] Abbasi, A., Hossain, L., and Leydesdorff, L. (2012). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics*, 6(3):403–412.

[Airoldi et al., 2007] Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2007). Combining stochastic block models and mixed membership for statistical network analysis. In *Statistical Network Analysis: Models, Issues, and New Directions*, pages 57–74. Springer.

[Amat and Perruchas, 2015] Amat, C. B. and Perruchas, F. (2015). Evolving cohesion metrics of a research network on rare diseases: a longitudinal study over 14 years. *Scientometrics*, pages 1–16.

[Batagelj et al., 1992] Batagelj, V., Ferligoj, A., and Doreian, P. (1992). Direct and indirect methods for structural equivalence. *Social Networks*, 14:63–90.

[Blondel et al., 2008] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

[Chinchilla-Rodríguez et al., 2012] Chinchilla-Rodríguez, Z., Ferligoj, A., Miguel, S., Kronegger, L., and de Moya-Anegón, F. (2012). Blockmodeling of co-authorship networks in library and information science in Argentina: a case study. *Scientometrics*, 93(3):699–717.

[Cugmas et al., 2016] Cugmas, M., Ferligoj, A., and Kronegger, L. (2016). The stability of co-authorship structures. *Scientometrics*, 106(1):163–186.

[De Haan, 1997] De Haan, J. (1997). Authorship patterns in dutch sociology. *Scientometrics*, 39(2):197–208.

[De Haan et al., 1994] De Haan, J., Leeuw, F., and Remery, C. (1994). Accumulation of advantage and disadvantage in research groups. *Scientometrics*, 29(2):239–251.

[Ferligoj and Kronegger, 2009] Ferligoj, A. and Kronegger, L. (2009). Clustering of Attribute and/or Relational Data. *Metodološki Zvezki (Advances in Methodology and Statistics)*, 6(2):135–153.

[Ferligoj et al., 2015] Ferligoj, A., Kronegger, L., Mali, F., Snijders, T. A., and Doreian, P. (2015). Scientific collaboration dynamics in a national scientific system. *Scientometrics*, 104(3):985–1012.

[Frenken et al., 2005] Frenken, K., Hölzl, W., and de Vor, F. (2005). The citation impact of research collaborations: the case of european biotechnology and applied microbiology (1988–2002). *Journal of Engineering and technology Management*, 22(1):9–30.

[Groboljsek et al., 2014] Groboljsek, B., Ferligoj, A., Mali, F., Kronegger, L., and Iglic, H. (2014). The role and significance of scientific collaboration for the new emerging sciences: The case of slovenia. *Teorija in praksa*, 51(5):866.

[Hollis, 2001] Hollis, A. (2001). Co-authorship and the output of academic economists. *Labour Economics*, 8(4):503–530.

[Hu et al., 2014] Hu, Z., Chen, C., and Liu, Z. (2014). How are collaboration and productivity correlated at various career stages of scientists? *Scientometrics*, 101(2):1553–1564.

[Hubert and Arabie, 1985] Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2(1):193–218.

[Katz and Martin, 1997a] Katz, J. S. and Martin, B. R. (1997a). What is research collaboration? *Research policy*, 26:1–18.

[Katz and Martin, 1997b] Katz, J. S. and Martin, B. R. (1997b). What is research collaboration? *Research policy*, 26(1):1–18.

[Kronegger et al., 2011] Kronegger, L., Ferligoj, A., and Doreian, P. (2011). On the dynamics of national scientific systems. *Quality & Quantity*, 45(5):989–1015.

[Kronegger et al., 2015] Kronegger, L., Mali, F., Ferligoj, A., and Doreian, P. (2015). Classifying scientific disciplines in slovenia: A study of the evolution of collaboration structures. *Journal of the Association for Information Science and Technology*, 66(2):321–339.

[Kyvik, 2003] Kyvik, S. (2003). Changing trends in publishing behaviour among university faculty, 1980-2000. *Scientometrics*, 58(1):35–48.

[Lancichinetti and Fortunato, 2009] Lancichinetti, A. and Fortunato, S. (2009). Community detection algorithms: a comparative analysis. *Physical review E*, 80(5):056117.

[Laudel, 2002] Laudel, G. (2002). What do we measure by co-authorship? *Research Evaluation*, 11(1):3–15.

[Lee and Bozeman, 2005] Lee, S. and Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social studies of science*, 35(5):673–702.

[Mali et al., 2010] Mali, F., Kronegger, L., and Ferligoj, A. (2010). Co-authorship trends and collaboration patterns in the Slovenian sociological community. *Corvinus Journal of Sociology and Social Policy (CJSSP)*, 1(2).

[Matias and Miele, 2015] Matias, C. and Miele, V. (2015). Statistical clustering of temporal networks through a dynamic stochastic block model. *arXiv preprint arXiv:1506.07464*.

[Melin, 2000] Melin, G. (2000). Pragmatism and self-organization: Research collaboration on the individual level. *Research policy*, 29(1):31–40.

[Melin and Persson, 1996] Melin, G. and Persson, O. (1996). Studying research collaboration using co-authorship. *Scientometrics*, 36(3):363–377.

[Moody, 2004] Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American sociological review*, 69(2):213–238.

[Ponomariov and Boardman, 2016] Ponomariov, B. and Boardman, C. (2016). What is co-authorship? *Scientometrics*, pages 1–25.

[Rand, 1971] Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850.

[Saporta and Youness, 2002] Saporta, G. and Youness, G. (2002). Comparing two partitions: Some proposals and experiments. In *Compstat*, pages 243–248. Springer.

[Tibshirani et al., 2001] Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423.

[Xing et al., 2010] Xing, E. P., Fu, W., Song, L., et al. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, 4(2):535–566.

[Xu and Hero III, 2013] Xu, K. S. and Hero III, A. O. (2013). Dynamic stochastic blockmodels: Statistical models for time-evolving networks. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, pages 201–210. Springer.

[Žiberna, 2014] Žiberna, A. (2014). Blockmodeling of multilevel networks. *Social Networks*, 39:46–61.

CHAPTER 15

SOME APPLICATIONS OF PARTITIONING LARGE NETWORKS

Vladimir Batagelj[1], Anuška Ferligoj[2], and Patrick Doreian[3]

[1]FMF, University of Ljubljana
[2]FDV, University of Ljubljana
[3]Pittsburgh

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,* **ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 16

DIFFERENT ANALYSES OF THE SAME NETWORK DATA SET

TBD

X Y Z

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,* **ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

CHAPTER 17

CONCLUSIONS AND DIRECTIONS FOR FUTURE WORK

PATRICK DOREIAN[1], ANUŠKA FERLIGOJ[2], AND VLADIMIR BATAGELJ[3]

[1] Pittsburgh
[2] FDV, University of Ljubljana
[3] FMF, University of Ljubljana

REFERENCES

1. J. S. Kilby, "Invention of the Integrated Circuit," *IEEE Trans. Electron Devices,* **ED-23,** 648 (1976).
2. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, Chapter N-1, McGraw-Hill, New York, 1962.
3. J. Lee, K. Mayaram, and C. Hu, "A Theoretical Study of Gate/Drain Offset in LDD MOSFETs" *IEEE Electron Device Lett.,* **EDL-7**(3). 152 (1986).
4. A. Berenbaum, B. W. Colbry, D.R. Ditzel, R. D Freeman, and K.J. O'Connor, "A Pipelined 32b Microprocessor with 13 kb of Cache Memory," it Int. Solid State Circuit Conf., Dig. Tech. Pap., p. 34 (1987).

TOPIC INDEX

AUTHOR INDEX