A NETWORK APPROACH FOR THE SOCIOLOGICAL STUDY OF SCIENCE AND KNOWLEDGE

MODELLING A DYNAMIC MULTILEVEL NETWORK

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy
in the Faculty of Humanities

2021

Alejandro Espinosa-Rada

SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF SOCIOLOGY

Blank page

Table of Contents

Word count: 48,963

List of tables

List of figures

Abstract

The thesis investigates the usage of *multigraphs* to understand citation patterns among researchers in the Chilean astronomical and astrophysics community. The usage of multigraphs to study scientific networks in local contexts has been acknowledged in early developments from the sociological study of science and knowledge but has been scarcely addressed in current empirical research. This research will show that multiple networks can contribute to investigating why scientific networks evolve, considering stable processes that mix social, cognitive, and situational dimensions.

In this research, processes of *group formation* using *multigraphs* are considered to enlighten the patterns of citations among researchers. The co-evolution of different networks is analysed, incorporating different levels (three-modes) and types of relationships that are jointly investigated. First, to understand how a group of academics generate interpersonal intercitations after the arrival of the Atacama Large Millimeter/submillimeter Array. And, secondly, to inquire how the local scientific community prepares for the arrival of the Vera C. Rubin Observatory.

For the analysis, it is used quadratic assignment procedures and stochastic actor-oriented models. This research offered methodological advances to understand multilevel networks exploring new goodness of fit, often used in statistical network models, for multiplex and three-mode multilevel networks. And suggest as an analytical strategy the analysis of samples of multilevel networks to investigate broader communities.

The research shows that the usage of citation-based measures is difficult to understand and that the consideration of different interpersonal relationships and the context allowed recovering the social dimension of the intercitation. The social relationships grounded on scientific collaboration and space proximity based on institutional affiliation are more accurately suited to understand the co-evolution of the networks and the intercitation among astronomers than cognitive-based networks when measured as the tendency to publish in similar journals. And, in the broader community, there is a tendency upon intercitation among researchers affiliated in the same external research centres creating closure in scientific niches

(i.e., research centres) as a community's tendency towards diversity and multi-connectivity.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses.

Acknowledgements

Thank you to the supervision team Dr Elisa Bellotti, Prof Martin Everett, and Dr Termeh Shafie. In particular, I am thankful to Elisa for her guidance, commitment, and support of my thesis. Her encouragement to build my academic's personal network and openness to explore new areas of research were crucial to developing this thesis and for my personal development as a researcher. To Martin for his supervision, advice, and encouragement. His guidance motivated me to discover some intersections between mathematics, networks, and sociology, and his advice was decisive in navigating this thesis and academia. And Termeh for her support, guidance, and for allowing me to participate as her teaching assistant in different courses from social network analysis at the University of Manchester.

The Mitchell Centre for Social Network Analysis and its members have offered me a vibrant and enthusiastic intellectual home. Thanks to Prof Emmanuel Lazega, Prof Peter Wade, and Dr András Vörös for the insightful comments and discussion during the viva, and for taking the time to reviewing the final version of this thesis. Also, thanks to Dr Johan Koskinen and Dr András Vörös for taking the time and effort as my annual reviewers to provide valuable feedback on the earlier stages of this research. In particular, to Johan, whose support in critical thinking towards statistics, generosity in exploring different areas of research and guidance allowed to push forward this thesis. And a huge thanks to the other members of the centre, especially for a couple of coffees with Prof Nick Crossley, the programming and technical help, and exchanges in the corridor with Dr David Schoch, and the virtual meetings during the pandemic with Dr Tomáš Diviák. You all provided part of the best at the Arthur Lewis Building (ALB), the Humanities Bridgeford Street Building (HBS) and virtual offices, respectively. Also, to my friends and colleagues of the centre, Francisca Ortiz, Pete Jones, Ella Guest, Dorottya Hoór, and Dominika Czerniawska, it was fun to hang out with you all, sharing also during the Mitchell Centre's seminars and the network conferences. And, a thoughtful thanks to my friends from the Department of Sociology and the Department of Social Statistics at the University of Manchester, Ali Siles, Denisse Sepúlveda, Andrea Lizama, Javier Castillo, Alexandrina Vanke, Hannah Haycox, Natalie-Anne Hall, Owen Abbott, Jaime García, Niamh Kavanagh, Kaidong Yu, Jessica Mancuso, Rohini Rai, Tugba Aydin, Rachel Katz, Neta Yodovich, Anh-Susann Pham, Tiba Bonyad, Joshua Bunting, Alexus Davis, Cedomir Vuckovic, Caitlin Schmid, Martin Greenwood, Sophie Atherton, and Mariana Dias, it was fun to be with you all, you make the best of daily life.

I also highly appreciate the different encounters, especially at the University of Groningen, and the guidance of Prof Tom Snijders. You have been a mentor and an inspiration in the most crucial moments of my PhD. Thank you for your time, for the academic advice and the help provided. Likewise, I am thankful for the collaboration and advice from Prof Christoph Stadtfeld, for our different encounters during networks' conferences, and for allowing me to visit the Social Networks Lab in ETH-Zürich. In the lab I had the chance to share with an exciting group – Prof Ulrik Brandes, Dr András Vörös, Dr Per Block, Dr Zsófia Boda, Dr Viviana Amati, Dr Timon Elmer, Kieran Mepham, Marion Hoffman, Álvaro Uzaheta, Simon Päpcke, Julian Müller, Sara Colella, and Felix Schönenberger - which push forward some of the ideas that I developed in this thesis. I am grateful to the colleagueship through the REDES list seminar organised by Francisca Ortiz during the COVID-19 pandemic, especially to Prof Isidro Maya-Jariego and Prof José Luis Molina. And I am also in debt to the kindness of Dr Matías Bargsted for receiving me twice at the Department of Sociology at the Pontifical Catholic University of Chile, to Daniel Pérez from Epistemonikus Foundation for his advice in the data collection of this thesis, and the whole community of the Chilean Society for Social Network Science (CHISOCNET). I am sure that, as Joaquín Torres García mentioned, I hope that in CHISOCNET, we will be able to invert the map to claim through the social network community that '*the tip of America, from now on, extending, insistently points to the South, our North*'.

Finally, I would like to thank my family. To my parents, Jorge Espinosa and Elena Rada, and my siblings María Elena, Sofía and Cristóbal, thank you for your unconditional love and constant support. And special thanks to my wife Francisca Ortiz, for your love, for growing together as partners, for inspiring, caring, supporting, and believing in me. I appreciate everything you have done for the last several years.

This thesis was supported by the National Commission for Scientific and Technological Research in Chile (ANID PFCHA/DOCTORADO BECAS CHILE/2017, grant number 72180242).

Blank page

Chapter 1.

Introduction

1.1. Introduction

Chile has been called the ‘capital’ of astronomy and astrophysics because during the decade of 2020-2030 it will have nearly 70 per cent of the world's ‘viewing capacity’ (Bronfman et al., 2002; Gibert, 2011; CONICYT, 2012; Barandiaran, 2015; Espinosa-Rada et al., 2019; Guridi et al., 2020). For this reason, the Chilean astronomical community has a competitive advantage, with the potential to shape the research front of the discipline (Espinosa-Rada et al., 2019) and to create a spillover to help to develop the country (CONICYT, 2012; Guridi et al., 2020).

From a network perspective, this case study is appealing because the development of this community involves the participation of researchers and organisations. Some of the particularities are that access to the telescopes shapes this scientific community at a national level. In which Chileans have 10 per cent of the observational times of the telescopes in the country. The community have been described as a small community (López et al., 2005; Gibert, 2011; and SOCHIAS census 2009, 2013, 2016, 2019), varying from only 21 researchers in 2000 (Gibert, 2011) to up to 255 professional researchers since the last census of the Chilean Astronomy Society (SOCHIAS) in 2019. The usage of these facilities depends on the local institutional affiliation guarantee given by a ‘white list’ of researchers with access to these observatories. The role and interdependency of the organisations and the astronomers in the development of the community can be represented as a *multilevel network* (Lazega & Snijders, 2016) shaped in an organisational field (Galaskiewicz & Wasserman, 1989; Scott, 1995; Powell et al., 2005).

This thesis uses the astronomical and astrophysics community in Chile as an empirical case study to understand the multilevel interrelation between researchers and organisations by proposing some advances in both empirical and methodological issues. It offers a perspective that connects with previous research in the sociology of science and knowledge (Mullins, 1972, 1973; Chubin, 1976), with contemporaneous advances from the social network perspective (Lazega & Snijders, 2016), highlighting some theoretical discussions and possibilities to model scientific networks from a sociological angle. Such issues and perspectives are discussed in three articles that constitute the body of the thesis.

The first article (Chapter 2) explores some of the limitations and methodological considerations of using citation as a proxy for social and cognitive relationships (Merton, 2000; White et al., 2004). As a previous theoretical background of the first article, this chapter (section 1.2.1) addressed that the ties in citation networks are difficult to understand, and there are different interpretations about their meaning (Mulkay, 1974; Gilbert, 1977; Nicolaisen, 2008).

The second article (Chapter 3) investigates the pattern of citations in combination with other types of relationships in a group of established academics in the development of the Chilean astronomical and astrophysics community. From a network perspective, the theoretical background of this study recognises the discussion of section 1.2.1 of this chapter, considering that the citations and co-authorships are proxies of social relationships deduced from the 'formal channel' of communications used through bibliometric information (Schrum & Mullins, 1988; Zuccala, 2006). Chubin (1976) conjecture that multiple relationships can help to recover the underlying structure using these indicators. In addition, the paper identifies the relevance of the evolution of the local discipline considering the processes of peer recognition (Zuckerman, 1967; Merton, 1968a) – as is reviewed more extensively in section 1.2.2 - and the *group formation* (Mullins, 1972, 1973) – reviewed in more details in section 1.2.3 and 1.2.4.

The last article (Chapter 4) moves one step forward and considers researchers and organisations beyond the group of established academics in the departments of astronomy and astrophysics. Theoretically, this issue is explored in section 1.2.5 of this chapter considering two issues: the first is regarding the delimitation of *cohesive subgroups*, and the other is how to investigate the presence of similar mechanisms among organisations.

In the remaining pages of this Introduction, I overview first the theoretical background of the thesis to explore how different networks allowed understanding scientific networks from a multilevel perspective. Then, I give some context of the astronomical and astrophysics community in Chile that is further explored in the empirical chapters of this thesis. Finally, I present how the three articles are connected following the theoretical background presented in this chapter.

1.2. Towards a Multilevel Perspective for Socio-Cognitive Networks

This section gives an overview of the theoretical backgrounds of the thesis exploring the combination of different networks for the understanding of socio-cognitive networks (Merton, 2000; White et al., 2004; White, 2011) – as the conflation of social, cognitive, and situational dimension in science. When only one type of relationships is analysed, it is likely to produce a 'structural confusion' (Holland & Leinhardt, 1974) in which the 'true' underlying structure will require a *multigraph* to describe the scientific networks (Chubin, 1976).

Holland & Leinhardt (1974) identify that a possible distortion in the measurement of social networks could be associated with different components of the relationships, which is, in principle, an operationalisation issue. From their perspective, a

> 'mechanism that produces distortion might be termed *'structural confusion.'* The *true structure* underlying a group may have different components such as admiration, friendship, and common side interests. Indeed, a single term such as friendship could have effectively different role connotations to group members who occupy different positions in the structure. This type of structure would really need a '*multigraph*' to describe it – for example, several different directed graphs on the same nodes considered collectively.' (Holland & Leinhardt, 1974: 208, emphasis is mine)

The ‘structural confusion’ is considered an issue in the study of science and knowledge, in which the combination of graphs could represent more accurately the ‘real structure’ underpinning scientific networks (Chubin, 1976). In the studies based on citation and other scientific relationships (e.g., co-authorship, colleagueship, trusted assessorship, mentorship), the argument was made, assuming that ‘If the structure changes with each *distinct graph*, then all must be constructed as approximations of the *true structure*. Taken together, these approximations should yield the least distorted representation of the true specialty structure.’ (Chubin, 1976: 455, *emphasise is mine*)[1]. Because of the disjunction of the social and cognitive dimension of scientific networks, creates a fictional detachment of structures and a disciplinary division between bibliometrics - often focused on the cognitive dimension - and the sociology of science – that emphasise the social dimension - (Merton, 2000; Gläser, 2001). The confusion relies on treating citation as part of only one of the dimensions, without considering its mixture or conflation as social and cognitive at the same time as a complex social-cognitive network. To overcome the ‘structural confusion’, *multigraphs* are considered in this thesis to explore social, cognitive, and situational components – often when actors are nested in laboratories or organisations - of scientific networks.

1.2.1 The ‘Structural Confusion’ in Citations

In the sociological analysis of scientific and knowledge networks, scientific relationships as a cognitive or social factor are not easy to understand. Citation is one of the puzzling elements in the delimitation between cognitive and social factors, in which some researchers treated it as a ‘cognitive’ dimension, while others consider that it is a ‘social’ dimension (Lievrouw et al., 1987; Merton, 2000; Leydesdorff, 2008). This thesis follows previous research considering that citation is an approximate measure of social processes that are social and intellectual at the same time (as in Crane, 1972: 20; Chubin, 1976: 451-452; White, 2011[2]).

---

[1] Crane mentioned that ‘Since scientists in research areas can have a number of different types of social relationships with one another, it is necessary to use a variety of indicators to measure social organization’ (1972: 41).
[2] Howard D. White used ‘socio-cultural’ instead, which might be more appropriate to emphasise the contextual dimension in which researchers are embedded.

In empirical research, the distinction between citation as a 'cognitive' or 'social' dimension is considered part of the operationalisation in the type of relationships for networks analysis. Some researchers that use networks to study scientific specialities, such as Mullins (1972), differentiate between cognitive (characterised by the paradigm development, problem success and puzzle-solving) and social dimensions (focusing on communicational, co-authorship, colleagueship, and apprenticeship). Small (1977) made a similar distinction, while less specific, emphasising that co-citations allowed distinguishing the cognitive and/or the social state of a speciality. Schrum & Mullins (1988) differentiated between 'interaction' considered as communication, information flow or general contact (e.g., co-authorship, 'in-house' citation), and 'interest' that can be inferred from co-occurrence on bibliographic or referential lists (e.g., citations). Moody (2004) was emphatic in distinguishing between citation and social networks, in which the first does not capture the informal interaction structure compared to the last. Leydesdorff (2008) traces a distinction, arguing that co-occurrences in bibliometric research should be treated differently from social networks that often refer to concrete relations. More recently, Basov, Breiger and Hellsten (2020) distinguished between the social ties (e.g., friendship, information exchange, or co-citations among actors) and the semantic associations (e.g., symbols, ideas, meanings).

Citations are perhaps one of the less clear types of relationships in which the 'structural confusion' can be further considered because citations are often a black box. Citations are usually investigated using the *products* of science (e.g., such as papers, books, technologies), and the trace of these *products* involves stabilisation of previous processes that then become 'black boxes' (Whitley, 1970)[3]. The *products* tend to give proper credits to other scientists through bibliographic references producing a traceable scientific network and receiving evaluation from other researchers. The citation does not address 'the reasons why scientists

[3] While stable relational patterns can reveal the presence of communication, it does not unravel the nature and specificity of those relations (Lievrouw, 1989), which can be further explored through mixed methods (Mitchell, 1969; Bellotti, 2015). Among many of the issues that can be further explored are, for example, what is at stake in the research, the content of controversies, the struggles between members of similar invisible colleges, among others. Topics that have been extensively addressed in the social studies of science and technology (e.g., Felt et al., 2017).

normally cite other papers, and why authors choose to cite particular papers rather than others.' (Gilbert, 1977: 114). There is no clarity on what citation is measuring, a perspective shared by Mulkay (1974). Compared with the citation-based measures, the co-word analysis (Callon et al., 1983; Latour et al., 2012) tend to give direct access to the content of the research topics regarding the use of concepts, words, and the co-occurrence of terms. It has been criticised that those terms could have different meanings in a different context. Leading that their use is not specific enough, requiring a more accurate boundary and stability in which their usage could be meaningful (Leydesdorff, 1997; van den Besselaar & Heimeriks, 2006) and explored with more details.

There are different interpretations about the usage of citations (Nicolaisen, 2008). A first perspective considers that citation is an institutional device that solves the problems of rights and priority claims and emphasises which work was fruitful for the contributions (Merton, 1973: chapter 14). Citations are granted as an argument implicitly attached to other oeuvres (e.g., Smith, 1981), and the references that achieve authority and are more visible becoming more recognised. The institutional perspective is criticised considering that there are references that are challenged, contradicted, or unimportant, which are aside from the arguments, and neither is clear how 'findings' reported in a paper can be matched as 'property' or 'income' (Gilbert, 1977). Citations can be used to understand the institution of science through maps of knowledge. The citations can be considered signals used to generate 'maps' of science, in which citations have similarities when they have related topics, titles, or commonly perceived citations (Morris & der Veer, 2009). For example, in co-citations, the focus is on 'citer's consensus' in an open-ended field where other authors perceived the work of two authors to be related (White, 2011).

A second interpretation is treating references as a 'tool of persuasion' oriented to the scientific community or part of it. If some consensuses were previously achieved, then the reference becomes scientific knowledge. This dimension emphasised the reward of recognition, validity, and significance of the article's work. The paper's recognition changes its status as a document that is treated as new, important, and valid. These qualities are not evident to the paper's audience, and these findings tend to rely on previously accepted references making it more effective to cite a paper with scientific authority. Considering the scientists' interest, the citation is often used as a persuading strategy (Latour & Woolgar 1986) oriented to other colleagues (justification of the position, demonstrate novelty, or how the findings illuminate or solve previous problems). Researchers tend to cite 'erroneous' papers and ignore 'trivial' or 'irrelevant' papers (Gilbert, 1977). From this second perspective, the references are embedded in the papers and act as an allegiance of a particular scientific community sector. According to this perspective, there are two common approaches to identify persuasion (White, 2004), one is according to the citation contexts, and the other is according to the choice of the cited works regardless of their content.

Some empirical research relativise that citation is a matter of persuasion. Different empirical studies classified the references according to their content and intention to disregard that citation is only motivated by interest or a 'tool of persuasion'. For example, Chubin & Moitra (1975), following Moravcsik & Murugesan (1975), classify the references considering that some of the citations are essential or supplementary for the main argument, and other citations negate other references. The essential or supplementary arguments are thereafter continued to be cited, while the negated documents decline their citations in time, which tend to be institutionally constrained through journals peer-reviewers or disciplinary consensus. Most of the time, the authors' intentions in using the references are not usually available to analyse how they are expressed in the content of the work, thus becoming highly complex to have a deeper understanding (Camacho-Miñano & Núñez-Nickel, 2009). If researchers know about the topic, there is a variety of interpretation of the given references, making this enterprise complex (Nicolaisen, 2008).

The third interpretation about citations is the handicap principle, in which the references are utilised as threat signals (Nicolaisen, 2008). Latour (1987) considers that reference interplays between transforming a fact into fiction or vice versa through two strategies. The first strategy is incorporating, or not, references that give insights into whether the citation is strong and serious as *stacking* of references strategy. The second strategy is to *modalise* the cited documents modifying the references to align with the article's argument. This perspective has been criticised because it gives an arbitrary element in which it seems that authors cite whatever is needed for their proposes. Nicolaisen believes that ‘the handicap principle ensures that citing authors honestly credit their inspirations and sources to a tolerable degree – enough to save the scientific communication system from collapsing’ (2008: 629). Nicolaisen assumed that scholars are well-informed actors that have a broader understanding of the literature and field covered by sets of references. Citing, for example, 'pioneer' references suggested from peer-reviewers, technical papers, or authors that become classics in a specific area of research. Other researchers recognise the references used, which could be potentially challenged (e.g., in peer-review journals or further publications).

From a network perspective, a different and complementary interpretation is that citation ‘can be *interpreted* as networks of interpersonal contacts’ (Lievrouw, 1989: 617). While 'not always involves underlying personal exchanges and that unknown references are an essential component…' (Milard, 2014: 2461). From this perspective, different studies have assumed that when researchers know each other, they tend to cite more often (Schrum & Mullins, 1988; White et al., 2004). The tendency is explained assuming that there is a homophilous effect (Lazarsfeld & Merton, 1954; Feld, 1981; McPherson et al., 2001) – or the effect that actors that share similar social attributes tend to be attracted to each other - in which colleagueship, among others, can drive similarities on research perspectives (White et al., 2004). Social ties allowed understanding patterns of citations but are not sufficient because a researcher might know someone and not cite her or cite someone they do not know (White, 2001), in which case intellectual factors can be considered part of the disposition of information and knowledge. For White et al., there is no ‘clear temporal arrow in the matter: citing may or may not lead to meeting, and meeting may or may not lead to citing’ (2004: 112). As was previously suggested, White et al. considered that ‘intellectual ties and social ties cannot always be neatly separated’ (2004: 112), especially when roles are considered (e.g., co-authorship, department colleagues or mentor/students). Researchers are also considered part of many social circles that varies in terms of their level of acquaintances, ranging from strangers, contactable researchers, peers, members of the same invisible colleges, the same team, or co-authors (Milard, 2014). White prefer to use the notion of socio-cognitive networks (as in Merton [2000]) which make clear the *intercitation* structure – ‘the record of who has cited whom within a fixed set of authors' (White, 2011: 275) - that can conflate both dimensions. Combining citation with other social relationships might allow overcoming the 'structural confusion' (Holland & Leinhardt, 1974).

The exploration of citations is still a puzzling endeavour, challenging to understand theoretically and requiring further scrutiny. This thesis's first article used a methodological perspective to disentangle the main citation components when used as a *direct citation*, *bibliographic coupling* or *co-citation*. The combination of direct citation, bibliographic coupling and co-citation is a strategy that is gaining popularity (e.g., Small, 1997; Persson, 2010; Wang et al., 2019). The assumption to explore their differences is whether the usages of citation itself can derive in a *multigraph* in which the 'true' underlying structure can be discovered (Holland and Leinhardt, 1974). The conjecture is that if the citation can be interpreted as a cognitive and a social element (Crane, 1972; Small, 1977), then many different graphs can be used to describe these relationships (Chubin, 1976) as a socio-cognitive network (Merton, 2000; White et al., 2004).

1.2.2 Processes and Peer Recognition in the Evolution of Science

Building on the social network perspective that considers that citation is related to other social components, this thesis moves one step forward to explore different relationships' interdependency and the patterns involved to understand citation co-evolution. This section explores one of these patterns, known as the *Matthew effect*, as the relevance of peer recognition processes that enhance the visibility of researchers. From a network perspective, the analytical perspective to understand the *Matthew effect* (Lazega & Jourda, 2016) was developed by Price (1976) - further explored by Barabási & Albert (1999) - and was investigated in fixed groups (White & Breiger, 1975; Breiger, 1976; Mullins et al., 1977).

R.K. Merton and H. Zuckerman conceptualise the peer recognition tendency. They use as a case study the Nobel laureates in science and other recognised scientists to explore the patterns and publication practices of the topmost *elite scientists* (Zuckerman, 1967; Merton, 1968a). Merton defines The Matthew effect as 'the accruing of large increments of peer recognition to scientists of great repute for particular contributions in contrast to the minimizing or withholding of such recognition for scientists who have not yet made their mark' (1988: 609). To avoid the misunderstanding of the effect in terms of the relevance of a single event, Merton emphasises that this effect is 'Conceived of as a *locally ongoing process* and not as a single event, the practice of giving unto everyone that hath much while taking from everyone that hath little will lead to the rich getting forever richer while the poor become poorer' (1988: 610, emphasise is mine). The Matthew effect, according to Merton (1968), violates the universalism norm of science - the propensity of a truth-claims subject to preestablished impersonal criteria.

In the study of Zuckerman, she identifies that Nobel laurates tend to collaborate more often with other researchers that are distinguishable and highly productive, and 'They are also in a position, even before receiving the prize, to exercise *noblesse oblige*, the generosity expected of those occupying undisputed high rank, by granting authorship to junior collaborators' (1967: 396). The mentor-mentee generosity of consolidated research in helping less advantaged researchers positively affect producing more collaboration and working as a team. And, working as a team and having more collaboration allow more success in science (Wang & Barabási, 2021: 88).

Price discovered different skewed distributions using citations (Price, 1965)[4] as tendencies similar to the *Matthew effect*. To explain the shape of the distribution, Price (1963) used the concept of *invisible colleges* hypothesised as the increasing formation of groups gathered in institutions and journals, that day-to-day share communication through publications, and that overlaps in other groups defining an invisible college emphasising on the concrete interactions. In his perspective,

> 'For each group there exists a sort of commuting circuit of institutions, research centers, and summer schools giving them an opportunity to meet piecemeal, so that over an interval of a few years everybody who is anybody has worked with everybody else in the same category. […] Such groups constitute an invisible college, in the same sense as those first unofficial pioneers who

[4] Previously identified and rediscovered in different contexts (e.g., Lotka, 1926; Simon, 1955).

> later banded together to found the Royal Society in 1660. In exactly the same way, they give each [researcher] status in the form of approbation from his peer, they confer prestige, and, above all, they effectively solve a communication crisis by reducing a large group to a small select one of the maximum size that can be handled by interpersonal relationships.' (1963: 85)

Price considered that the evolution of knowledge was done by an *elite of researchers*[5] gathered into groups creating interpersonal relationships. Afterwards, the group create relationships with other groups through formal or informal communication using, among others, papers, manuscripts, letters (today e-mail) or meeting trough conferences, seminars or other academic activities. The group and the communication with other researchers create an *invisible college*, in which 'the apex of the triangle is not a single beloved individual but an invisible college; its locale is not a dusty attic of a teaching laboratory but a mobile commuting circle of rather expensive institutions' (1963: 90). Using the network of bibliographic references afterwards, he indicates the nature of the *scientific research front* that is built on recent work, questioning 'whether there is a probability that the more a paper is cited the more likely it is to be cited thereafter' (Price 1965: 512). He further suggests that 'one of the major tasks of statistical analysis is to determine the mechanism that enables science to cumulate so much faster than nonscience that it produces a literature crisis' (Price, 1965: 512).

Merton, Zuckerman, and Price rely on elites to explain the effect that they were observing. These elites are often considered as a scientific research group that certifies knowledge, and the manifestation of this is indicated in different dimensions, such as achieving accumulative advantages because of their belongingness to major research centres, having more rewards (e.g., prizes or citations), funding (Merton, 1988), or the control of committees, allocate research funds, informal influence (Mulkay, 1976). More recently, the skewed distribution of the *preferential attachment* has been revisited in the context of scientific networks

[5] For Price, different scientific roles, such as the scientific elite, leaders of groups, or lower-level scientists, were considered part of informal communication networks of scholars shaped by an elite of researchers with different affiliations.

(Fortunato et al., 2018) in which, among others, the 'hub actors' (as a type of *elite*) accumulate connections on time, leading to a 'richer-get-richer' situation (Barabási & Albert, 1999; Newman, 2001a; Clauset et al., 2009)[6].

According to the preferential attachment, the main explanation of the skewed distribution is that scientists cannot read all the papers published. Therefore, it is often the case that discovering papers leads to the bias that papers that are already cited tend to be encounter more often in the readings and therefore are more cited emulating the *Matthew effect* (Wang & Barabási, 2021: 184-185). The relation between the preferential attachment and the *Matthew effect* is that the recognition of the papers is often given to the most prestigious authors but not all co-authors. Price (1976) formalises this pattern in which the overall growth of the scientific literature contributes to the pool of papers available. The *preferential attachment* mechanism is then the tendency of researchers to select one of these papers that do not have a uniform distribution.

Further exploration considered other patterns that explain why citations increase as a process of accumulation by the information available in science. The first explanation was suggested in Price as *exponential growth* mechanism. He explained this pattern as the contemporaneousness and immediacy of science and the 'recognition that so large a proportion of everything scientific that has ever occurred is happening now, within living memory. To put it another way, using any reasonable definition of a scientists, we can say that 80 to 90 per cent of all the scientists that have ever lived are alive now' (Price, 1963: 1). Another common citation pattern is the *first-move effect*, which considers that the 'first papers' to appear in the literature tend to accumulate more citations regardless of the content (Newman, 2009). And another common explanation is the *fitness* effect mechanism, in which two equal cited papers will attract more citations if one of them is considered as higher quality (Bianconi & Barabási, 2001). In the three scenarios, the explanation for accumulation does not consider the specific context in which the citation is done nor the relevance of interpersonal relationships as

[6] This empirical tendency has received many controversies in the use of the theory that underlines the mechanism (Scott, 2011; Bonacich, 2004), the methodology involved (Handcock & Jones, 2003; Broido & Clauset, 2019), and more recently considered as a sub-cultural clash between nomothetic and ideographic perspectives in the understanding of this pattern (Jacomy, 2020 for a review of different controversies).

in *intercitation* contexts. Similar to Crane's observation, 'When individuals in a system *are not in communication with one another*, the probability that a member of the system will adopt an innovation remains constant and the pattern of growth is linear' (1972: 23, emphasise is mine).

The peer effect can be associated with open-ended fields (White, 2011) exploring collaboration as a *social dimension*. The *preferential attachment* mechanism was identified in *co-authorship* in the disciplines of biology, physics and mathematics (Newman, 2004). And it was interpreted as the tendency of research that collaborates more in the past in having more co-authors in the future and, therefore, authors with more collaborators will increase their social circles of collaborators creating hubs in science (Newman, 2004). Wagner & Leydesdorff (2005) considered that the *preferential attachment* mechanism of collaboration could be explained by dividing the tendency into three stages. The first ('the hook') stage is associated with the arrival of *newcomers* or *transients* into the field from other disciplines. Then, the *continuants* mediate the entrance of juniors, which is in the middle (sometimes acting as 'gatekeepers'). Finally, the tail is associated with the *continuants* and *terminants* (the 'hubs'), which are at the end of their career and an indicator of science institution. They further theorise that 'when choosing between two possible links, they will seek someone who is already highly connected and therefore has access to resources and reputation' (Wagner & Leydesdorff, 2005: 1615), emphasising the decision of individuals and the relevance of more prestigious researchers.

A different approach uses social networks to identify social boundaries to explore the *Matthew effect* considering an elite of researchers or organisations. Some empirical research identify an elite group of biomedical research in which there is a group of 'leaders' that the rest of the researchers are more 'aware' and that are more 'visible' in comparison with the 'followers' that might allow the 'leaders' to have access to new ideas, techniques, colleagues and students (Breiger, 1976, expanding from White & Breiger, 1975). In the case of elite universities, for example, in a longitudinal study of Stanford University faculty members (Rawling & McFarland, 2011), it was identified that collaboration and colleagueship secure more grant submissions and are more likely to achieve grants when interpersonal relationships are considered. Alternatively, multi-university collaboration is said to produce the highest-impact papers, but there is stratification in science among fewer higher rank universities that collaborate with other universities in similar positions (Jonas et al., 2008). Lazega & Jourda (2016), through the analysis of an elite of oncologists in France, identifies how members of organisations can borrow relational capital from other colleagues through indirect contacts and affiliation ties. This social bounded context allowed exploring other processes and relational structures to understand the evolution of science.

1.2.3 Processes of Group Formation in the Evolution of Science

In this section, the discussion is narrowed to the *group formation processes. In forming groups,* different relations are created in more specific contexts and emphasising different elements involved in the evolution of science, such as the situations, potentiality to create information or communicate, among others. This theoretical review connects with a classic discussion in the sociological study of science and knowledge, in which an active generation of researchers was trying to understand why different relational structures emerge from their local context. This section argues that this delimitation is highly related to the contemporaneous understanding of processes and structures from a network perspective, consonant with current methodological advances in social network modelling.

From a theoretical perspective, to identify why and how knowledge grows as *processes* in science, Diane Crane (1972) uses the concept of *invisible colleges,* examining the cognitive culture of the scientific communities with more details. From her approach, she was concerned about the *cumulative advantages* analysed by Price (1963) – called *preferential attachment* more recently (Barabási & Albert, 1999) - and theorised empirically by the *Matthew effect* (Merton 1968) as was reviewed in the previous section. Even when the regularity was acknowledged through a distribution, Crane considered that it was not sufficient to explain why the scientific knowledge growth and takes the form of this highly skewed distribution, neither an explanation of how the social relationships were affecting the production of ideas. From a network perspective, it can be related to identifying additional elements in the formation of *groups* – and entities with different levels - within scientific communities that can affect knowledge growth.

From this regard, Chubin & Studer considered local context, 'one must remember that institutional structures out of which the ideas arise may actually be 'distorting' our perceptions of scientific development' (1979: 186, *emphasis is mine*).

From the perspective of the sociological study of science and knowledge, different models were explored to understand the growth and internal dynamics of the scientific knowledge considering social networks (e.g., Crane, 1972; Mullins, 1972, 1973; Mulkay et al., 1975; Chubin, 1976). This exploration motivated the conceptual delimitation of invisible scientific *colleges* (Price, 1963; Crane, 1972; Lievrouw et al., 1987; Zuccala, 2006), *paradigms* (Kuhn, 2012), *scientific communities* (Hagstrom, 1965), *research networks* (Mulkay et al., 1975), *scientific collectivities* (Woolgar, 1976), *scientific specialities* or *fields* (Chubin, 1976, 1985), among others (Hagstrom, 1976; Shubin & Mullins, 1988; Morris & der Veer, 2009; Raimbault & Joly, 2021) that were further analysed through networks. As can be noticed, the terminology has been perplexing, and some of the particularities that these different versions shared are that they tend to consider emergent *relational structures* in the formation of scientific groups that arise from local interactions in social settings evolving from groups and then amorphous *networks* as the foundation of the growth of scientific knowledge.

As a case in point, in the dynamic model of Mullins (1972, 1973), he identifies different stages in which a *scientific speciality* grows in a (new) discipline starting from the formation of *groups*. The explanations behind the evolving process were the universities' context in the United States and the country's conditions allowing to have such particularities (Mullins, 1983). From this perspective, the first stage was called the *normal stage*, in which researchers were creating their innovations but were isolated from the usual activities guided by dominant ideas (the *paradigm*). The *network stage* was the second phase in which different network layers appear, connecting the researchers with more people, specifically those connected with the research and training centre in which the original ideas arise. The third stage is called *the cluster stage*, in which other centres hire students that were part of the original movements. Finally, the *speciality stage* in which the ideas become part of the everyday activities of science and then established as part of state of the art (e.g., works become routine, textbook appears, ideas are developed outside the centres of origins). The last stage becomes the background in which new groups may develop, and all the stages tend to overlap in regular activities of science.

A different model to explain the evolution of specialities was suggested by Chubin (1976). For him, the flow of ideas starts in *workgroups* characterised for their propinquity in time t (e.g., feedback from colleagues in the same laboratory or department). Then, moves in time $t+1$ into *clusters* of researchers regularly meeting to discuss new ideas and findings. Finally, it moves in time $t+2$, creating an amorphous network structure that links clusters and their parts (representing the entire specialities). In time $t+3$, it is assumed the relevance of *marginals*, or 'outsiders', and time $t+n$ corresponds to the expansion of the originated knowledge to non-specialists[7]. From this perspective, a paradigm is said to govern first a group of practitioners whose community structures can be unravelled, considering the changes over time from the small scale compared to cumulative tendencies.

The model of Chubin was similar to Mullins but lessening the number of early stages while incorporating more explicitly the relevance of 'marginals' and non-specialists in later stages. The model of Mullins is used as a baseline to explore with more detail further developments. The following sections argue that Mullins's model facilitates tracing a '*backbone'* between the theoretical discussions from the early sociology of science and knowledge related to social networks and current modelling strategies (e.g., Snijders, 2016) to study scientific networks. Mullins (1972, 1973) perspective incorporates some elements that many of the forthcoming models ignored: the relevance of local configurations in the evolutionary process and the interdependency of *multigraphs* that aimed to resolve the 'structural confusion' as emphasised by Chubin (1976).

1.2.4 Micro-mechanisms in Scientific Networks

[7] From this perspective, 'marginal innovators' can be highly influential (Granovetter, 1973), leading to the question to what extent the core and the periphery are interrelated, with the potentiality of creating the *birth of new specialities* (Crane, 1969; Chubin, 1976). And, when the 'marginals' increase their number, they can challenge the scientist's speciality to reject the prevailing paradigm (Kuhn, 2012).

The model of Mullins (1972, 1973) assumed dyads[8] as a minimum structure named 'paradigmatic group'. In Figure 1, Mullins present the 'paradigmatic group', in which he shows two types of relationships. The first relationship between actor A and B is the colleagueship of researchers in the same place, and the other arrows are the informal communication with external actors. He presents from this structure two actors that have no necessarily social connections (actor C and D) but are having communications with others. He emphasises that 'The minimal requirement of such an entity is two or more *established scientists* who have shifted from one viewpoint to another (Gestalt shift), and who might [i.e., actor A and B] or might not be [i.e., actor C and D] in communication with one another' (Mullins, 1972: 54-55, emphasise is mine[9]). Mullins stressed that 'specific social development precedes the literature' (1973: 246), and social relationships are different from the actual content of the work in which researchers might be engaged (Lievrouw et al., 1987). This *relational structure* assumed that an idea could occur independently by several people (where discoveries are after claimed in the race of publications) (Merton, 1973: chapter 14), and these researchers move to a similar cognitive situation according to similar problems. The social dimension follows, in which these researchers began to talk with available others to discuss the 'puzzles' and involve others in studying these problems. For Mulkay et al. (1975), the publication of a paper (or the communication of results in different formats) is crucial as a statement for further research in the earliest stage, expecting reactions from others. The model of Chubin emphasises the *group*, mentioning that 'The smallest unit and the one in which ideas are first broached is the work group. Relations here are based on propinquity – that is, communication among colleagues in the same laboratory or department. Perhaps feedback from colleagues will generate a new draft of the paper' (1976: 456).

[8] Attributed initially to Simmel (1950: part 2, chapter 3), dyads' is often related to the distinction between three possible stages between two actors. The first possibility is not having a tie between two actors. The second is that only one of the actors create a tie (asymmetric tie), and the third stage is two actors sharing a mutual tie (Wasserman & Faust, 1994).

[9] The emphasis on established scientists was considered one of the particularities of this perspective, allowing a further understanding of these groups' development for science policy and their career patterns (Chubin, 1985). Crane (1969) select an elite identifying common attributes, while others, such as Crawford (1971), use nominations. These actors can either be distinguished 'endogenously' using information from the network criteria or 'exogenous' variables to identify them.

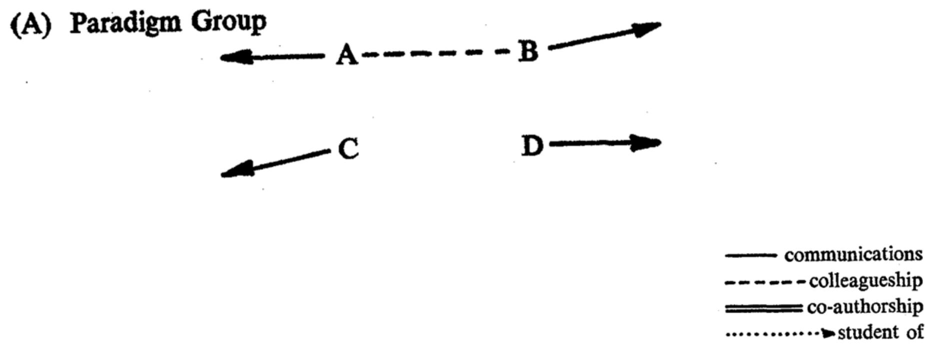


Figure 1 The *paradigm group stage* in the model of Mullins (1972)

The *groups* are considered in these perspectives as the baseline for the argumentation of Mullins (1972). He considers that groups depend on cognitive, social (interpersonal) and situational dimensions associated with departments or laboratories, which, together, can be operationalised as a *multigraph* because he incorporates in the representation different types of relationships (i.e., communications, colleagueship, co-authorship, and studentship) as co-constitutive of the same process. For Kuhn, '[a] paradigm is what the members of a scientific community share, and, conversely, a scientific community consist of [researchers] who share a paradigm' (2012: 175), which is social and cognitive at the same time. The *paradigm groups* in 'normal science' does not necessarily evoke a new paradigm but have the potentiality to become one. The *paradigmatic groups* might share an interest in solving a new problem (or a problem that they did not know before how to resolve), can communicate with others that share common or complementary perspectives, in which there a previous construction of trust ties, share speciality or complementarity between the researchers[10]. For example, Rawlings et al. emphasised how cohesion overlap intellectual and collaboration, in which 'teams that are more intellectually diverse have greater potential for interpersonal influence' (2015: 1691).

[10] The existence of many fields, growing and declining, and the links to some of the concepts, allowed the co-existence of many areas of research and diffusion of innovation, in which controversies allowed creating more differentiation between groups often related with the solution of significant problems and the appearance of anomalies (Crane, 1972: 37-39).

The *groups* are the minimum stable relational unit of analysis that allows identifying and studying the emergence of the local subnetwork or *relational structures* for the developments and evolution of knowledge, the internal dynamics inside disciplines, and the formation of potential (new) disciplines[11]. The social dimension of *groups* is often based on 'a number of people who interact with one another in accord with *established patterns*.' (Merton, 1968b: 339, emphasise is mine), giving to its frequency of interaction, in which people defined themselves as 'members', and is defined by others as belonging to the group[12].

Institutional affiliations often shaped group because researchers participate in daily situations in which actors share informal communication. The connection between researchers that share *foci* in which joint activities are organised involve informal communication (Feld, 1981). Scientific teams may share the same institutional affiliations, but the spatial proximity encourages informal communication (Katz & Martin, 1997) (e.g., laboratories, astronomical observatories, research centres and university departments) in which the colleagueship ties are promoted (Mullins, 1972). Actors share the same space of relation in which they incorporate the cognitive dimension of the organisational forms, where they share the same reference and knowledge space and institutional proximity that constrain their environment (Boschma, 2005).

Another subsequent *relational structure* in Mullin's model (1972) is called the *network stage*. As is presented in Figure 2, actor A and C are co-authoring a work. Actors A and B are colleagues in the same department or laboratory, and authors A, B, and C have informal communication, while actor C is starting external communications. According to Mullins, he considers that ties are stabilised in patterns of at least one co-authorship or different types of triads[13]

[11] Disciplines tend to create boundaries. Scientists are often organised in networks across different disciplines with more or less overlapping, for example, as Bellotti et al. (2016b) reviewed in exploring the Italian academia.

[12] Stable relationships are then different from *events* or *contingent interactions* that are more situational, not stable, but can become a pattern. The *continuum* between *events* and *states* have a long tradition in the network perspective (e.g., Homans, 1950; Boissevain, 1968; Borgatti & Halgin, 2011b; Crossley, 2011). The complexity of other interactions and the identification of relationships that can be present in the context of science was addressed from ethnographic studies in the context of studies about science and knowledge (e.g., Collins, 1974; Knorr-Cetina, 1982) and, more generally, in the tradition of the network anthropologists (e.g., Mitchell, 1969).

[13] The theory about triads is associated with Simmel (1950: part 2, chapter 4).

(similar to the labelling scheme assigned by Holland & Leinhardt, 1970; Davis & Leinhardt, 1972) of scientists engaging in informal communication, or colleagueship, over some time (Mullins, 1972: 58). The stable patterns of at least one co-authorship are consonant with the definition of *group* in Merton (1968b), while the previous usage of 'groups' does not necessarily consider this characteristic and is assumed simultaneously as a stage of potentiality and the presence of possible interactions in the form of informal communication. The particularity of this stage is that the stability of a tie could be simultaneously within the *group* (e.g., inter-department or project of research) or between groups (e.g., inter-organisations), involving different types of relationships such as informal communication, colleagueship, co-authorship and apprenticeship (as a *multigraph*)[14]. According to Mulkay et al. (1975), in their interpretation of the stages, a *second stage* could achieve an intellectual consensus, in which newcomers acts as an apprenticeship to an established research network or are led by a mature scientist.

One of the particularities of Mullins (1972) *network stage* is that at the social level, there are member exchanges in different institutional arrangements, and recruitment of researchers (often younger scientists), creating rapid growth and turnover. Chubin considers this *relational structure* as a cluster of researchers, 'who regularly exchange information and who may even assemble to discuss new ideas or findings' (1976: 456). The pattern of been embedded in collaboration networks have been studied in specific settings (e.g., Friedkin, 1978; Tuire & Erno, 2001; White et al., 2004; Rawlings et al., 2015; Stark et al., 2020) and using ethnographies (e.g., Knorr-Cetina, 1982; Collins, 1998), in which researchers 'share ideas, use similar techniques, and otherwise influence each other's work' (Moody, 2004: 213). In this stage, the entire network increases the number of ties and decreases the number of disconnected or independent researchers.

[14] According to Lievrouw et al. (1987), the 'informal communication' should be investigated through qualitative methods, while others (Schrum & Mullins, 1988; Zuccala, 2006) considers that the relationships can be approximated from the 'formal channel' of communications used through bibliometric information. From the perspective of this thesis, the second approach is used in which the institutional affiliations approximate the 'informal communication', while it is agreed that mixed methods (e.g., Lievrouw et al., 1987) are ideal for a better understanding of the processes.

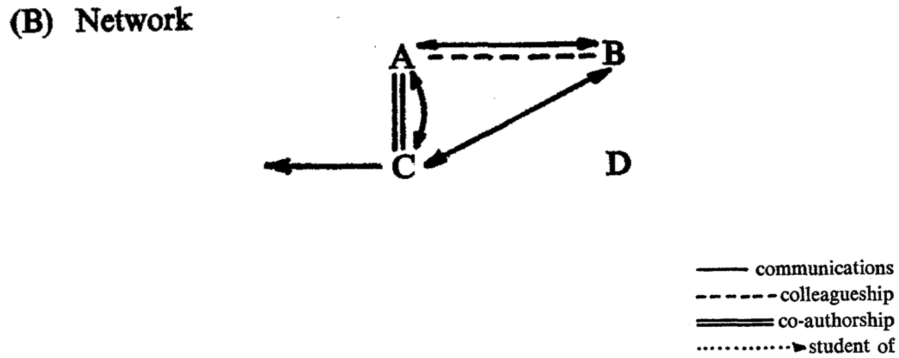


Figure 2 The *network stage* in the model of Mullins (1972)

The first two stages in the Mullins model and the proposal of Chubin allowed emphasising the relevance of local patterns in the subnetworks for the stability of more complex structures. In these models, the emphasis was on *group formation* instead of the *Matthew effect*, and both perspectives can complement each other if they are mutually contemplated as simultaneously operating local patterns. These patterns are often considered as mechanisms (Hedström & Bearman, 2009) that are semi general because they can explain certain types of phenomena. The mechanisms are somehow general (not restricted to a specific time, place, identity of actors, the content of beliefs, or type of actions) and have specific pattern linking different types of entities requiring stability in the social forms. Mechanisms assume a constellation of entities and their organised activities, creating regularities that bring a particular type of outcome. Compared with less stable and contingent interactions on different spatial and temporal frames that may produce different outcomes, the mechanisms require a certain level of stability of the relationships between entities in specific contexts (Merton, 1968b).

Similar to Mullins (1972, 1973) early model that identifies dyads and triads, recent models allowed distinguishing how and why different processes emerge from local processes using statistical network models (e.g., Block et al., 2019). Some of the common elements that share these analyses are that they assume that the global structures depend on the presence of simultaneously operating local structures or configurations (Robins et al., 2005) – expressed in statistical models such as the exponential random graph models (Lusher et al., 2012). Other models consider that the decisions of actors in the creation of social ties (Snijders, 2001) as the processes responsible for the formation of network structures - as in statistical perspectives such as the stochastic actor-oriented model (Snijders et al., 2010) or the dynamic network actor models (Stadtfeld & Block, 2017). Mullins (1972) model considered dyads and triads without exploring the specificity of the type of forms and their relevance to shaping the networks. In the sociological study of science and knowledge, *group formation* was considered an indispensable element. These patterns often represent local social regularities or micro-mechanisms that can explain the network and consider different entities involved in the analysis.

A comparatively new research area addressed the exploration of these micro-mechanisms and their relevance for shaping the networks. For example, to investigate homophily between departments, spatial proximity, topics, disciplines or considering social attributes such as gender or race (Kronegger et al., 2012; Cimenler et al., 2015; Harris et al., 2015; Peng, 2015; Dhand et al., 2016; Luke et al., 2016; Zinilli, 2016; Fagan et al., 2018; McLevey et al., 2018; Wang et al., 2018). Or to address the tendency of creating transitivity (as a type of triadic structure) within scientific networks (Kronegger et al., 2012; Zinilli, 2016; Sciabolazza et al., 2017; McLevey et al., 2018). And the tendency of popular actors to receive more ties in comparison with the other researchers within the network as it was suggested in the *Matthew effect* (Peng, 2015; Dhand et al., 2016; Harris et al., 2015; McLevey et al., 2018; Zhang et al., 2017). The organisations have been considered a relevant research area to understand collegiality (Wang et al., 2013; Gondal, 2018; Purwitasari et al., 2020; Stark et al., 2020). Simultaneously using different micro-mechanisms has facilitated the further consideration of the different dimensions underpinning the network structure, but less has been done to explore these micro-mechanisms using *multigraphs*.

From the different models reviewed until now - and using Mullins model as a reference -, cognitive, social (interpersonal), and situational (often in departments or laboratories) dimensions are considered building blocks for the emergence of different *relational structures*. Zuccala (2006) considered these three main components (i.e., *subject speciality*, *social actors*, and *information use environment*) for the evolution, decline, and formation of scientific networks. This thesis's second article addresses the co-evolution of some of these entities to understand why scientists create interpersonal *intercitations* when different social elements are simultaneously considered, such as the *Matthew effect* and the relevance of *groups* in the contexts of a scientific discipline using cross-level micro-mechanisms. The second article explores how the micro dimension represents the network from a methodological perspective, 'operational definitions of this relation [communications used to link scientists] to determine their relative goodness-of-fit as aggregate representations' (Chubin, 1976: 451) in which it is explored different goodness of fit for complex networks as a diagnostic often used in current statistical models for the study of social networks.

1.2.5 Ill-defined Structures in Scientific Fields

Moving beyond dyads or triads as *groups* require further consideration for the delimitation of boundaries concerning researchers and disciplines because it is no longer straightforward to identify an intuitive type of *relational structure* without considering potential ties of actors at longer distances (e.g., from other disciplines or indirect ties) and that often involve *marginal* actors.

This issue was addressed, among others, by Diana Crane, in which she considered a broader definition of *invisible colleges*, in comparison with Price, considering scientists with *a common interest* in amorphous research areas. For example, she mentioned how physicists indicate that their disciplines' subfields constantly shifted their boundaries considered 'fluid' (Crane, 1972: 13). Using Kadushin (1966) concept of 'social circle', Crane (1972: 13) identify that this concept is the best way to describe the social organisation of the members of research based on impersonal networks in an invisible college. She recognises that the boundaries of the social circle are challenging to define and mentioned that

> 'Indirect interaction, interaction mediated through intervening parties, is an important aspect of the social circle. It is not necessary to know a particular member of a social circle in order to be influenced by him. Not only can a scientist be influenced by publications written by authors whom he has never met, but he can also receive information second-

hand through conversation or correspondence with third parties' (Crane 1972: 13-14)

She aimed to identify the interaction between the cognitive and socials component of science by exploring groups in communities that share common interests (as is suggested by Kadushin, 1966). Kadushin (and Crane) starts from the shared interests and then moves to locate the people involved (Bott, 1968: 315). In a similar perspective, Mulkay et al., following Barnes (1954), considered that 'a relatively intensive concentration of interest ties as a *research network* even though, because some and perhaps many ties are not reciprocal, the network will not necessarily have a natural boundary.' (1975: 189, emphasise is mine). Schrum and Mullins (1988) considered co-occurrence on bibliographic or references to identify the common *interests* from a methodological perspective.

Mullins recognise these natural boundaries or amorphous structures and named them part of his model's third stage as *clusters*[15] or *invisible colleges* (Hagstrom, 1976). As is presented in Figure 3, the model of Mullins (1972) becomes more complex in which co-authorships increase (e.g., actor B, E and H, or C and D), there are students involved (e.g., actor B and C, or D and F), colleagueship (e.g., A with B, or E with G) and informal communication (e.g., A with C, or D with F), which together present a growing *multigraph*. The particularities of these structures were that 'A cluster forms when scientists become self-conscious about their patterns of communication and begin to set boundaries around those who are working on their common problem' (Mullins, 1972: 69), recognised by those who are inside or outside the cluster and considered as more stable than the dyadic and triadic structures which constitute them having their own culture (e.g., own history, set of beliefs, theories). Hagstrom considered that the *clusters* 'vary considerably in size, interconnectedness, internal stratification, clarity of boundaries, and visibility to members and non-members' (1976: 758). After comparing with alternative delimitations, Hagstrom (1976) prefers to call the cluster an *invisible college* to clarify terminology, and that was

[15] The *cluster* of Mullins (1972) is further considered as the equivalent of *invisible college* (Lievrouw, 1990), and more recently, *cohesive subgroup* (Everett & Borgatti, 2019).

further re-defined by Lievrouw as 'a set of informal communication relations among scholars or researchers who share a common interest or goal' (1990: 66), which does not imply formal institutional structure (Zuccala, 2006). Concordant with Mullins (1972), the *clusters* were assumed to have not achieved yet formal structure, and the growth rate was less rapid than the *network phase*. According to Mullins's model, this structure could not maintain itself beyond co-authorship and the informal communications among the researchers involved if they change. They gather into summer schools meeting (as a more contingent situation or focus of activity), but there was no formal institutional society.

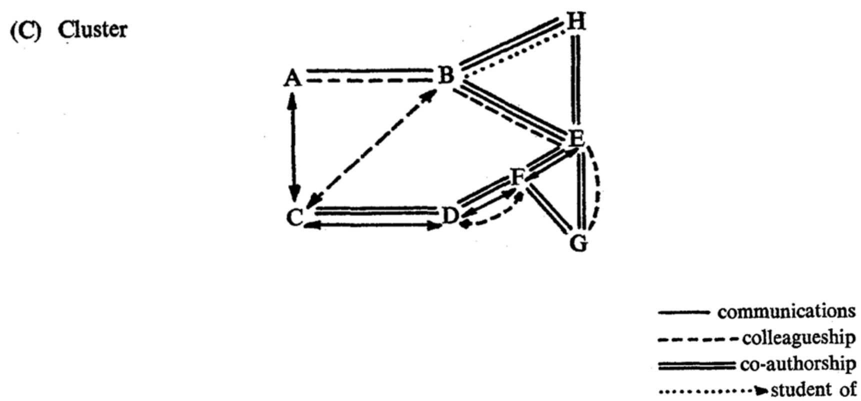


Figure 3 The *cluster stage* in the model of Mullins (1972)

One of the most critical elements for the *clusters* (in Mullins terminology) is the perception of belongingness into the *invisible college* (Hagstrom, 1965) that generate conceptual and methodological divisions. A further argument suggested that neither membership, interactions, nor formal membership is sufficient to explain membership to the communities (Gläser, 2001). The perception argument contrasts with the social dimension, dividing the intellectual and social dimension distinguished before, in which bibliometricians tend to use the role of third parties in citations to reinforce this argument (e.g., co-words, bibliographic coupling or co-citation) (Morris & der Veer, 2009), disregarding the social dimension. One of the implications is that the perception from third parties allowed clarifying the *invisibility* of the colleges. Instead, the socio-cognitive network considers the overlapping between intellectual and social dimensions (Merton, 2000; White et al., 2004). For example, this perspective is concordant with Zuccala's (2006) definition of *invisible colleges* that combine the shared interest *or* interactions among subject speciality of research engaged in formal or informal communication, or an overlapping model introduced by Rawlings et al. (2015) or Stark et al. (2020) in which the flow of knowledge depends on both dimensions.

The delimitation of boundaries involved a frosty debate about the definition of these ill-defined boundaries. In which the discussion was often raised according to methodological preferences and stressing the relevance of scientific practices. Some researchers highlighted the relevance of creating boundaries according to bibliometric studies to consider broader researchers beyond the scientific domain (Small & Griffith, 1974; Gläser, 2001; Raimbault & Joly, 2021). Other researchers combined the bibliometric approach with questionaries (White & Breiger, 1975; Breiger, 1976; Mullins et al., 1977; Bourdieu, 1988). From another perspective, researchers emphasised the relevance of ethnographic studies and the scientific practices (e.g., Collins, 1974, 1981; Knorr-Cetina, 1982; Wenger, 1998) to explore the difficulties of creating scientific boundaries. The exploration of specialities (Chubin, 1976) or research networks (Mulkay & Edge, 1973; Mulkay et al., 1975) as a unit of analysis were considered as coherence structures (Whitley, 1983) which in practice were difficult to discern. For example, Collins (1974: 177-178) suggested in his ethnographical study about a gas laser (known as TEA) that a way to shape a network could start from the contact to a laboratory, in his case, the Canadian defence research laboratory as an ego-network (the *core-set*), and then trace the other actors involved in the diffusion of knowledge[16]. Woolgar (1976) argued that identifying *collectivities* becomes challenging when memberships are doubtful, in which actors that are not in the 'core' are considered part of the periphery limiting their relevance[17]. Collins (1974) and Woolgar (1976) perspectives emphases on the relevance of broader environments.

[16] In the following years, Collins (1981, 1988) will identify the *core-set* through controversies and the social contingencies of those involved in experimentations and observations, but in which it is not possible to know who is inside or outside the core-set.

[17] Notice the parallel between these alternatives and the quasi-groups (Meyer, 1966: 115-116) or *stars* and *zones relationships* described by Barnes (1969: 60-61) - then described as *personal network* (Boissevain, 1974: 26-27) -, which have been recently (re) used to trace 'hard-to-reach' or 'hidden populations' (Gile & Handcock, 2010) or to extract partial network in extensive networks (Stivala et al., 2016).

The delimitation of these ill-defined structures is still an ongoing debate in the current network perspectives[18]. A popular terminology, and used more recently to identify these *invisible colleges*, is 'community detection', in which actors in a 'community' would be more closely connected than actors in other subgroups. Traditionally, these have been based on some forms that distinguish between 'inside' and 'external' subgroups, roles, or common properties (Fortunato, 2010; Fortunato & Hric, 2016). The aim of 'community detection' is to identify the partitions using the graph's information (Fortunato, 2010). When a highly connected community is considered, then is assumed to be a *cohesive subgroup* - which is a more general concept -, leading that the 'community detection' would be a particular case of 'blocks'[19] (Everett & Borgatti, 2019), used for the understanding of scientific *specialities* (e.g., White & Breiger, 1975; Breiger, 1976; Mullins et al., 1977; Burt & Doreian, 1982).

The fourth and last stage of the network evolutionary model of Mullins (1972) was the *speciality* stage (Figure 4) that adopts the form of a formal organisation (e.g., recruitment procedures, a test of membership, journals, meetings, other locations that support its work, among other). A *speciality* is an 'institutionalized cluster which has developed regular processes for training and recruitment into roles which are institutionally defined as belonging to that specialty' (Mullins, 1972: 74). Speciality is the last stage of the Mullins model, and for the case of the 'phage workers', they were then absorbed into molecular biology and further becoming part of the normal activity of science. As the last stage of the process, its resolution becomes part of the curriculum of teaching programmes, with validated rules and processes for technical solutions, increased graduate students, and attendance in meetings, among others (Mullins, 1973). The students themselves colonise organisations, and this speciality is institutionalised into journals, positions and centres (Shubin & Mullins, 1988). Then

[18] In scientometrics, the exploration of *speciality* is still an ongoing agenda identifying meso-level structures from the macro perspective using techniques such as co-words, co-citation or bibliographic coupling (Gläser, 2001), and as is have been considered from the science of science (Fortunato et al., 2018; Wang & Bárabasi, 2021). The debate is concordant with the current understanding of network delimitations and the specification problem in network analysis (Laumann et al., 1983).

[19] From a sociological perspective, 'blocks' has been treated as positions that represents role sets (e.g., White et al., 1976) or as social niches (e.g., Lazega, 2001)

institutionalisation is the crucial element to achieve the status of a speciality, and the 'research areas which have become established take a long time to die out altogether' (Mulkay et al., 1975).

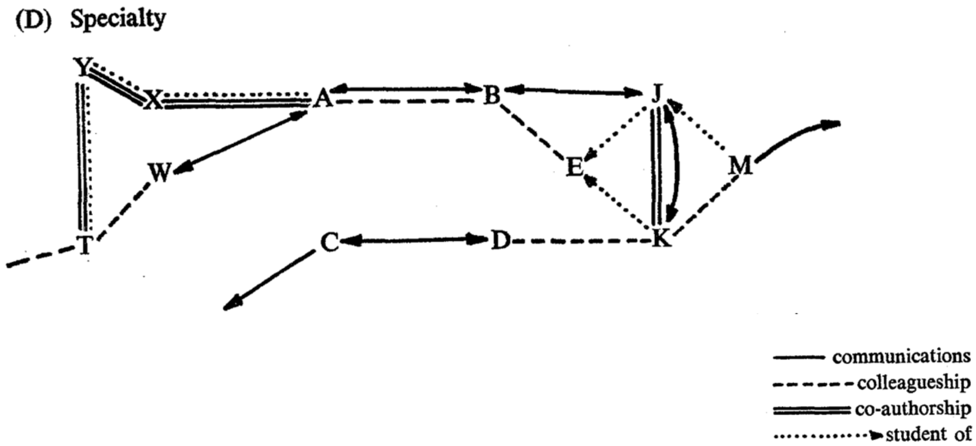


Figure 4 The *specialty stage* in the model of Mullins (1972)

The final stage of Mullins (1972, 1973) model assumed that after achieving the *speciality,* it becomes part of the standard canon of scientific activities in a circuit of different universities. The particularity of becoming part of the establishment of the universities are, on the one hand, that the dissemination of specialities into organisations generates that *speciality* becomes more challenging to observe in isolation in which researchers might be part of other *paradigms groups* and *networks* (in Mullins's terminology). The specialities have many overlaps between specialities in the work of researchers, articles, scientific instruments and methods, and journals (Gläser, 2001). And, because they are part of the curriculum of universities and expected knowledge from graduate students, organisations are the guarantors of achieving these expected minimums.

Instead of using the *speciality*, further developments have suggested considering organisations – primarily because of the relevance of disciplinary departments (Jacobs & Frickel, 2009) - to identify some social boundaries. They were assuming that the interrelationships between the researchers and their institutional affiliation in laboratories, departments, among others, are co-constitutive and assumed to have a 'dual position' (Breiger, 1974; Lazega et al., 2008; Bellotti et al., 2016a; Lazega & Jourdá, 2016) because researchers are often 'nested' into these places.

Some of the perspectives that use organisational fields consider that these boundaries only exist when organisations are institutionally defined, in which

> '[t]he process of institutional definition, or 'structuration', consist on four part: an increase in the extent of interaction among organizations in the field; the emergence of sharply defined interorganizational structures of domination and patterns of coalition; an increase in the information load with which organizations in a field must contend; and the development of a mutual awareness among participants in a set of organizations that they are involved in a common enterprise' (DiMaggio & Powell, 1983: 148)

In this context, an organisational field in the aggregate constitutes an area of institutional life in which organisations share key suppliers, resources and product consumers, regulatory agencies, and other organisations that produce similar services or products. From some recent network perspectives, the consideration of organisational fields is appealing to identify exchange through networks (e.g., Galaskiewicz & Wasserman, 1989; Stadtfeld et al., 2016) and understand scientific networks (e.g., Powell et al., 2005; Lazega et al., 2008).

Some organisations share interests that tend to be structured into a similar field, in which there are similar forces that made them more similar to one another in time. This particularity has been conceptualised as a homogenisation process through isomorphism, considered as 'a constraining process that forces one unit in a population to resemble other units that face the same set of environmental conditions' (DiMaggio & Powell, 1983: 149). The isomorphism can be by competition (with other groups) or institutional (at the disciplinary level). For the first case, the organisations' homogenisation is the rationality that emphasises the market competition, niche changes, and fitness measure. Institutional isomorphism competes for resources and customers and political power and legitimacy, either social or economic fitness.

The consideration of *cohesive subgroups* and *organisational fields* allowed creating an ill-defined boundary for the study of scientific networks. In the contemporary multilevel analysis of networks (Snijders, 2016), different alternatives have been made to analyse the differences within or between *cohesive subgroups* in the study of scientific networks. Some studies compare different disciplines or similar substantive research considering different *cohesive subgroups* (e.g., Kronegger et al., 2012; Ferligoj et al., 2015; Sciabolazza et al., 2017; Akbaritabar et al., 2020). Other studies have investigated the analysis of inter-organisations (e.g., Powell et al., 2005; Lazega et al., 2008), in which the relevance of the organisations is highlighted. These investigations explored common micro-mechanisms or constraints that can be present in relatively broader and defined populations.

In this section, two aspects were raised: The first is the difficulties in generating boundaries among researchers in the tradition of the sociological study of science and knowledge. The second was the consideration of organisational fields as a perspective that allowed exploring similar constraints among organisations. The third article of this thesis addresses the identification of common mechanisms between a sample of organisations to identify whether the patterns are similar or vary according to the organisations' position in the field. The positions of the organisations can be either in the core or the periphery of a scientific field. From a methodological perspective, the contribution explores an alternative operationalisation of a *cohesive subgroup* that relies on two simple premises. First, assumed that actors are situated in the same organisations, which is reasonable to expect a minimum level of colleagueship because they are co-workers. Second, it incorporates researchers who share the same interests – publishing in the same discipline- as the members of these organisations or as an 'outsider' actor citing researchers from these institutions.

1.3. Astronomy and Astrophysics in Chile

For the empirical exploration of the local *relational structures* and the *processes* that allowed the evolution of scientific networks, this thesis analyses the case of the Chilean astronomy and astrophysics discipline. Some of the main reasons to explore this case are that the size of the researchers working in this area is relatively small (López et al., 2005; Gibert, 2011) (~200 professional researchers in 2017 according to the Chilean Society of Astronomy), and the access of the telescopes held in this country is restricted to the members of 'Chilean Institutions' included in a 'white list' in charge of SOCHIAS[20]. The development of the discipline in this country is relatively recent, in which the first astronomical program was created in 1965 (University of Chile), the second astronomical program appeared only in 1990 (Pontifical Catholic University of Chile), and in 2012 there were nine astronomical departments (CONICYT, 2012). According to SOCHIAS, in 2020, 21 organisations were investigated in areas related to astronomy and astrophysics in this country[21]. Currently, in the context of astronomy, 'Chile, local universities played a key role, including the development of endogenous capabilities for astronomical research.' (Guridi et al., 2020: 5). As can be noticed, this specific context relies on the discipline[22] and the organisations to administrate the access to time observation[23].

[20] Extracted from: https://sochias.cl/access-to-chilean-telescope-time/ (last time visited: 30/03/2021). Noticed that post-doctoral, visiting academics, and PhD student can apply for the time of observation if they are sponsored by academics hired in Chilean institutions. For example, to apply for ALMA time of observation – currently, the largest radio astronomical observatory held in Chile – 'Each proposal must have at least one permanent *Chilean faculty member* among the proposers (PI or co-I)' (extracted from: http://www.das.uchile.cl/~alma_crc/ [last time visited: 11/04/2021]). For CNTAC, there is a temporal restriction as well: 'For the purpose of applying for CNTAC time, a Chilean astronomer is defined as a resident scientist working in a Chilean institution. 'Resident Scientists' are those scientists who maintain continuous residence in Chile for at least 9 months.' (extracted from: http://www.das.uchile.cl/das_cntac_rules.html [last time visited: 11/04/2021]).

[21] Extracted from: https://sochias.cl/astronomia-en-chile/universidades/ (last time visited: 11/04/2021).

[22] Previous research has noticed the isolation of astronomy from other disciplines (Leydesdorff & Rafols, 2009; Jansen et al., 2010; Van Noorden, 2015), and in the Chilean case (Cárdenas et al., 2015).

[23] Considered one of the most relevant researchers' assets (McCray, 2000; Jansen et al., 2010), which according to the evaluation processes – as in the Atacama Large Millimeter/Submillimeter Array - is based on a collective review that is biased in favour of less risky proposals (Espinosa, 2015), disciplinary bounded and with stables paradigms (Heidler, 2011, 2017). Elites of researchers in astronomy – such as in the *astro-informatic* speciality – emphasise the relevance of the priority in the discoveries (Espinosa-Rada et al., 2019).

The local community has access and the monopoly of an impressive amount of time observation. The main reasons are because, in Chile, there is the Atacama Desert, which is the driest non-polar place in the world located and surrounded by two mountain chains giving unique geographical conditions for astronomical observations (Guridi et al., 2020), which the Office protects quality and transparency of the skies through the Protection of the Skies of Northern Chile (OPCC). This condition allowed this country to hold nearly 40 per cent of the world's earth-bounded telescopes' astronomical capacity and the most important astronomical observatories (list of current observatories in Appendix A). In the decade of 2020, it will be home to almost 73 per cent of the world's total astronomical infrastructure since the incorporation of new telescopes and 96 per cent of the southern hemisphere's scientific observation capacity (CONICYT, 2012; Unda-Sanzana, 2018 in Guridi et al., 2020). In 1997 (Decreto 1766), the government ensured that the Chilean astronomy community receives 10 per cent of the observing time of the international observatories built in the country and will own 10 per cent of the Large Synoptic Survey Telescope [LSST] computer cluster. The disposition of resources geographically based and the propinquity in a country allowed to identify the location-dependence and the emergence of ‘national core’ of researchers (i.e., a national subset of an international speciality) (Gläser, 2001: 201-203) that are shaped by the access to these technologies.

Previous research emphasised the problems of generating a local community in an underdeveloped country. This situation has been considered as a risk of ‘producing a quasi-colonial form of dependencies on foreign partners’ (Guridi et al., 2020: 2; also, in Espinosa-Rada et al., 2019), or conjectured that ‘Chile’s trajectory shows just how hard it is to build ‘scientific community’ and ‘research infrastructure’, both important to how scientists practice their work’ (Barandiaran, 2015: 143). According to Barandiaran, it is difficult to assume that Chile has a ‘national scientific community’ because there is no stable funding that can sustain the community, the funding available reinforces the quantitative productivity indicators, there is no sufficiently ‘dense communication network among scientists’, among others (2015: 146-147). According to Guridi et al. (2020), this situation changed after the fieldwork of Barandiaran in 2009, in which the disproportional benefits to foreign organisations were reverted despite initial missteps.

In terms of the 'national scientific community', members of this community seem to maintain fluid communication. For example, in SOCHIAS's newsletter[24], it is possible to track the history of the community, the arrival of new members, the creation of new departments, some relevant events, discoveries, among others. The astronomers working in the Chilean institutions have access to resources administrated by the local community. Different sources for local astronomical development are available that mix local and foreign funding (e.g., ESO–Government of Chile Joint Committee funded projects since 1998, ALMA-ANID since 2005, Gemini-ANID since 2005, QUIMAL since 2009, China-ANID since 2015, among others). The observational time is allocated by the Chilean Telescope Allocation Committee (CNTAC) and the National Commission for Science and Technology (CONICYT) through the APEX and Gemini SUR Committees bounding and promoting the internal developments of astronomy and astrophysics in this country.

Two periods are further analysed, which presents some differences in their context that might be related to the arrival of new technology and the government strategies to seek competitive advantages to leverage Chile depending on the development of the local scientific community. One of the periods corresponds to a few years after the Atacama Large Millimeter/submillimeter Array (ALMA) arrival in 2011 - which is currently the largest radio astronomical observatory - full operative in 2013. In this period, the government created the Atacama Astronomical Park in 2013 as a strategic zone for the exclusive usage of astronomical observatories in the Chajnantor Plateau despite being one of the richest zones for mineral extraction (the main economic activity of this country). A few years before, other telescope classes were settled in this country, incorporating some of the largest observatories in optical astronomy (e.g., the Magellan Telescope or the Very Large Telescope) (Guridi et al., 2020), representing a different branch of this discipline. Previous research identifies how the time observation is allocated in ALMA, and the results indicate that there was a clarity

[24] Extracted from: https://sochias.cl/material-de-interes/newsletters/ (last time visited: 11/04/2021)

upon the quality of the research projects for the allocation of the time observation that often corresponds to established researchers (Espinosa, 2015). The Chilean astronomical community was highly isolated from other disciplines (Cárdenas et al., 2015). During this period (approximated between 2013-2015), Chilean astronomers had access to some of the largest telescopes for optical and radio astronomy at the same time, and for a couple of years (in average 2 or 3 years), each astronomer, after having the allocation time for observation, had the exclusivity of using the information before it is openly available.

Some years later, a different scenario and context occurred in which a new class of observatories based on survey telescopes was starting to become part of the national discussion. In 2017, the Chilean government had an interest in the development of astronomy to spur economic activity to national advantage (Arancibia et al., 2020; Guridi et al., 2020) because of the arrival of the Vera C. Rubin Observatory (*a.k.a.*, the Large Synoptic Survey Telescope [LSST]), considered to be the research front of the discipline (Espinosa-Rada et al., 2019). This class of telescope represent the age of the digital era in astronomy (McCray, 2014; Hoeppe, 2014), which is a new variation of telescopes (e.g., the Sloan Digital Sky Surveys [SDSS] or the 2df Galaxy Redshift Survey [2df GRS]). Compared with previous telescopes (optical or radio), this type of observatory creates an automated mapping of the sky that will deliver thousands of *petabytes* of images and data integrated into virtual observatories such as the Virtual Observatories (VO) in which the information become available immediately. One possible implication of this is that the new 'data-turn' in astronomy can change some of the current and predominant astronomers' basic research practices. In which previous research mentioned that astronomers change their norms and behaviours considering that sharing, ownership, and access of data become more challenging once the data become available for the whole world – open access- (McCray, 2014) and where there is no need to write a proposal for time allocation (Heidler, 2011) as it is usual in current astronomy. In comparison with the previous period, at this moment, I conjecture that it was a period of change and some internal variation – corresponding with the second stage of the theoretical model of this research, but more open to their social environment. This period corresponds to the preparation of the national community for the arrival of the LSST.

1.4. The Present Studies

In this section, the three empirical articles that constitute the thesis's corpus are briefly presented, highlighting how they are connected and some of their contributions.

The three studies explore multigraphs to deal with what is called the 'structural confusion'. The confusion is investigated through a deeper understanding of the ties between different types of networks (i.e., using *direct citation*, *co-citation*, *bibliographic coupling*, and *collaboration*) and, with more emphasis in this thesis, collective actors in a different level (*i.e.*, institutional affiliations and journals in this case) using *multigraphs*. In each article, at least one type of relationship can be assumed as *social* and *cognitive*, considering that both relationships constitute a social-cognitive network operationalised differently but conflated in interpersonal or *intercitation* relationships contexts. The three studies aim to make a methodological contribution.

As was presented in this introductory chapter, the last two articles consider different *processes* and *structures* at a different level of abstraction, similar to the model of Mullins (1972, 1973) and the different exploration about the *processes of group formation* (e.g., Mulkay et al., 1975; Chubin, 1976; Woolgar, 1976), and the consideration of peer recognition according to the *Matthew effect* (Price, 1963; Zuckerman, 1967; Merton, 1968a). More precisely, the second article investigates the relevance of *local processes* analysing different relationships (i.e., citations and co-authorship). This study investigates the relevance of collective actors (i.e., journals and organisations) in the co-evolution of a discipline, in which the identification of certain micro-mechanisms, such as different types of *groups* or the *Matthew effect*, allowed exploring how together they can resemble the network of the Chilean astronomy and astrophysics discipline. The third article moves one step forward to explore the variation within and between *cohesive substructures*, considering the *local relational* processes and the relevance of micro-mechanisms in the network formation. Also, investigates with more flexibility 'outsiders' actors assumed to share common interests in the discipline of astronomy and astrophysics and their relationship with the 'core actors' to explore the variations in the potential isomorphism between organisations. From a methodological perspective, the second study uses an *analysis of multilevel network* perspective, and the third study a *multilevel analysis of network* (Snijders, 2016) that deals with the different understanding of 'levels' from a multilevel perspective.

Study 1. Citations are challenging to understand because they mix or conflate social, cognitive, and situational dimensions often when researchers are aware of each other (Chubin & Studer, 1979; Schrum & Mullins, 1988; White et al., 2004). In the first article (chapter 2), a methodological perspective is used to investigate whether the combination of different types of representations that use citations (i.e., *direct citation*, *bibliographic coupling,* and *co-citations*) can recover an underlying 'real structure' (Holland & Leinhardt, 1974) of this network. The combination of these different types of relationships have been suggested in recent scientometrics literature to have a more accurate representation of citations (Small, 1997; Persson, 2010; Wang et al., 2019) – called the *author normalised weighted direct citation*. The research question that motivates the first paper is,

> *What are some of the consequences of the combination of different citation-based networks into a common representation?*

The first article addressed some methodological considerations in combining derived citations networks (i.e., *direct citation*, *bibliographic coupling* and *co-citations*) between authors – which is called the *author normalised weighted direct citation*. To explore the *author normalised weighted direct citation*, are considered different similarity measures often used in scientometrics to study citation-based measures such as the *Jaccard index*, *cosine similarity*, and *association strength* (Ahlgren et al., 2003; van Eck & Waltman, 2009; Egghe & Leydesdorff, 2009). For the exploration, a quadratic assignment procedure (QAP) and multiple regression quadratic assignment procedure (MR-QAP) (Krackhardt, 1988; Dekker et al., 2007) are estimated. The (MR) QAP is used considering the different citation-based measures to understand the implication of the normalisation and how they give more relevance to different dimensions of the citation network when the components are combined. This contribution suggested that combining different normalisations requires distinguishing between an analytical interest in the shared cognitive dimension according to the community's perception or the communicational trace between the actors.

Study 2. For the analysis of peer recognition and group formation, contemporaneous models can analyse both types of processes simultaneously to understand the evolution of scientific networks (e.g., Ferligoj et al., 2015; Kronegger, 2012; Zinilli, 2016; Purwitasari et al., 2020). While recent studies have addressed the use of networks that consider two types of 'levels' (i.e., organisations and actors; Wang et al., 2013; Gondal, 2018; Purwitasari et al., 2020). In this second article, we advance in the understanding of the evolution of the scientific network methodologically for three 'levels' (i.e., organisations, actors, and journals), and multiple relationships (i.e., co-authorship and citation) between them. New and already available measures for diagnostics for statistical models for social networks are explored. The diagnostics (Hunter et al., 2008) helps to identify how the decisions of actors (Snijders, 2001) or the local neighbourhood of actors – as a subnetwork – can explain the emergence of the entire network (Robins et al., 2005) as a type of linkage between micro and macro level (Snijders & Steglich, 2015; Stadtfeld, 2018). Two research questions oriented this paper,

How a group of academics generate interpersonal intercitations considering the co-evolution of a multilevel network?
How well the micro-level represents macro features at the network level in a three-mode multilevel and multiplex networks?

The second article analyses the period of formation of the astronomical community after the arrival of ALMA – the largest radio astronomical facility - to understand the co-evolutionary interdependency of scientists and entities of different levels and the interpersonal intercitation patterns in a group of academics. The main methodological contribution is to expand available goodness of fit (Lospinoso & Snijders, 2019) for a three-mode (i.e., relationships within researchers, between organisations and journals) and multiplex network (i.e., citation and co-authorship) in the context of *stochastic actor-oriented models* (Snijders, 2001; Koskinen & Edling, 2012; Snijders et al., 2013). Some extensions and already available measures of the goodness of fit are then proposed for dyadic similarity-based mechanisms (i.e., E-I index, Yules Q, IQV and dyadic similarity distance-based for reciprocal ties), relational-based mechanisms (i.e., effective size, constraint and two overlapping triadic censuses) and proximity-based mechanisms (i.e., mixed multilevel degree distribution, mixed multilevel geodesic distribution, and mixed multilevel quadrilateral census). The results suggest that social relationships grounded on scientific collaboration and space proximity based on institutional affiliation are more accurately suited to understanding the networks' co-evolution in a scientific network than cognitive-based networks such as the journal network.

This contribution was made in collaboration with Elisa Bellotti, Martin Everett, and Christoph Stadtfeld. Alejandro wrote the complete draft, and he conducts the analysis, creates the codes, and suggests new measures for the goodness of fit. Elisa contributes to the complete supervision of the investigation and made theoretical clarification of the paper. Martin contributes to the investigation's full supervision and suggests incorporating some of the paper's goodness of fit. Christoph collaborates in exploring different potential micro-mechanisms, reviewing the adequacy of different modelling specifications, and interpreting some of the mechanisms contemplated in the model.

Study 3. The *cohesive subgroups* are challenging to delimitate in the current development of social networks (Fortunato, 2010; Fortunato & Hric, 2016; Everett & Borgatti, 2019), and from a sociological perspective in the study of science and knowledge, there was a frosty debate about their interpretation and delimitation (e.g., Collins, 1974; Small & Griffith, 1974; Woolgar, 1976; Mullins et al., 1977; Callon et al., 1983; Gläser, 2001). The last study offers an alternative of *cohesive subgroups* that relies on two assumptions: first, it aggregates researchers situated in the same organisations where it is reasonable to expect a minimum level of communication between co-workers in which colleagueship can be expected. Second, it incorporates researchers who share the same interests – publishing in the same discipline- as the members of these organisations. Using the notion of organisational fields from a network perspective (de Nooy, 2003; Powell et al., 2005; Ramos-Zincke, 2014), the study compares the presence of similar constraints distinguishing between those organisations that are in the core of the field in comparison with other institutions that are in the periphery (Borgatti & Everett, 2000). From a methodological perspective, when networks become bigger, they are often sparse, and some models for the analysis of social networks (e.g., Snijders, 2001; Lusher et al., 2012) become difficult to estimate. Therefore, this study advances in proposing an alternative that uses as a strategy a sample of *cohesive subgroups* – using as a methodological strategy a *second-zone multilevel sampling from a second-mode focal actor* – for networks that are 'not too small'. Two questions are considered in this paper,

> *How do the regular join patterns of intercitations among researchers in organisations – as meso-level social forces – vary within scientific communities?*
> *Do researchers in core organisations have similar patterns compared to other institutions on the periphery?*

The last article analyses the transitional period in which astronomers and astrophysics were preparing for the arrival of the Observatory C. Rubin (a.k.a., Large Synoptic Survey Telescope). This telescope will be one of the biggest of his kind and represent a branch of astronomy of the so-called 'data-turn' in astronomy (McCray, 2017). In 2017 the Chilean government considered the development of big data in astronomy as a potential area of interest for the development of this country (Espinosa-Rada et al., 2019; Arancibia et al., 2020). The main objective of this study is to extend the analysis of the case study to a sample of networks to address how the regular join patterns of *intercitations* among researchers in organisations, as meso-level social forces, have similar constraints within different organisations depending on their position (as core or periphery [Borgatti & Everett, 2000]) in the field, and considering interpersonal *intercitational* contexts. In this approach, it is used as a strategy a *meta-analysis stationary stochastic actor-oriented model* (Snijders & Baerveldt, 2003; Snijders & Stiglich, 2015; Block et al., 2019) and used as a methodological approach a *second-zone multilevel sampling from a second-mode focal actor* in networks that are 'not too small'. The second-mode focal actor is a *cohesive subgroup*, where institutional affiliation is first identified, and then all their members. All the researchers cited or cited from the organisation's members are considered, including their institutional organisations. The results indicate that researchers in this community are not preserving endogamic recognition within their organisations. And, actors tend to cite other researchers affiliated in the same research centres, creating closure in specific research niches due to diversity and multi-connectivity in this scientific community.

Chapter 2

Authors standing on the shoulders of other authors: Unpacking the author normalised weighted direct citation

Alejandro Espinosa-Rada*

**The Mitchell Centre for Social Network Analysis/Sociology, The University of Manchester*

Abstract

In this article, I review the author normalised weighted direct citation metric to decompose its main components and identify how each of its elements changes its prevalence according to different similarity measures. For the exploration, I decompose each of the author citation-based elements through an illustrative example to identify their particularities. Using an empirical case, I use multiple regression quadratic assignment procedures to identify each component's contribution according to the normalisations. The results indicate that according to the similarity, each of the citation components used to create the author normalised weighted direct citation can have a different contribution requiring further consideration in selecting the normalisation. I suggest that a possible interpretation in selecting the normalisations is whether there is an interest in the shared cognitive dimension according to the community's perception or the authors' communicational trace.

Keywords

Author Citation; Author Co-citation; Author Bibliographic Coupling; Author Normalised Weighted Direct Citation; Scientific Networks; Scientometrics.

2.1 Introduction

The use of scientific publications is the primary type of communication in science that allows the mapping of science. This mapping science can be done at different scales such as disciplines, research fields within disciplines, subfields, and research topics at the lowest level (van den Besselaar & Heimeriks, 2006). Researchers in (sub) fields share common base knowledge, in which they can identify a similar set of research questions, methodologies and shared overlapping references. Nevertheless, exploring these social-cognitive networks (Merton, 2000; White et al., 2004) requires a suitable representation, and recent developments suggested combining different perspectives. However, while the mixture of measures is suggestive in principle, it is not clear what are some of the consequences of the combination of different citation-based networks into a common representation.

There are different approaches to identify social-cognitive networks. Some of these include (1) *citation-based* measures, which depend on the references of the documents; (2) *text-based* measures that focus on the text and contents of the works (e.g., titles, abstracts, or full text); and (3) *hybrid* measures that employ both approaches (Ahlgren & Colliander 2009; Liu 2017). All these representations are part of interlocking multilevel networks (i.e., multi-modal, or linked networks) in which, for the same work it is possible to construct different types of networks. Example of these networks are bibliographic coupling, co-citation networks or direct citation networks (a.k.a. intercitation or cross citation) for the citation-based measures, and co-word of terms for the text-based measures. These networks are intrinsically related, and the derived networks (Batagelj & Cerinsek 2013; Batagelj 2020) can help us explore different dimensions of the scientific networks.

These representations have different interpretations that depend on references, but they focus on different social-cognitive dimensions. For Garfield, Sher & Torpie (1964) and Price (1965), *direct citations* – or the tendency of an author to cite another author – allows us to identifying development patterns within a particular field. The direct citations represent the self-organisation of current literature topics among authors (Klavans & Boyack 2017), often used to investigate how researchers '*stands on the shoulders of other authors*'. Also, coupling authors into a *co-citation* analysis – as the tendency of two authors to be cited together by later authors - represent the intellectual structure of a given scientific field (McCain, 1990). Co-cited authors tend to be grouped based on similar topics, methodologies, and social affinities as perceived by citers (White 2003), thus positioning the authors in a common intellectual space. The highly cited authors represent the field's knowledge base (Zhao & Strotmann, 2014). Another approach is a *bibliographic coupling* – two authors are coupled if they are citing the same authors in their references. The bibliographic coupling aims to identify research fronts or core documents (Glänzel & Czerwon, 1996) that reveal what active researchers are currently working on (Zhao & Strotmann, 2008, 2014; Klavans & Boyack, 2017).

A more recent alternative uses the three citation-based measures together through a normalised weighted direct citation to address the representations' differences. The assumption is that combining indirect citation (bibliographic coupling and co-citation) and direct citation allowed identifying cited authors that are out of the citing paper topic (Persson 2010; Wang et al. 2019). These measures can complement each other, obtaining additional information to improve the measures' reliability (Glänzel & Czerwon 1996). This tendency of combining different measurement is becoming more popular in recent years, for example, combining citation with the information of the venue of the publication and keywords (Bu et al., 2016) or the combination of multiple metadata for knowledge representations (Bu et al.,2018).

Much has been said about the accuracy and relevance of the different citation-based measures (i.e., direct citation, co-citation, and bibliographic coupling). Recent developments suggested the possibility of merging the three measurements to complement each other, known as the combined linkage or *normalised weighted direct citation* (Small, 1997; Persson, 2010; Wang et al., 2019). However, there is a lack of clarity in the implications of doing the combinations, which overlaps with two relevant discussions in scientometrics. First, there is an ongoing debate in understanding the accuracy of these measures and how they capture different dimensions through citations (Shibata et al., 2009; Boyack & Klavans 2010; Glänzel & Thijs 2017; Klavans & Boyack 2017). Second, because these networks depend on other networks, there is also an ongoing debate on dealing with the citation networks' projection (Ahlgren et al., 2003; Leydesdorff, 2008; Egghe & Leydesdorff, 2009; van Eck & Waltman, 2009). These two issues require firstly more clarity in what the citation-based measures mean and secondly the implications for the projection of the matrices often involved in the construction of these measures. However, there is less understanding of how sensible the weighted direct citation is when different similarity measures are considered for the normalisations. The citation-based alternatives have different meanings, and each citation-based option is reviewed to understand the weighted direct citation network in an empirical setting using multiple similarity measures. Clarifying these two dimensions allows identifying the implications of combining these representations.

In the following, I will address these three issues, in turn, to understand an empirical case study from the perspective of the author's works to identify the manifestation of the socio-cognitive interrelationships (Merton, 2000; White et al., 2004) of the authors in a scientific field. To do so, I disentangle with a detailed illustrative example of the construction of the four different measures identifying some of the implications and characteristics that arise from the matrix transformations of citation networks. Then, I highlight some of the current discussions according to normalisation processes often used in scientometrics literature. Next, through a quadratic assignment procedure (QAP), I use the different citation-based measures to understand their implication on this case of study. Then, I identify how these measures relate to each other and how the normalisation gives further relevance to the combined measure dimensions according to the normalisations. Finally, I highlight the main findings of the explorative analysis. I suggest that a possible interpretation of the normalisation difference may vary according to interest in the shared cognitive dimension according to the community's perception or the authors' communicational trace.

2.2 Scientific and Knowledge Networks

Most of the analyses for scientific networks use a standard strategies projection of two-mode networks[25]. Following this approach, and considering a W matrix of

[25] According to Borgatti and Halgin (2011a), 'two-mode graphs' refers to a representation of different types of entities or 'modes', and *affiliation* networks is a particular case that emphasizes

works-by-author[26] in which $w_{ij} = 1$ if the ith work is written by the jth author, and $w_{ij} = 0$ otherwise, then it is possible to analyse the relationship of both 'modes' directly as a two-mode network, or we can use both sets of nodes separately, or separately and then jointly (Everett and Borgatti, 2013, 2018). One of the most common strategies in scientometrics studies is to use the separation of nodes as an ordinary one-mode network using matrix projections. The projections of a matrix W, are formed by taking the product of the matrix W with its transpose to form WW', whose ijth cell gives the number of authors that both works i and j share and $W'W$, the number of common oeuvres (i.e., a body of writings done by a person [White & Griffith, 1981]) that share author i and j (i.e., collaboration network).

In this approach are that the value of the ijth cell in WW' or $W'W$ are often not considered. Also, there are some problems with the interpretation of scientific networks. For example, in co-authorship, papers with many authors produce large complete sub-graphs that obscure some collaboration structures by over-representing works with many values (Batagelj & Cerinsek, 2013). Some alternatives address these issues, but I first disentangle the projections used as the baseline of direct citation, co-citation and bibliographic coupling as the building block for the normalised weighted direct citation.

2.2.1 Direct Citation Networks

Previous research has emphasised that direct citation is more accurate for capturing research fronts (Shibata et al., 2009) and trace well the socio-cognitive and historical development of knowledge (Klavans & Boyack, 2017). The directed citation network is based on the product of two two-mode matrices. The first matrix is based on a given matrix X of authors-by-works in which $x_{ij} = 1$ if the ith author's *oeuvre* cite a jth work, and $x_{ij} = 0$ otherwise. For simplicity, I will assume

relations such as participations or memberships. For example, people and demographic characteristics (case-by-variables matrix) can be represented as two-mode graphs but is not considered an affiliation network.

[26] As it was mentioned by White and Griffith: ''Author' in this context means something like what the French call an *oeuvre* – a body of writings by a person – and the person himself.' (1981: 163)

that each author in X is a first solo author of a set of citing work for illustrative proposes. Also, and together with a second matrix W as defined previously, the product of $XW = C$, in which C is the matrix used for the direct citation. For this particular case, the ijth values of C is the number of ways an author i cite (or can reach) another author j. Which is the trace from the jth work in X, and the same work ith in W. The matrix C is the baseline for *co-citation*, and *bibliographic coupling*.

In comparison with co-citation and bibliographic coupling, direct citations focus on the *links* and the relevance of the structures (Hummon & Doreian, 1989), as an iteration and flow of information that is aggregated in the direct author citation to investigate how researchers '*stands on the shoulders of other authors*'. Garfield uses the metaphor that 'If one consider the book at the macro unit of thought and the periodical article the micro unit of thought, then the citation index in some respects deals in the submicro or molecular unit of thought' (1955: 122) because researchers often cite other contributions due to particular ideas instead of complete concepts. However, recall that this limited measure is blind to the citation's intention (e.g., Mulkay, 1974; Gilbert, 1977; Nicolaisen, 2008; Milard, 2014), and the act of citing and the value involved in the tie is difficult to interpret. For example, I can cite an author once in the reference, but the hypothetical author can be more relevant for the paper's argument. In contrast, a different second author can be mentioned several times and might have a secondary relevance in the same work.

In the following example, I illustrated how two non-negative matrices, X (Table 1) and W (Table 2), are connected through the same work as is represented in a temporal two-mode graph (Figure 5). The temporal two-mode graph $G = (V, U, T_1, T_{t-1})$, has a set of authors $V = \{a_1, \dots, a_n\}$ and a different set of works $U = \{p_1, \dots, p_k\}$. And, has two different types of temporal connections, the ties T_t is a set of edges that connect authors in V to works in U in time t, and T_{t-1} as a different set of edges in which work in U connects to an author V in time $t - 1$. Notice that in the example (Table 1), author 4 is not present in matrix X (time t) but appears in W (time $t - 1$), this node is assigned with the value of 0 for the missing edges in this illustration (as in Table 1), and not incorporated in the empirical analysis in the forthcoming section. Also, consider that X and W are incident matrices, referred to as occurrence matrices in the literature of scientometrics.

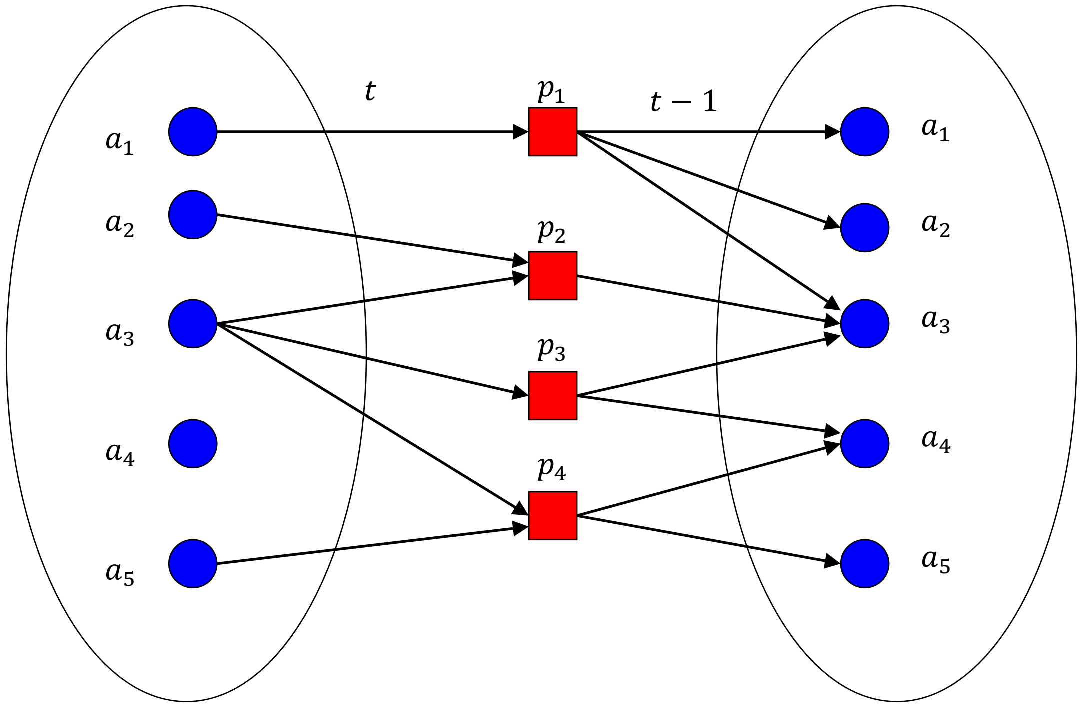
Table 1 Author Citing Papers

	Paper 1	Paper 2	Paper 3	Paper 4
Author 1	1	0	0	0
Author 2	0	1	0	0
Author 3	0	1	1	1
Author 4	0	0	0	0
Author 5	0	0	0	1

Table 2 Paper Cited and Their Authors

	Author 1	Author 2	Author 3	Author 4	Author 5
Paper 1	1	1	1	0	0
Paper 2	0	0	1	0	0
Paper 3	0	0	1	1	0
Paper 4	0	0	0	1	1
Total:	1	1	3	2	1

Figure 5 A Temporal Two-mode Network



Author citing paper in time $\boldsymbol{t}$ (author sender and paper receiver), and paper cited and their authors in time $\boldsymbol{t-1}$ (paper sender and author receiver).

Through the multiplication of the two-mode matrices X and W, the direct citation network is derived in Table 3 and represented as a graph in Figure 6, in which the diagonal corresponds to the *authors' self-citations*. As can be noticed, the direct citation represents the *walks* of the authors and the parallel dissemination of the written paper to several other authors (i.e., *diffusion by replication* [Borgatti, 2005]) that are aggregated considering the scientific documents of early works and their researchers.

Table 3 Direct Citation as the Number of Walks of Distance Two from Author ***i*** to Author ***j***

	Author 1	Author 2	Author 3	Author 4	Author 5
Author 1	**1**	1	1	0	0
Author 2	0	**0**	1	0	0
Author 3	0	0	**2**	2	1
Author 4	0	0	0	**0**	0
Author 5	0	0	0	1	**1**

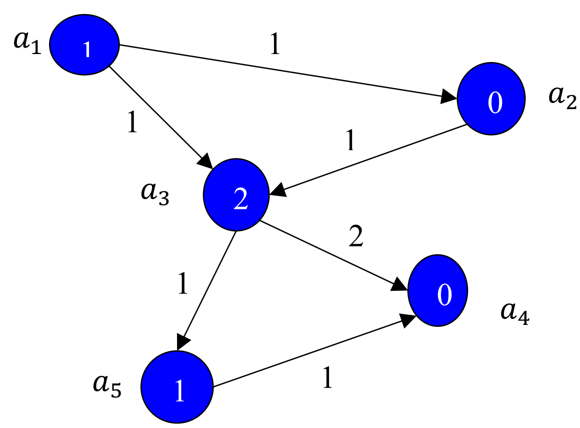


Figure 6 Projection of the Number of Walks of Distance Two from Author ***i*** to Author ***j***

In this representation, the matrix C (Table 3) identifies the number of walks in which an actor in time t reach at a distance of two another author in time $t-1$. Also noticeable is that author three cite in two different papers (p_3 and p_4) author four, while the other authors are cited only through one document (Figure 5). From this representation, the number of walks increases because of the number of publications that refer to the same author[27]. For example, author three contributes

[27] Another area for mapping networks focused on finding the main paths considering the direct citation. The main paths often adopt a direct acyclic graph (DAG) structure (while in rare circumstances, there might be loops, especially when there are 'forthcoming' references) (Hummon

to 45% of the total walks from t to $t-1$. This representation will consider that author four is more cited without considering the self-citations (Figure 6). However, in this illustrative example, the popularity of author four is given by the number of publications that traverse the relationship between the authors. The matrix C is used as the building block for the upcoming representations.

2.2.2 Co-citation Networks

The author co-citation allowed mapping scientific networks (Rosengren, 1968; Small, 1973; Marshakova, 1973 for co-citation; White & Griffith, 1981 for author co-citation). This measure assumes that if two authors are cited together by later authors, then it is more likely that their intellectual oeuvres are perceived as related in a field even if they do not create a direct citation. Co-citation often represents well-established authors' work. It is assumed that oeuvres need to mature to appear (Zhao & Strotmann 2008) and reflect the past structure of knowledge (van den Besselaar & Heimeriks 2006). Hence, this measure (and *bibliographic coupling*) can be seen as a manifestation of an intellectual relationship that is perhaps unobservable through direct citation. We could deduce the co-citation matrix using *direct citation*

$$coCi = C'C$$

Following the illustrative example, through the citation network the co-citation matrix is constructed (Table 4 and Figure 7).

& Doreian, 1989; Batagelj et al., 2014). Here, I limited the review to the most straightforward representation used as the building block for other citation representations. Also, is used direct citation as a one-time window ($t-1$), which can expand for longitudinal couplings (Small, 1997).

Table 4 Co-citation Full Counting Matrix

	Author 1	Author 2	Author 3	Author 4	Author 5
Author 1	**1**	1	1	0	0
Author 2	1	**1**	1	0	0
Author 3	1	1	**6**	4	2
Author 4	0	0	4	**5**	3
Author 5	0	0	2	3	**2**

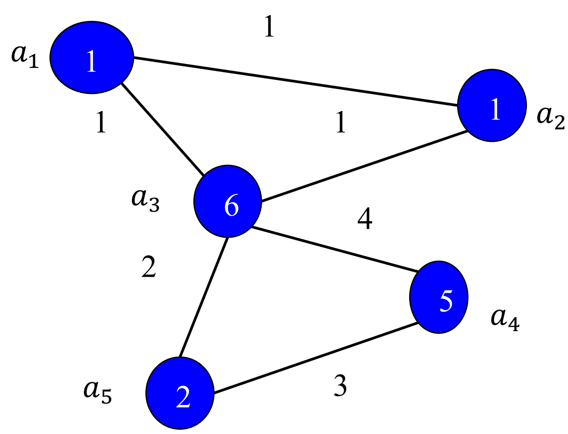


Figure 7 Co-citation full counting network

Intuitively, this representation assumes that two authors would be co-cited if both appear in the references of a common work from later works. In this network, author 3 is publishing paper 2, paper 3 and paper 4. Therefore, the number of walks from co-citing author 3 and author 4 at time $t-1$ from the perspective of authors in t at a distance of two are from $A3 \rightarrow P2 \rightarrow A3$, $A3 \rightarrow P3 \rightarrow A3$, $A3 \rightarrow P3 \rightarrow A4$ and $A3 \rightarrow P4 \rightarrow A4$. Which are coupled directly in paper 3 and 'indirectly' couple by authors through the other papers written by author 3. Similarly, the number of walks coupling author 4 and author 5 are from $A3 \rightarrow P3 \rightarrow A4$, $A3 \rightarrow P4 \rightarrow A4$ and $A5 \rightarrow P4 \rightarrow A5$ that are directly co-cited in paper 4, but 'indirectly' couple by author considering the other walks. A third example is from author 3 to author 5, who is not directly connected by any paper but is 'indirectly' connected by author 3 through the walk $A3 \rightarrow P3 \rightarrow A3$ and $A3 \rightarrow P4 \rightarrow A5$.

In the example, one author (i.e., author 3) can increase the number of walks between two other actors when there are more cited actors in the references, increasing the co-cited authors' weight.

2.2.3 Bibliographic Coupling

The bibliographic coupling introduced by Kessler (1963), and further expanded for authors by Zhao and Strotman (2008), is considered a complement of co-citation. This measure relies on the coupling of authors by other researchers in the community, without the cited authors' active engagement as in the direct citation. Bibliographic coupling describes better the research-front topics (Glänzel & Thijs 2017), but sometimes this might depend on the subject of interest (Boyack & Klavans 2010). The bibliographic coupling has the limitation that two articles might cite different references and still be coupled together. Also, because the inter-citations are very sparse, the coupled bibliography is often dominated by a few citations (Liu 2017). This measure is mathematically similar to co-citation,

$$biCo = CC'$$

However, bibliographic coupling focuses on how two authors are coupled if they are citing the same authors in their references (consider that the C' is equivalent to changing the direction of all the ties in C and remaining constant the diagonal). In this example (Table 5 and Figure 8), author two and author three are tied with a value of 2 because they are both citing paper 2.

Table 5 Bibliographic Coupling with Full Counting Matrix

	Author 1	Author 2	Author 3	Author 4	Author 5
Author 1	**3**	1	2	0	0
Author 2	1	**1**	2	0	0
Author 3	2	2	**9**	0	3
Author 4	0	0	0	**0**	0
Author 5	0	0	3	0	**2**

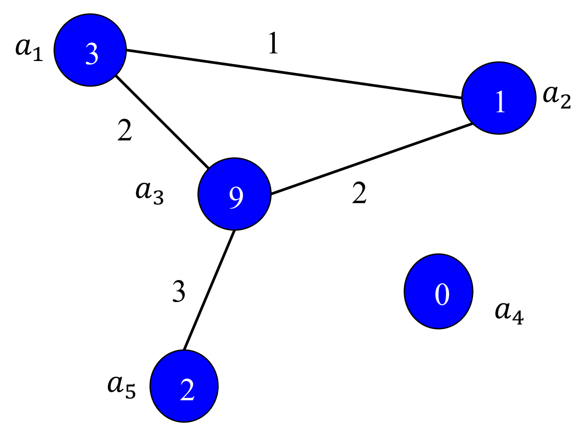


Figure 8 Bibliographic Coupling Full Counting Network

Therefore, the assumption is that because paper 2 is written by author three, then there is a walk of distant two from $t-1$ to t from $A3 \rightarrow P2 \rightarrow A2$ and $A3 \rightarrow P2 \rightarrow A3$. Similarly, author three and author five have a bibliographic coupling because there are three walks of distant of two that indirectly connect them from $t-1$ to t (i.e., $A4 \rightarrow P3 \rightarrow A3$, $A5 \rightarrow P4 \rightarrow A3$ and $A5 \rightarrow P4 \rightarrow A5$). Notice that this walk depends on time t. Therefore, author 4 disconnect the network because it only appears in $t-1$. Also, self-citation (such as author 1 citing itself and author 3) assumed in this representation that there is a direct connection between its work and the cited work. However, some references, and therefore their authors, might be out of topic.

Certain studies suggest that bibliographic coupling and co-citation analysis are more accurate for the long-term manifestation of relevant topics (Klavans &

Boyack 2017). In terms of clustering, bibliographic coupling tends to cluster recent papers and the old papers to a lesser extent. On the contrary, co-citation clusters old works but cannot cluster recent documents that are not yet cited, and direct citation tends to cluster more important documents (Boyack & Klavans 2010).

2.2.4 Normalised Weighted Direct Citations

Direct citation, bibliographic coupling, and co-citations tend to have different meanings, trace the walks' flow differently, and emphasise different dimensions of the temporality of the works and their authors. Small (1997) suggested using the three perspectives combined into a single measure to recover multi-year citation. A more recent approach has suggested a potential combination of the direct citation with weighted co-citation and bibliographic coupling to compute a new measure. In the example of Persson (2010, Figure 9), paper A is citing directly paper B (i.e., a direct citation). However, the link is enhanced because they are both citing paper C (i.e., bibliographic coupling), and D cites A and B (i.e., co-citation).

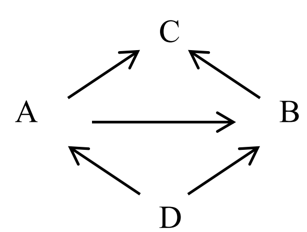


Figure 9 Diagram of the Normalised Weighted Direct Citation (extracted from Persson, 2010)

The generalisation of the normalised weighted direct citation (Wang et al., 2019) stands that author A citing author B is represented as,

$$NWDC_{AB} = \sum_{r \in R} w_{r_a} * w_{r_b} (X_r + Y_r + Z_r)$$

Where, X_r is the normalised direct citation[28], Y_r the normalised co-citation strength, and Z_r is the normalised bibliographic coupling, and R is the set of work relationships. In which the work relationship is expressed as,

$$R = \{r | r = A_i \rightarrow B_j, work\ A_i \in work\ set\ A, work\ B_j \in work\ set\ B\}$$

Also, work set A is author A's oeuvre, and the set B are the author B's oeuvre. Then, r_a are the work of author A in work relationship r, and r_b are the work of author B in work relationship r. As was noticed in the illustration before, the connections between authors is based on the intermediary works and, for simplicity, each author was considered to have one oeuvre without co-authorship. The generalisation of Wang et al. (2019) allowed expressing a more accurate situation that also included the set of works in A. However, they also included w_{r_a} as the A' contribution to paper r_a, and w_{r_b} the contribution of author B to paper r_b, which considered that author in A′ can be in different positions of the byline hierarchy of the work (e.g., the first author of a work, in the middle or the last author). If the position of the author A and B does not matter, then $w_{r_a} = w_{r_b} = 1$, which is often the case when a solo or first author extraction is considered.

Using the notation above for the entire matrix, and without any type of normalisation or positional contribution in the work byline hierarchy, then, the expression only takes into consideration the strength of the citation communication in which the equation is just

$$NWDC = C + coCi + biC$$

[28] Notice that in Wang et al. (2019), the direct citation is not normalised, assuming that this component is more relevant than the others and adds according to the relationship's strength. For Small (1997), direct citation weighted twice than any indirect tie. I will not make that assumption due that is not clear what the intensity of the connection could mean for each author, and instead, I will also normalise this measure for the following exploration. Further consideration should explore potential limitations, for example, to maintain the direction of the relationships or the intensity of the relationship without normalisation.

Table 6 Weighted Direct Citations with Full Counting Matrix

	Author 1	Author 2	Author 3	Author 4	Author 5
Author 1	**5**	3	4	0	0
Author 2	2	**2**	4	0	0
Author 3	3	3	**17**	6	6
Author 4	0	0	4	**5**	3
Author 5	0	0	5	4	**5**

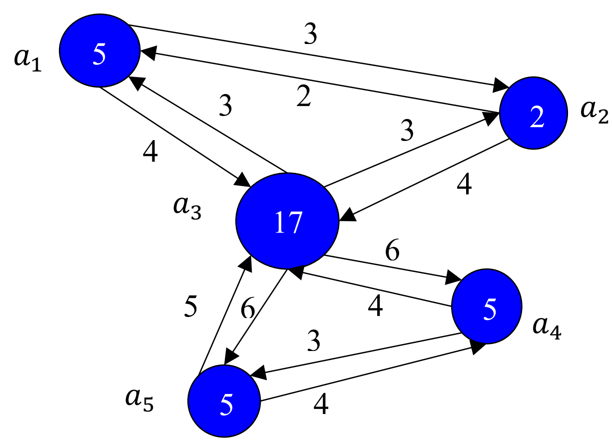


Figure 10 Weighted Direct Citations with full counting network

The ties are now enhanced in the illustrative example, adding the different measures, and considering their weight (Table 6 and Figure 10). As can be noticed from the arcs, the number of papers plays an indispensable role in the co-citation and bibliographic coupling, increasing mutual co-occurrences, and one author is capable of bias in the number of walks increasing co-occurrences. For example, considering the tie from author three and author four, author three is responsible for all the walks between itself and the other author. Author three add two direct ties through paper 3 and paper 4 (i.e., $A3 \rightarrow P3 \rightarrow A4$ and $A3 \rightarrow P4 \rightarrow A4$), and then four possible ties through co-citations passing through paper 2, paper 3 and paper 4. Bibliographic coupling in this particular case plays no role in the connection because author four is not present in time t. However, when the weighted direct citation from author three and author five is considered, there are many internal variations in the type of walks and have the same weighted score as the ties reviewed before. From author three and author five, there is only one direct citation (i.e., $A3 \rightarrow P4 \rightarrow A5$), author three is also responsible for the two co-citations (through paper 2 and 4, and the second arc passing through paper 3 and 4 to connect author three and five). Authors four and five contribute to three arcs passing through papers 3 and 4 to connect authors three and five. In both cases, the number of papers increases the reachability between the authors.

Up to now, I have not normalised any of the measures considered. Instead, I explore the weighted direct citation directly without further normalisation. Many limitations arise from the example. One of them is that I did not focus on the co-authorship network's potential role, which can partially explain some of the walks, increasing the possible paths. For example, author two and author three can be part of the same paper changing the citation's interpretation, in which case, it is not possible to distinguish which of the author decided to create some of the ties without further information or additional assumptions. Also, specific authors can be cited many times in the list of reference, which will increase its weight considerably, and the network will no longer reliably inform about the relevance of the number of intervening documents.

To avoid walks' overabundance in the analysis, some researchers differentiate between citations that rely on *text* or socio-cognitive dimensions and social networks of concrete relations, or proxies, between *agents* (Leydesdorff, 2008). This differentiation affects how the matrices are treated for citation networks, but the suggested difference for citation networks disappears when the matrices are dichotomised but are relevant when the weighted and normalisation are considered. For example, Leydesdorff & Vaughan mentioned that 'if an author is cited twice in one (set of) papers and three times in another, the number of *affiliations* [e.g., Borgatti & Halgin, 2011a] – as this measure is called in social network analysis – is 6, while the number of co-occurrences remains only 2' (2006: 1625). Considering the illustration described, let suppose that there are three papers and two authors in an W asymmetrical matrix (Table 7).

Table 7 Asymmetrical Occurrence Matrix

	Author 1	Author 2	Total:
Paper 1	2	0	2
Paper 2	3	0	3
Paper 3	0	1	1

Following classical projections (WW'), the asymmetrical occurrence matrix often used from a social network perspective is derived from a matrix multiplication (Table 8) as was previously described.

Table 8 *Affiliation* of the Occurrence Matrix (Projection)

	Paper 1	Paper 2	Paper 3
Paper 1	4	6	0
Paper 2	6	9	0
Paper 3	0	0	1

If the second option is used, then the co-occurrence matrix is derived from the minimum overlapping (Table 9). In the diagonal of the matrix, the marginal is imputed (i.e., the total number of citations in the example) to recover some of the lost information during the matrix transformation (Leydesdorff, 2008), retrieving the underlying distribution of the original matrix. For this matrix, each possible mutual dyads of papers are compared (paper 1-2, 1-3 and 2-3) to identify the minimum in which they co-occurred using Morris (2005) nonbinary overlap function. For example, for the dyad of papers 1 and 2, paper 1 is citing the first author two times, and paper 2 is citing the same author three times in the occurrence matrix. Therefore, the minimum overlapping of that mutual dyad is 2 in the resulting co-occurrence matrix (equation reviewed in next section).

Table 9 The Co-occurrence Matrix (Minimum Overlapping)

	Paper 1	Paper 2	Paper 3
Paper 1	2	2	0
Paper 2	2	3	0
Paper 3	0	0	1

Consequently, normalisations aim to overcome some of these issues when the network projections are used because the number of ties in the transformed network can be distorted with the size of the elements involved and the walks considered. This behaviour of the matrices was identified at the beginning of the formulation of some of these measures (e.g., Small, 1973; White & Griffith, 1981; Salton & McGill, 1983). For example, when there are too many authors in a paper or many references in a document, it will give more predominance to the researchers, which will seem similar to other researchers that also appear more. Also, these measures are sensitive to the productivity of an author. Notice that the total number of citations of authors in W in Table 2 is no longer available in other representations.

The theoretical assumption is that the walk's weight in *intellectual* networks should be treated differently as information events, primarily when co-citation and bibliographic coupling are used. The network focused on how often an author appears in the representation. Borgatti (2005) identifies this issue more generally, considering that there are expectations of the flows according to different centrality measures in social networks (which should be considered in the author citation-based representations). Therefore, the normalisation used implicitly can assume that the flowing of 'unit of thought' in the citation network has a socio-cognitive dimension. The normalisations can either use the weight as a proxy between agents through conventional projections or consider the overlapping between their ideas. These are two theoretical distinctions used to estimate similarity measures.

2.3 Normalising Weighted Citation-based Networks

It is not straightforward to decide how to normalise the different matrices to identify the two authors' underlying ties. There is a recent consideration in simultaneously using the two projections, which does not necessarily imply losing information (Everett & Borgatti, 2013, 2018). According to Leydesdorff, in the internet research 'one often can no longer retrieve the entire document set [i.e., the original matrix] that is needed to construct the co-occurrence matrix, but one can construct these matrices directly, for example, by searching in a domain with Boolean ANDs' (2006: 1616), which is a limitation for the scientometrics analysis.

At least three different options are often used in scientometrics to overcome some of the issues that arise from the transformation of the matrices. One of them is to dichotomise the matrix (e.g., when $|x_{ij}| > 0$, as formulated by Breiger [1974]) considering a specific threshold, which leads to the arbitrariness of deciding which value is the most suitable for $x_{ij} = 1$. However, dichotomisation might overcome some of the challenges of using weighted networks (e.g., the number of cliques that often arise). The *backbone* can reduce the original network and preserve the most significant edges considered more significant (Neal, 2014). Different alternatives are available in the literature to create a more reasonable dichotomisation. However, there is no straightforward method to distinguish which is the best option while new options are becoming available (Schoch, 2021), and due to that, few authors can bias the number of possible walks. This issue requires further scrutiny for citation networks. While the abundancy of alternative strategies is becoming available, a further question is whether to create the threshold initially or if it is preferable to use a weighted network with the full information instead.

A second alternative is assuming the relevance of the position of the authors in the papers. One of the main alternatives is the normalisation process for co-authorship proposed by Newman (2001c) that has been recently discussed with different alternatives to conduct a fractional approach within different bibliometric measures (Batagelj & Cerinsek, 2013; Perianes-Rodriguez et al., 2016; Leydesdorff & Park, 2016; Batagelj, 2020). The fractional counting approach assumes that each action should have equal weight, regardless of the number of authors, citations, or references of a publication (Perianes-Rodriguez et al., 2016). This perspective might be reasonable in scientific disciplines in which alphabetisation is more common (e.g., mathematics, economics, or high energy physics) or in situations in which the scientific group are more preeminent in the decision of citing. For the directed citation network, on the other hand, we could approximate the weighted contribution of each author using some known weights, *fractional approach* (Leydesdorff & Park, 2016) or assuming a *harmonic counting* to quantify the byline hierarchy (Hodge & Greenberg, 1981; Hagen, 2013). However, previous knowledge is needed to identify the practices of citing according to different areas of knowledge, and Wang et al. (2019) explored this alternative for *author weighted direct citation* using harmonic counting.

A third option, and a popular alternative often used to analyse scientific networks, is based on specific types of local similarity between nodes to identify how similar the two authors are. This strategy is often used as an input for multidimensional scaling to visualise maps of scientific networks (visually represented as close relationships between the authors' positions). One of the drawbacks of this alternative is that there are many different options available, and it is not simple to discern which one is a better option (Ahlgren et al., 2003)[29]. However, in scientometrics, some of the normalisations most used are the *Jaccard* and *cosine* that have been traditionally the most compared similarity measures for symmetric and asymmetrical matrices (Leydesdorff, 2008), and *association strength* is also suggested as a preferable candidate in scientific networks (van Eck & Waltman, 2009; Egghe & Leydesdorff, 2009).

Considering some notations, and generalising the previous matrices, CB is expressed as any of the three citation-based matrices reviewed before (i.e., C, $coCi$, biC). Also, let $(cb)_{ik}$ denote the element in the ith row and kth column of CB in which $n_{ij} = \sum_{k=1}(cb)_{ik}(cb)_{kj}$ is the number of common neighbours k between node i and j. And, considering the row sum of j in the matrix CB (i.e., the out-degree), then $k_j = \sum_{i=1}(cb)_{ij}$, and the column sum of the i node in the matrix CB (i.e., the in-degree) is $k_i = \sum_{j=1}(cb)_{ij}$.

---

[29] Other common strategies emphasise *links*, often used in the social network perspective (for a review, see Wasserman & Faust, 1994; Borgatti & Halgin, 2011a; Borgatti, Everett & Johnson, 2018).

There are two different options to specify k_i and k_j, which are often used for further normalisations in scientometrics (Ahlgren et al., 2003; Leydesdorff, 2008; van Eck & Waltman, 2009). One option is not considering the self-edges from i to i (i.e., self-citation, the number of walks considering the mediating papers from the reviewed networks or the diagonal in the overlapping co-occurrence matrix), as it is regularly used in the network perspective (Borgatti & Halgin, 2011a; Newman, 2018; also, Ahlgren et al., 2003 in scientometrics). The other option is imputing information from previous matrices into the diagonal of the derived matrices. For the case of the direct citation, to calculate k_i and k_j extra information is extracted from the W matrix imputed in the self-edges $C_{ii} = \sum_{j=1} W_{ij}$. And, for the cases of co-citations and bibliographic coupling, the information is extracted from the C matrix imputed in the self-edges $coCi_{ii} = \sum_{j=1} C_{ij}$ and $biC_{ii} = \sum_{j=1} C_{ij}$, in which the self-edge is considered for the estimation of k_i and k_j. The second strategy is preferred in the literature of scientometrics (Leydesdorff, 2008; van Eck & Waltman, 2009; Zhou & Leydesdorff, 2015).

For the case of *Jaccard*, this index is defined as the ratio between the number of times a relation between two authors is observed together divided by the number of times k_i or k_j are observed.

$$J_{ij} = \frac{n_{ij}}{k_i + k_j - n_{ij}}$$

And, when the data is not binary but have integers (as C), there are some options for weighted networks using *Weighted Jaccard* (a.k.a., *Ružička*) (see Schubert 2013; Schubert & Telcs, 2014), defined as

$$J_{w_{ij}} = \frac{\min(n_{ij})}{\max(n_{ij})}$$

In particular, some of the characteristics of the *Jaccard index* are that it does not take into account the shape of the distribution because it relies on the *intersection* of two sets considering the sum of the two sets (Leydesdorff, 2008), while strongly skewed distribution can be addressed using *weighted Jaccard* (Schubert & Telcs, 2014), which is often the case for measures based on citations.

The *cosine* (a.k.a., *Ochiai coefficient* or *Salton's index/measure*), on the other hand, is defined as the ratio between the number of times of relationships between two authors and the geometrical mean, in which k_i and k_j are observed. Its interpretation is according to the angle between the two elements that are normalised, in which their directionality is explored (0 if they are orthogonal [perpendicular and therefore independent] or 1 if they are pointing in exactly the same direction) and not according to its magnitude.

$$Co_{ij} = \frac{n_{ij}}{\sqrt{{k_i}^2}\sqrt{{k_j}^2}}$$

In scientific networks, *Salton cosine similarity* is often recommended instead of *Pearson correlation* due that the later normalise to the mean distribution, which is not the case of *cosine*, which can be seen as a nonparametric version sufficiently able to deal with skewed distributions and the prevalence of zeros in citation matrices (van Eck & Waltman, 2008; Zhou & Leydesdorff, 2015). Notice that *cosine* and *Jaccard* are based on relative overlapping patterns. And, when only the co-occurrence matrix is available, the *Ochiai* coefficient is equivalent to the cosine similarity in the occurrence matrix (see Zhou & Leydesdorff, 2015), in which case in the *Ochiai* coefficient is expressed as

$$Oc_{ij} = \frac{\min{(n_{ij})}}{\sqrt{{k_i}^2}\sqrt{{k_j}^2}}$$

Finally, despite the popularity of *cosine* and *Jaccard*, the *association strength* (a.k.a., *probabilistic affinity index*, *proximity index*, *pseudo-cosine*) is also an alternative for similarities for scientific networks (van Eck & Waltman, 2009). This measure builds on the ratio between the observed tie and the expected values of the same tie based on its degree, assuming statistical independence.

$$A_{ij} = \frac{n_{ij}}{k_i k_j}$$

One of this measure's characteristics is that it corrects for the size effect (van Eck & Waltman, 2009). On the contrary, *Jaccard* and *cosine* do not correct for the size effect and, as a consequence, they have on average higher values of ties that occur more frequently. In comparison with these measures, *association strength* does not depend on the frequency in which the ties occur.

From a more substantive interpretation, *cosine* and *Jaccard* are based on overlapping measures, while *association strength* is based on the expectation of observed ties compared to expected ties (van Eck & Waltman, 2009). This distinction can be related to the socio-cognitive dimension, in which the overlapping measures can be considered to be closer to a *text* or *cognitive* interpretation of citation. Simultaneously, the difference between observed and expected ties relies more on the structural social dimension of the citations.

2.4 Database

For this comparison, I will use the Chilean community's information extracting the complete record of all researchers that were institutionally affiliated in Chile in 2017 and published in the topic of 'Astronomy and Astrophysics' in the Microsoft Academic database. One of the features that have Microsoft Academic in comparison with other well establish databases (such as Google Scholar, Web of Science, Astrophysics Data System and Scopus-Elsevier) is that it can extract complete records of all references of each paper (which is currently a limitation in databases such as the Web of Science or Scopus Elsevier), and recent studies show that cover a significant amount of citation in comparison with similar databases (Martín-Martín et al., 2021; Visser et al., 2020).

With the references' information, it is possible to distinguish between directed citations (from paper p to paper q) with complete information of each of the refereed papers, allowing extracting complete co-author information. However, I limit the analysis to Chileans' references to analyse a local geographically environment and the particularities that this community has when their institutions are settled in Chile. After extracting the data, I manually disambiguate the institutional affiliation and the authors of the database. The total numbers of researchers (cited and citing) are $a = 1{,}021$, and the total number of papers (cited and citing) are $p = 3{,}105$, some basic description of the cited authors and papers in Table 10.

Table 10 Descriptive of the Chilean Astronomers and Astrophysics Database

	Size	Mean	Standard Deviation	Median	Minimum	Maximum
Cited papers	2,505	23.132	16	23.978	1	199
Cited authors	943	5.621	9.618	2	1	102

The main reason to demarcate the country and the discipline is that the astronomical community in Chile is considered to be small (~255 astronomers in 2019). Also, according to the census of astronomers of the Chilean Astronomical Society (SOCHIAS), the 70 per cent of the earth infrastructure for astronomical observations will be settled in this country, and these astronomers have 10 per cent of the total observation of these telescopes if they are working in an institution that is also settled in the country. Therefore, creating this boundary will be used as a proxy to explore the citation between actors that, because of the small size of this community, they are likely to know each other because of their joint participation in committees[30], they are likely to compete for the same national funding and for

[30] For example, the Chilean Telescope Allocation Committee (CNTAC), Chilean Telescope Allocation Committee for APEX telescope, the Chilean Telescope Allocation Committee for Gemini Sur telescope, ALMA-CONICYT Committee, CAS-CONICYT Committee, NAOC-SOCHIAS Committee, GEMINI-CONICYT Committee, ESO-Chile Committee, QUIMAL Committee, among others.

the allocation of time observations of these telescopes. This feature allows identifying not only the cognitive dimension but also the social elements.

2.5 Quadratic assignment procedure

Different normalisations are compared in empirical settings often used Pearson and Spearman correlation to identify if there is much difference between the various measures and whether these measures are monotonically and/or linearly related (van Eck & Waltman, 2009). I will also use the quadratic assignment procedure (QAP) to control the interdependency often present in dyadic network data (Krackhardt, 1988; Dekker et al., 2007). The main characteristic of QAP is that allowed to correlate whole matrices and estimate its significance comparing the observed matrix with a correlation of thousands of reference set of matrices in which their labels (rows and matching columns) are permuted. This strategy allows to maintain the network structure but knowing that the permuted matrices are independent of the observed networks. Then, the *p-values* are estimated, comparing how different was the proportion of correlations of the independent matrices with the observed correlation.

Also, I use the multiple regression quadratic assignment procedure (MR-QAP) using double semi-partialing (Dekker et al., 2007) to identify the contribution of different citation-based matrices to the normalised weighted direct citation measure. Parameters of MR-QAP can be interpreted in the same way as ordinary least squares (OLS) analysis, are suitable for weighted networks, deal with fluctuations of time, and make statements about effect size (Elmer & Stadtfeld, 2020). For this case, the full model can be expressed as,

$$NWDC \ = \ \beta_0 + \beta_1 C + \beta_2 coCi + \beta_3 biC + e_{ij}$$

Where $NWDC$ is the normalised weighted direct citation as the dependent variables, C stands for direct citation, $coCi$ co-citation and biC bibliographic coupling, and parameter β_k are coefficients, and e_{ij} the error terms of authors i and j.The dependent variable should be interpreted at the dyadic level of how similar are two authors, and because this socio-cognitive network is not independent, the standard errors obtained through OLS should not be considered.

2.5.1 Empirical analysis of the normalisations

In this section, I analyse the contribution of the metrics of direct author citation, bibliographic coupling, and the co-citation to the weighted direct citation when different normalisations are used. In the following, I identify the correlation of each combination of citation-based measures according to the different normalisations to identify empirically how these dimensions are related. Then, and for each combination of the citation-based measures, I conduct an MR-QAP to identify which dimensions have more prevalence when Jaccard, cosine and association strength are used. I identify for each combination of the citation-based measures how strongly related are these measures with each other using correlation. I use Pearson correlation to identify the linear relation between the matrices and Spearman correlation for the monotonical relations (reported in Table 11). For this data, there are apparent differences between the use of different normalisations and their relation.

Considering Table 11, when weighted Jaccard is used in this empirical case, the normalised weighted direct citation is very strongly and monotonically (and linearly) related with co-citation, while direct citation and bibliographic coupling are strongly related (monotonically and linearly) between each other. This can be interpreted as a predominance of overlapping between members that appear in the co-citation to also appears in the normalised weighted direct citation, and the same between direct citation and bibliographic coupling. Moreover, because the weighted Jaccard does not consider the slope of the distribution (e.g., number of times the citation has been cited), bibliographic coupling and direct citation seem to be more related than the other measures.

Table 11 Correlation Based on Quadratic Assignment Procedure with Normalisation

	Direct Citation	Bibliographic Coupling	Co-citation	Normalised Weighted Direct Citation
Correlation considering Weighted Jaccard				
Direct Citation	-	0.630***	0.050***	0.306***
Bibliographic Coupling	0.845***	-	-0.055***	0.342***
Co-citation	0.168***	0.142***	-	0.851***
Normalised Weighted Direct Citation	0.750***	0.753***	0.740***	-
Correlation considering Cosine/Ochiai				
Direct Citation *(Cosine)*	-	0.120***	0.241***	0.354***
Bibliographic Coupling *(Ochiai)*	0.057***	-	-0.153***	0.476***
Co-citation *(Ochiai)*	0.318***	-0.115***	-	0.743***
Normalised Weighted Direct Citation	0.567***	0.527***	0.732***	-
Correlation considering Association Strength				
Direct Citation	-	0.084***	0.484***	0.513***
Bibliographic Coupling	0.049***	-	0.068***	0.300***
Co-citation	0.416***	0.146***	-	0.865***
Normalised Weighted Direct Citation	0.942***	0.252***	0.657***	-

Note: *** $p < 0.001$. The numbers of draws to use for the quantile estimation are 3,000. Upper right triangle are Spearman correlations, and lower left triangle are Pearson correlations.

For the case of cosine and Ochiai index, co-citation still prevails in its strong correlation with normalised weighted direct citation overlapping more. However, when the imputed diagonal is considered in the numerator, and because this measure deals better with zeros and skewed distributions, bibliographic coupling and direct citation do not longer correlate much. The bibliographic coupling has a moderate (monotonically and linear) correlation with weighted direct citation, which reveals what active researchers are currently working on (Zhao & Strotmann 2008, 2014; Klavans & Boyack 2017). Recall that the interpretation relies on the number of citations in the occurrence matrix (k_i and k_j), implying that it is normalised to the total number of citations instead of maximum overlapping in the network. Nonetheless, from an empirical perspective, this measure emphasises still co-citation more strongly and increases the relevance of bibliographic coupling, which can be interpreted as a predominance for indirect relationships of the academic field rather than emphasising direct ties among authors.

Table 12 Regression Based on Quadratic Assignment Procedure

Dependent Network	Predictor	Weighted Jaccard		Cosine/ Ochiai		Association Strength	
		B ($SE\ B$)	β	B ($SE\ B$)	β	B ($SE\ B$)	β
NWDC	Intercept	0.026*** (0.000)	0.000	1.409*** (0.000)	0.000	0.010*** (0.000)	0.000
	Direct Citation	0.887*** (0.010)	0.287	1.186*** (0.014)	0.458	1.148*** (0.002)	0.912
	Bibliographic Coupling	0.980*** (0.008)	0.460	0.622*** (0.021)	0.460	1.183*** (0.010)	0.233
		R^2 =0.515		R^2 =0.443		R^2 =0.911	
NWDC	Intercept	0.026*** (0.000)	0.000	0.409*** (0.000)	0.000	0.010*** (0.000)	0.000
	Direct Citation	1.710*** (0.005)	0.553	0.713*** (0.013)	0.276	0.963*** (0.002)	0.765
	Co-citation	0.977*** (0.003)	0.706	0.816 *** (0.024)	0.684	1.067*** (0.006)	0.364
		R^2 =0.934		R^2 =0.657		R^2 =0.964	
NWDC	Intercept	0.026*** (0.000)	0.000	0.409*** (0.000)	0.000	0.010*** (0.000)	0.000
	Bibliographic Coupling	1.276*** (0.004)	0.599	0.811*** (0.0.13)	0.599	0.947*** (0.013)	0.186
	Co-citation	0.991*** (0.003)	0.716	1.013*** (0.021)	0.850	1.977*** (0.008)	0.673
		R^2 =0.990		R^2 =0.941		R^2 =0.526	

Note: B for the unstandardised beta, $(SE\ B)$ for the standard error of the unstandardised beta, β for standardised beta, and $***\ p < 0.001$. The numbers of draws to use for the quantile estimation are 5,000. Each matrix was log transformed, and $R^2 = R^2_{Adj}$ for all cases. NWDC: *Normalised Weighted Direct Citation*.

When association strength is used to normalise the citation-based measures, the correlation between the normalised weighted direct citation and co-citation is also strong, consistent between the three cases reviewed. Nonetheless, in contrast with the other cases, direct citation increases its relevance with a moderate correlation with the unified measure and co-citation, which was previously considered as the tendency of self-organisation of current literature topics among authors (Klavans & Boyack 2017) or as patterns of developments in particular fields (Garfield et al., 1964; Price, 1965). Association strength measure also corrects the size effect and is less sensitive to the frequency in which the ties occur, capturing more predominantly the walks of the direct citations.

The correlations previously highlighted are further explored in Table 12 using MR-QAP. For different combinations of the citation-based measures considering each normalisation, it is possible to identify the relation of some of the matrices to explain the normalised weighted direct citation. In terms of explained variation, for weighted Jaccard, the use of direct citation or bibliographic coupling with co-citation explain 93% and 99% of the total variation, respectively (indicating the presence of high collinearity between the measures). As previously explored in the correlation, direct citation and bibliographic coupling have a similar contribution when used interchangeably with co-citation. These might suggest that in this particular case, and as we notice for the correlation between the networks (Table 11), direct citation and bibliographic coupling tend to be highly related. The interpretation for this normalisation might be aligned with the tendency of weighted Jaccard to give more prevalence on the intellectual structure of the field combining different temporalities into the representation, slightly giving more emphasis to the tendency of the patterns of new developments expressed in this particular case through direct citation and bibliographic coupling, but in combination with the knowledge baseline of the co-citations.

When cosine or Ochiai index are used to normalise the four measures, the model that explains more variation is when bibliographic coupling and co-citation are used as predictors (R^2 =0.941). In this particular case, the emphasis is given for indirect references instead of direct relations because co-citation assumes that two authors are cited together by later author and bibliographic coupling consider that two authors will appear linked if they are citing the same authors of the same references. However, there is no consideration to the direct citation among the authors, which gives more predominance to the perceived network in this scientific field. The intensity of co-citation is more predominant than bibliographic coupling ($\beta_{stnd} = 0.850, p < 0.001$ and $\beta_{stnd} = 0.599, p < 0.001$ respectively), which bias favours the work of well-established authors.

Finally, the association strength emphasises more in favour of the pattern of development on this field and the arcs of the network through direct citations compared to the perceived ties when the association strength is used as a normalisation. This situation can be noticed through the relevance given to direct citation when bibliographic coupling (R^2 =0.911) or co-citation (R^2 =0.964) are controlled to explain the normalised weighted direct citation. For both models, the intensity of direct citation prevails ($\beta_{stnd} = 0.912, p < 0.001$ and $\beta_{stnd} = 0.765, p < 0.001$ respectively). From this particular case, local concrete links between authors appear to be more relevant than the global intellectual structure perceived by the researchers in this field. Which, for the empirical case considered here, can suggest that researchers might be citing more frequently recent literature.

2.6 Discussion

I have reviewed through an illustrative example the main building blocks (i.e., direct citation, bibliographic coupling, and co-citation) to create the author weighted direct citation. The main point of using these measures together is to complement each other, allowing recovering the different emphasis of time considered in the various citation versions, as was reviewed during the illustrative example, and penalising works out of topics. I disentangle how different walks arise between authors traversing works or how researchers 'stand on the shoulder of other authors', which identifies the patterns of developments of multiple fields as the building block of further analysis. For co-citation, it was considered how the past is emphasised as the knowledge base, and bibliographic coupling, on the contrary, gives more prevalence to active researchers and their current works. As can be noticed, the relevance of time differs in the three cases.

A further distinction is between the socio-cognitive citation network that divided into two different dimensions. The first relies on minimum overlapping to emphasise the *citation networks' cognitive dimension*, while the second used the conventional projections to give more prevalence to the *social dimension*. This distinction is also considered implicitly in the network's normalisations when *Jaccard* and *cosine* are used because they depend on relative overlapping measures. On the contrary, *association strength* compares the observed and expected ties, which is more similar to the social dimension.

In the empirical case reviewed, Jaccard and cosine normalisation emphasise the authors' indirect connections in the field. Simultaneously, the association strength gives more relevance to the direct citations when used in the empirical example. As previously identified, similarities do not come without controversies (e.g., Ahlgren et al., 2003; Leydesdorff, 2008; Egghe & Leydesdorff, 2009; van Eck & Waltman, 2009). As I have argued through an empirical case, the suitable measure can be sensitive to the normalisations used, which in practice does not give much support to the combination of these three measures without further consideration of the normalisation's impact in the first place. The reasons are that *Jaccard*, and *Cosine* rely on overlapping elements, which for this particular case, emphasise more on the role of third parties in the assignation of co-concurrency in which the past citation was more predominant than the citation of active research in both cases. However, *association strength* does not rely on overlapping and instead rely on the strength of the relationships (van Eck & Waltman, 2009), which is consistent in this empirical case giving more relevance to the direct citation in the merged measure.

These representations are often used to conduct further fine-grained analysis identifying mechanisms operating in the network, detecting invisible colleges, among others, and the baseline network is crucial for the exploration of some of these features. While the idea of combining these measures seems appealing, up to now have been normalised using fractional counting (Persson, 2010; Wang et al., 2019), which assumes that each author should be treated equally in the byline hierarchy. Fractional counting and harmonic counting require previous knowledge of the relevance of the positions in different knowledge areas but are theoretically straightforward in their motivations. When this information cannot be assumed, an additional option is to rely on similarities in citations' patterns. In this case, I have explored how the transformation and normalisation of the network change the representation baseline. Specific dimensions are related to each other, having different meanings and becoming empirically sensitive to the normalisation used.

Theoretically, this delimitation might also represent different types of distinctions in the scientific fields under inquiring according to external forces (i.e., other individuals coupling researchers) or as a trace of direct walks in the network (i.e., author citing authors). It might be reasonable that the aforementioned external forces or direct traces operate in the representation of the socio-cognitive network, and the weighted author direct citation can capture some of its common elements. Nevertheless, it is still unclear which normalisation to use and the theoretical explanation to motive the decision. A first attempt can be an interest in emphasising the fields' external elements as the common interest in the shared cognitive dimension according to the community's perception or the authors' communicational trace. In both cases, the field's shared understanding seems to have a relevant role in justifying the weighted normalised direct citation. There is much more to learn about the combination of different measures, e.g., using different strategies that capture the *backbones* of the network, the combination of fractional/harmonic counting, other normalisation measures, and/or the simultaneously uses of works and authors. And further research should be done to identify differences in terms of short and long-term manifestations of the citation ties. More research is needed to explore the potential consequences of merging these measures.

Acknowledgments

The author would like to thank members of the Mitchell Centre for Social Network Analysis at the University of Manchester for helpful comments and suggestions that considerably improved this work. Additionally, he acknowledges the assistance given by Daniel Perez from Epistemonikos Foundation in the data collection.

Financial Support

The study was supported by the National Commission for Scientific and Technological Research in Chile (ANID PFCHA/DOCTORADO BECAS CHILE/2017, grant number 72180242).

Conflict of interest

The authors have nothing to disclose.

Chapter 3

Co-evolution of a Multilevel Scientific Network: Extended Features for Goodness of Fit in Three-Mode Networks

Alejandro Espinosa-Rada*, Elisa Bellotti*, Martin Everett*, Christoph Stadtfeld†

**The Mitchell Centre for Social Network Analysis/Sociology, the University of Manchester*

†Social Networks Lab, ETH Zürich

Abstract

This paper aims to understand how a group of academics create interpersonal intercitations considering the co-evolution of a multilevel network. To achieve this, it explores the relevance of the closures by affiliation and closures by association mechanisms, expanding some diagnostics to distinguish the contribution of the cross-level effects in the representation of relevant features in a complex three-mode multilevel and multiplex network. This approach uses the stochastic actor-oriented model for one-mode and bipartite networks to link micro-macro processes using a dataset of a scientific community from 2013 to 2015. New and already available measures for diagnostics are used for statistical models for social networks to identify how micro-mechanisms trigger different structures at the macro level. The results suggest that social relationships grounded on scientific collaboration and space proximity based on institutional affiliation are more accurately suited to understand the co-evolution of the networks in a scientific network than cognitive-based networks measured as the similarity in publishing in the same journals.

Keywords

Scientific Networks; Multilevel Networks; Citation Network; Collaboration Network; Multiplex Networks; Network dynamics; Stochastic Actor-Oriented Model; SIENA

3.1 Introduction

Three main areas determine the presence of ties in dynamic social networks (Rivera, Soderstrom and Uzzi, 2010) from the perspective of the network's theory of networks (Borgatti and Halgin, 2011b). The first is dyadic similarity (assortative) mechanisms in which the dynamic process of the network rely on the compatibility and complimentary of the actor attributes. The second is relational mechanisms which capture the importance of direct and indirect connections between the actors. The last mechanisms are based on proximity, in which the source of the network is at the level of physical and cultural environments. In this paper, proximity mechanisms are explored to understand how a group of academics generate interpersonal intercitations considering the co-evolution of a multilevel network.

Few studies use the three types of mechanisms in a dynamic actor-oriented perspective simultaneously, and fewer of them explore the co-evolutionary interdependency of actors and entities of different levels in scientific networks. None of them considers interpersonal *intercitation* contexts (White et al., 2004; White, 2011; Milard, 2014) considered as the tendency of creating 'in house' citations within a fixed setting – as a set of authors that share a similar context (Chubin & Studer, 1979; Schrum & Mullins, 1988). In previous studies on scientific networks, dyadic similarity and relational mechanisms are often used (Ferligoj et al., 2015; Kronegger, 2012; Zinilli, 2016; Stark et al., 2020). Purwitasari et al. (2020) also studied the relationship between author and topics as part of the cultural environment of researchers using a longitudinal scientific network controlling for the other mechanisms. While there are a few studies available, this approach is still narrow in the context of scientific networks, and none of them conducts diagnostics to distinguish the contribution of the cross-level mechanisms as meso-level social forces in the representation of relevant features for three-mode multilevel networks at the level of the system as a whole. As far as we are aware, there are no studies that use statistical diagnostics to distinguish how well the micro-level represents macro features at the network level in a three-mode multilevel and multiplex networks.

Micro-level and macro-level features can be distinguished using the different types of mechanisms, in which the micro-processes are the local unit of analysis responsible for the emergence of the network, and the macro-processes correspond to the properties of the network as a whole (Robins et al., 2005; Snijders & Steglich, 2015; Stadtfeld, 2018) in this case. Simulations are often used to explore the linkage between the micro and macro level in a dynamic network. The simulation replicates the observed system constrained by the estimated parameters corresponding to specific statistics in a random network using a Markov Chain Monte Carlos approximation, thus creating a sample of possible networks. Then, these simulated samples of networks are contrasted with the observed network to replicate some substantive features of the network itself. Different diagnostics are then expanded and conducted to identify how the micro-mechanisms represent additional macro features, allowing to identify whether the inclusion of specific mechanisms results in better models. Some of the diagnostics that are considered in this study are for dyadic similarity types of mechanisms, such as an E-I index distribution, Yules Q and IQV index for the same covariates, and a Euclidian distribution for similar covariates. Some measures considered are in line with the study of social capital (Burt, 1992), and we propose an overlapping multiplex triadic census as a diagnostic of relational mechanisms for parallel or multiplex networks of the same group of actors. Additional fits are explored for proximity-based mechanisms for two-mode and three-mode multilevel networks, such as mixed degree distributions, the mixed geodesic distance distribution, and a mixed quadrilateral census.

As a case study, the Chilean astronomers are explored as a unique scientific community suitable to derive complex structures, which is also a relatively small group for the analysis. This community has a national scientific discipline institutionally bounded with privileged access to some of the most relevant telescopes in the world. The years analysed corresponds to the period of formation of this astronomical community after the arrival of the *Atacama Large Millimeter/submillimeter Array* (ALMA), which is currently the biggest radio astronomical facility earth bounded.

The following sections explore the co-evolution of multilevel networks to identify why scientists are attracted to create interpersonal intercitations in a scientific group. The theoretical background of the research is first presented to distinguish the different tie-based mechanisms in the context of the study of scientific networks. Some extensions of the goodness of fit are then proposed for dyadic similarity-based mechanisms, relational-based mechanisms and proximity-based mechanisms. Next, the data and the methodology used for the analysis are described. The measurement and relevance of the closure by affiliation and closures by association mechanisms are presented as two proximity-based mechanisms. Finally, the usefulness of the micro-mechanisms is investigated using the case study to explore the suggested diagnostics is demonstrated. This article concluded with a discussion of using a dynamic multilevel approach in the analysis of scientific networks.

3.1 Theoretical background

3.1.1 Mechanisms for the sociological study of scientific networks

Dyadic similarity-based mechanisms are mostly a pairwise phenomenon studied as actors' tendency to create ties with other actors with similar or different social attributes (Lazarsfeld & Merton, 1954; McPherson et al., 2001). In the sociological study of science and knowledge, different types of social characteristics are either ascribed (e.g., gender, age, nationality and ethnicity) or acquired and, therefore, accumulated (e.g., academic hierarchy, citations and resources) (Merton, 1988). The relationship between these attributes was studied in some scientific networks considering the tendency of homophilous ties (the trend of actors in creating relations with others that share similar social characteristics) or heterophilous ties as the contrary tendency. For example, gender homophily is a reliable mechanism for team formation (Ruef et al., 2003; Bear & Woolley, 2011). This homophily tendency in science can create social boundaries. For instance, men tend to develop close social ties with other men, and women scientists tend to lack closeness and reciprocity compared to male–male relationships (Etkowitz et al., 2000; Jha & Welch, 2010). A complementary strategy is the heterophily tendency in science, in which scientists are likely to collaborate with others with different skills, knowledge and know-how, which are complementary. As Moody remarked, it 'is easier to bring in a new author than it is to learn new material oneself' (2004, p. 217).

Other extensively used types of mechanisms in the literature of scientific networks are based on relations. Most of these approaches use macro topologies and big networks to characterise the structures of scientific communities (e.g., Barabási & Albert, 1999; Newman, 2001a; Moody, 2004; Price, 1965; Watts & Strogatz, 1998). And some of these mechanisms can also be tractable in the sociological study of science and knowledge. For example, because of the unequal distribution of recognition in science, Merton, in collaboration with Zuckerman (1967), emphasises that there is a Matthew Effect (Merton, 1968a, 1988) in which scientists with more peer recognition tend to accumulate even more credit and that scientists with few or even no credit tends to obliterate their tendency to earn the same peer recognition. These patterns have been generalised in the study of networks as the tendency of actors to receive more connections in science, such as citation (Price, 1965) and collaboration (Newman, 2001b). According to Mullins (1972, 1973), the evolution of scientific networks involves social, cognitive, and situational dimensions that allowed the emergence of groups in the form of dyads and triads to understand specialities, in line with recent empirical research (Kronegger et al., 2012; Zinilli, 2016; Stark et al., 2020).

Proximity-based mechanisms have been recently studied in multilevel scientific networks, predominantly in cross-sectional networks (Lazega et al., 2008; Bellotti, 2012; Bellotti et al., 2016a). The primary approach is to study how institutional sharing affiliation creates scientific networks. For example, inter-individual and inter-organisational dependencies in science interact in a joint multilevel approach where the 'dual position' corresponded to a form of status, allowing the identification of the strategies used for individuals to appropriate, accumulate and manage their resources and the ones from their organisation (Lazega et al., 2008). Some previous studies have demonstrated the advantages and disadvantages of being part of big or small institutions in scientific networks (Bellotti, 2012; Lazega et al., 2013; Bellotti et al., 2016a) when there is a joint interdependency within inter-individual and inter-organisational leading to a different pattern of collaboration between scientists. Individuals also have different *foci* in which joint activities are organised, creating clusters and becoming tied interpersonally (Feld, 1981). For example, scientific teams may share the same institutional affiliations, but the spatial proximity encourages informal communication (Katz & Martin, 1997). Actors share the same space of relation in which they also incorporate the cognitive dimension of the organisational forms, where they share the same reference and knowledge space and institutional proximity that constrain their environment (Boschma, 2005).

3.1.2 Micro–macro linkage in dynamic networks

Compared with macro-mechanisms, relatively few approaches use micro-mechanisms in the study of scientific networks. These micro-mechanisms can either be tie-based (conditional probability of a connection given the rest of the network) or actor-based (actor choices to optimise utility based on their outgoing ties) (Block et al., 2019). These micro-mechanisms rely on local configurations (Robins et al., 2005) or actors' actions (Snijders, 2001) that generate the network, thus avoiding direct interpretations of aggregated structures in distributions or features of the entire network. With this distinction, one will be able to prevent ecological fallacies between the analytical levels (Robinson, 1950) in interpreting scientific networks, avoiding set conclusions about the characteristic of one level deduced from the other. This study argues that using the micro-mechanisms gives a wider variety of different processes in creating links that facilitate the interpretation and explanation of the emergence of scientific relationships. It should be noted that the micro-mechanisms can be linked with network macro-mechanisms if researchers are interested in the connections with topologies at the level of the network as a whole.

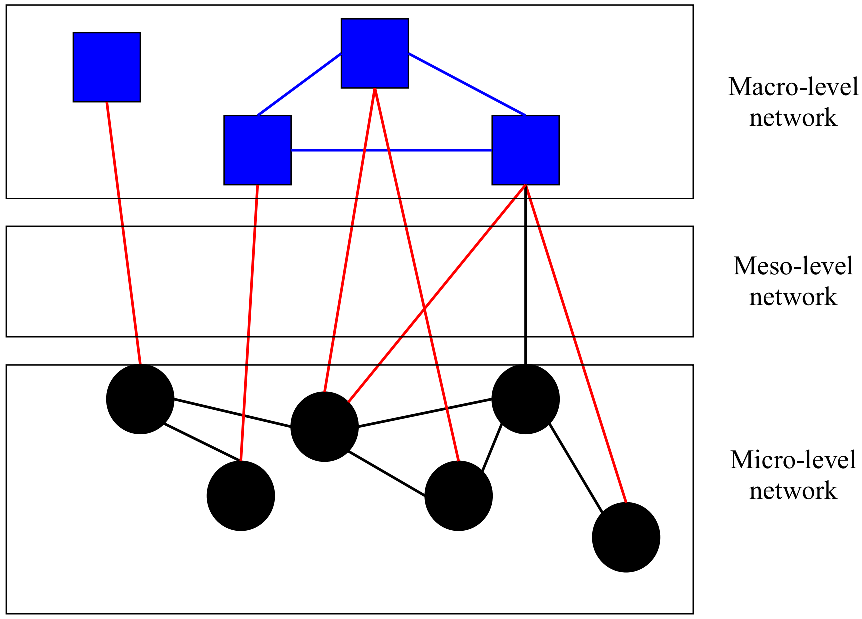


Figure 11 A Multilevel Network

A *multilevel network* is a complex structure in which actors can have ties within and between levels. In this case, each set defines a micro and macro level, respectively. A bipartite network (Breiger, 1974) – or even a tripartite network (Fararo & Doreian, 1984) - can be established between nodes from two adjacent levels as a meso level (Wang et al., 2013; Bellotti et al., 2016a) (Figure 11). This network representation enriches the complexity of the analysis incorporating cross-levels dependency (e.g., Lazega et al., 2008; Bellotti, 2012; Snijders et al., 2013; Wang et al., 2013; Broccatelli et al., 2016), referred in here as *meso-level social forces* to emphasise the role of the meso-level network. Little is known about the processes that generate the system when co-evolved in an interdependent process of multiplex and (tri)bipartite networks.

Another gap in the study of scientific networks is the linkage between these micro-processes that produce structural features at the level of the network considered as a 'macro level'. A formal way of evaluating the micro-macro connection uses statistical goodness of fit for social networks (Snijders & Steglich, 2015; Stadtfeld, 2018). The goodness of fit will compare the observed values with a simulated population of networks that are constrained by estimated parameters to replicate the entire network. This approach has been extensively applied in static (Hunter et al., 2008; Robins et al., 2009) and dynamic networks at the end of the periods (Lospinoso, 2012; Lospinoso & Snijders, 2019). One of the fundamental limitations of using this approach is that it is unclear which auxiliary statistic to use to make the diagnostic. If the model cannot reproduce relevant features, there would be little confidence in the inferences. Substantial deviation from these features would indicate that the model is not a good proxy for generating the processes or the data is not sufficiently accurate, needing further scrutiny (Wang et al., 2020). Therefore, the features not included in the model should be selected using substantive reasons or knowing a priori the measures of interest of the observed network.

There is not always clear which are the reasons to select the feature to evaluate. Hunter et al. (2008) published one of the first papers proposing this goodness of fit, and most of the applications use this reference to justify the selection of the features for the diagnostic. Their paper argues that degree should be included to explore the aggregated structures of interest because of the attention paid in the literature. Secondly, there are some statistics (i.e., curved or alternating parameters) that improve the convergence of specific statistical models, especially for problems of convergence in exponential random graph models (Lusher et al., 2012) and less often in stochastic actor-oriented models (Snijder 2001; Snijders et al., 2010). Thirdly, geodesic distance is relevant in the social network theory to understand the speed and robustness of diffusion across networks. Currently, there are options beyond relational mechanisms (e.g., indegree distribution, outdegree distribution, geodesic distribution, clique census and triad census) for the behavioural and dyadic similarity-based mechanisms (Lospinoso & Snijders, 2019; Wang et al., 2020) and proximity-based mechanisms for two levels (Hollway et al., 2017; Wang et al., 2020).

3.1.3 Features for goodness of fits

This paper presents some alternatives for 'aggregated micro features' rather than 'proper macro' features (Snijders & Steglich, 2015) that do not pretend to be exhaustive options. These alternatives are for the goodness of fit for dyadic similarity-based mechanisms (i.e., E-I index, Yules Q, IQV and dyadic similarity distance-based for reciprocal ties), for relational-based mechanisms of two networks from the same set of actors (i.e., effective size, constraint and two overlapping triadic censuses) and proximity-based mechanisms for two or three-mode multilevel networks (i.e., mixed multilevel degree distribution, mixed multilevel geodesic distribution and mixed multilevel quadrilateral census). Some network models have an actor-oriented perspective which tends to be closer to the social theory (Snijders, 2001; Stadtfeld & Block, 2017; Block et al., 2019). Therefore, this article uses some measures often used to describe ego networks when alter–alter ties are available (Crossley et al., 2015; Perry et al., 2018; McCarty et al., 2019) that could be meaningful for the descriptive analysis of the network.

Two different measures are used for dyadic similarity mechanisms as additional alternatives to edgewise similarity (Lospinoso & Snijders, 2019). The first measure is the E-I index of Krackhardt and Stern (1988) that identify homophily in the network. This index differentiates between actors that share the same covariate (I) compared to the other actors with a different covariate (E).

$$EI = \frac{E - I}{E + I}$$

This measure allows the estimation of an index for categorical variables that oscillates between perfect homophily (-1) and perfect heterophily (+1). Because of its particularity, this measure can be expanded into different types of centrality measures, normalisations and consider other substantive properties (Everett & Borgatti, 2012). For this case, the E-I index is calculated at the actor level for the reciprocal ties. Since the E-I index only looks at the relationships that were formed and not the pool of potential actors, the Yules Q is added to account for this extra information; In which X is the number of non-chosen alters that have the same categories of actor i, and Y is the number of non-chosen actors j that have different categories.

$$Q = \frac{IY - EX}{IY + EX}$$

Another measure often used is Agresti's Index of Qualitative Variation (IQV), which is a normalisation of Blau's H index.

$$IQV = \frac{1 - p_1^2 - p_2^2 - p_3^2 - \cdots - p_r^2}{1 - 1/r}$$

Where r is the number of categories of the covariate v, and p_i is the proportion of ties to actor i. Compared to the E-I index, this measure also has a straightforward interpretation, where the minimum is 0 if all ties are in one category and the maximum is 1 if each category has the same number of connections.

Another measure for dyadic similarity-based mechanisms is a variation of the dyadic similarity distance measure (Lospinoso & Snijders, 2019). The dyadic similarity calculates the outer product between v_i and v_j; where v_i is the attribute of the focal actor, and v_j is the attribute of the alters of ego.

$$d(i,j) = 1 - |v_i v_j^T|$$

Then, the measure is normalised using the range of the numerical covariate for this dyadic similarity measure (Lospinoso & Snijders, 2019).

$$norm(d(i,j)) = \frac{d(i,j)}{\max\big(d(i,j)\big) - \min\,(d(i,j))}$$

This similarity measure is 1 if the two actors have the same value and 0 if one has the highest and the other the lowest possible value, which allows for a generalisation of the measure. Dividing by the number of reciprocal ties, the emphasis is on how similar the actor is compared to its reciprocal ties.

For relational mechanisms, useful alternatives are often used to analyse ego networks and structural holes. These measures are the effective size and constraint to estimate social capital, according to Burt (1992). Effective size measures the size of the network of actor i controlling for the redundancy of ties were,

$$\sum_h \left[1 - \sum_h p_{ih} m_{jh}\right]$$

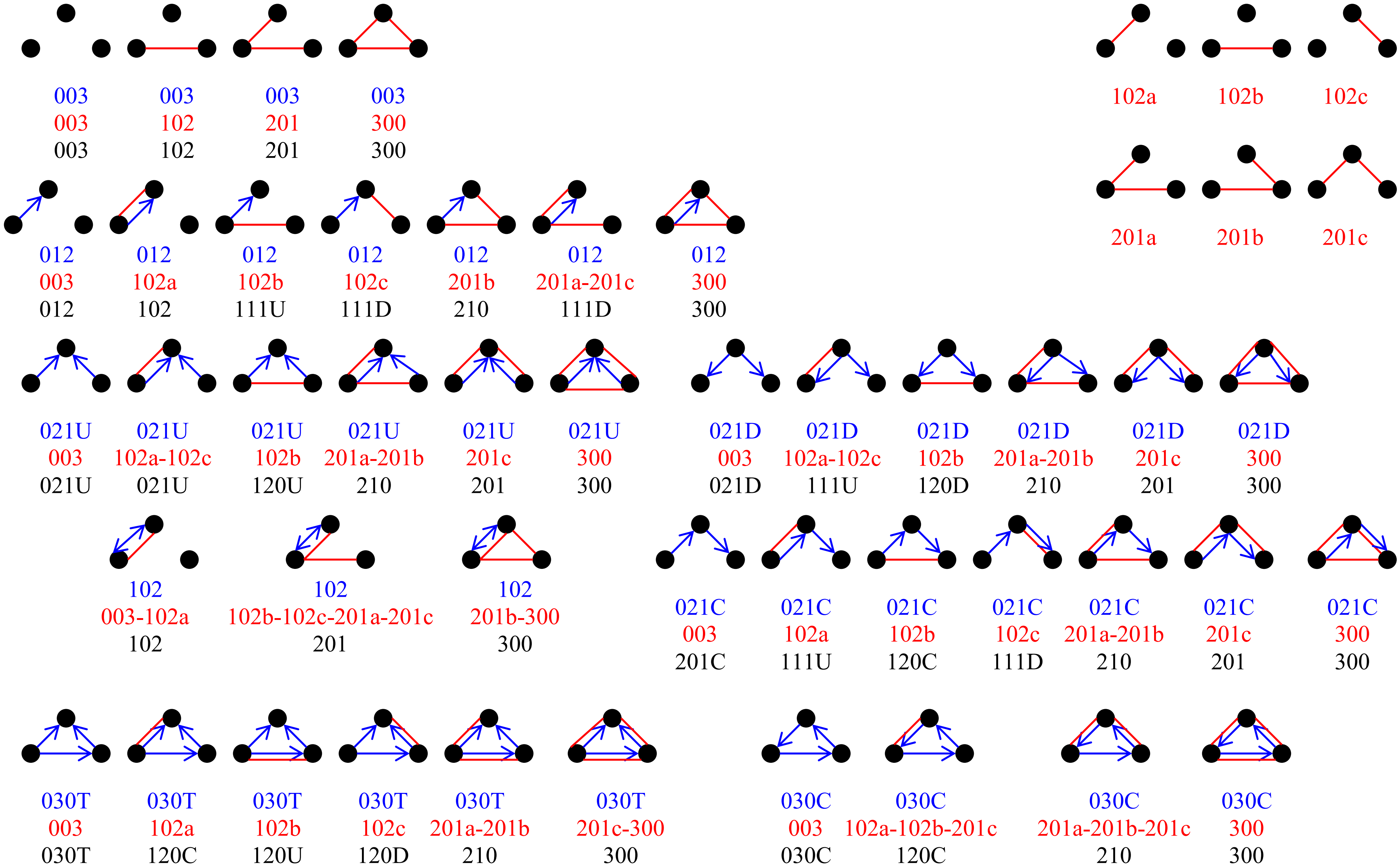
In which $h \neq i, j$, and p_{ih} is the proportion of $i's$ network time and energy invested in the relationship with actor h, and m_{jh} is the marginal strength of the contacts of j in its relationship with h. Dyadic constraint focuses on how an actor j impose structural constraint to the actor i to exchange resources.

$$\left(p_{ij} + \sum_h p_{ih}\, p_{ih}\right)^2$$

Further explanations and details of these measures are in Borgatti (1997) and Everett and Borgatti (2020).

The overlapping triadic census is suggested to study the triadic isomorphic classes of a multiplex network, calculating the triadic census (Davis & Leinhardt, 1972) of two (or more) overlapping matrices. The two matrices could be directed ($d_{ij} \neq d_{ji}$), undirected ($d_{ij} = d_{ji}$), or one of each. Still, adding undirected and directed matrices overrepresented specific configurations. This overrepresentation occurs because undirected networks are restricted to four triadic isomorphic classes (labelled as 003, 102, 201 and 300 in the MAN convention [Holland & Leinhardt, 1976]) while directed networks have 16.

A second alternative to study the overlapping of two networks is the mixed multiplex triad census (Figure 12). In the following, only the possible classes of triads in an undirected network (red) in certain triads of the directed network (in blue) are presented for simplicity. The corresponding triad of the overlapping triadic census (black) is given, which reduce most of the possibilities in the form of the MAN convention.


003 003 003
003 102 102
003 201 201
003 300 300
102a
102b
102c
012 003 012
012 102a 102
012 102b 111U
012 102c 111D
012 201b 210
012 201a-201c 111D
012 300 300
201a
201b
201c
021U 003 021U
021U 102a-102c 021U
021U 102b 120U
021U 201a-201b 210
021U 201c 201
021U 300 300
021D 003 021D
021D 102a-102c 111U
021D 102b 120D
021D 201a-201b 210
021D 201c 201
021D 300 300
102 003-102a 102
102 102b-102c-201a-201c 201
102 201b-300 300
021C 003 201C
021C 102a 111U
021C 102b 120C
021C 102c 111D
021C 201a-201b 210
021C 201c 201
021C 300 300
030T 003 030T
030T 102a 120C
030T 102b 120U
030T 102c 120D
030T 201a-201b 210
030T 201c-300 300
030C 003 030C
030C 102a-102b-201c 120C
030C 201a-201b-201c 210
030C 300 300

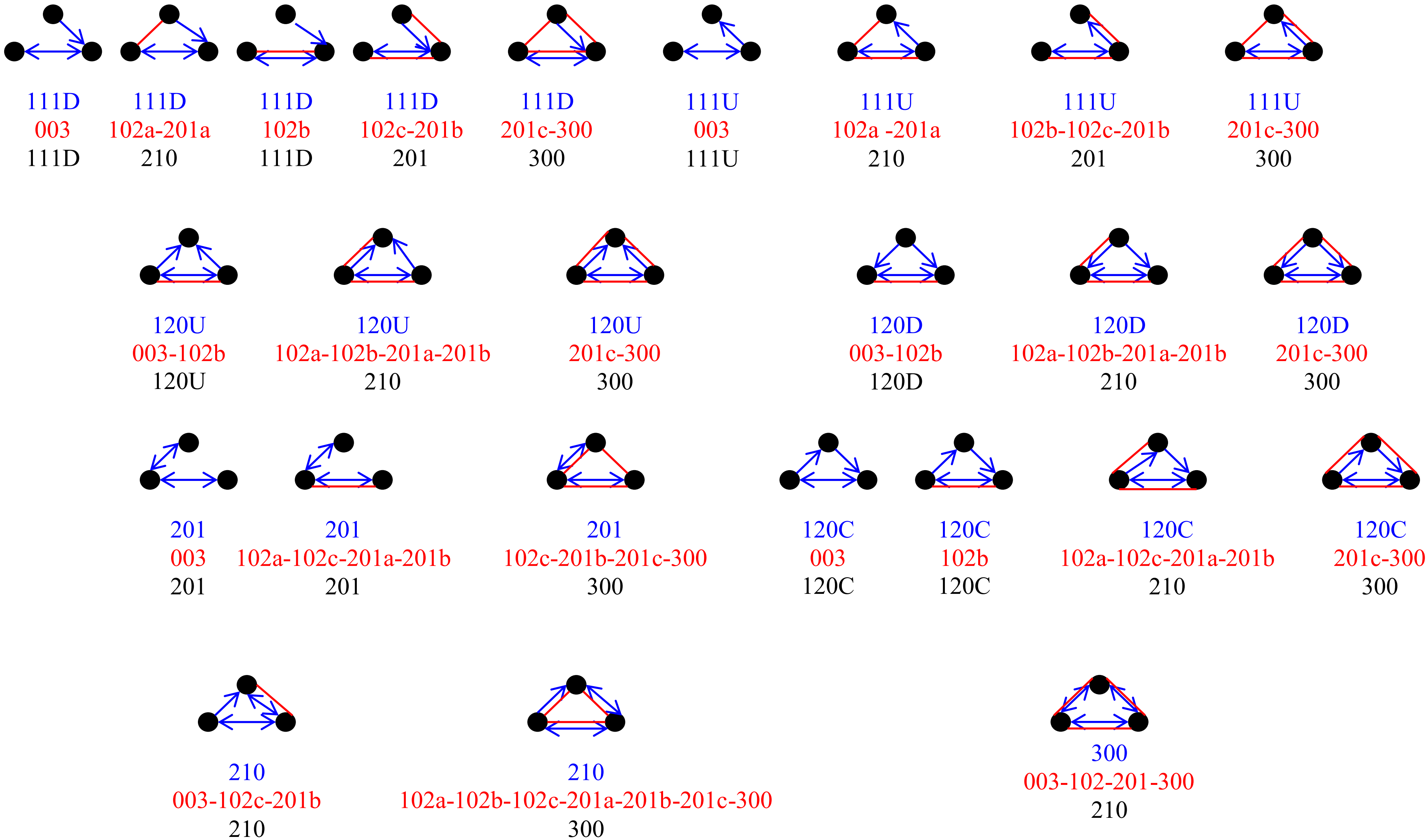


Figure 12 Mixed Multiplex Triad Census for Two One-mode Networks

Some measures are expanded to achieve multilevel features for proximity-based mechanisms. This line of research has been widely used in the study of two-mode networks (Breiger, 1974; Borgatti & Everett, 1997; Latapy et al., 2008) and three-mode networks (Fararo & Doreian, 1984). Recently, some multilevel features have been suggested for static (Lazega, 2008) and dynamic networks (Everett et al., 2018). And, there is a new extension of the goodness of fit for bipartite networks using stochastic actor-oriented models that distinguish between 'belonging' and 'not belonging' to entities of second levels (Wang et al., 2020).

Certain extensions are proposed for two-mode and three-mode multilevel networks. One of them is the mixed multilevel geodesic distribution, that can be calculated using a meta-matrix (Krackhardt & Carley, 1998; Carley, 2002) (reachability could be another option). The network should be jointly represented in a common structure to identify the geodesic distances between the actors of the different levels. For example, as a bipartite network, the actor level with a higher level ($A = n\,x\,m$) can be represented. A second subgraph can be the actor level with the lower level ($B = n\,x\,k$) also creating a two-mode network. Finally, an actor matrix as a directed relationship is possible ($D = n\,x\,n$, where $d_{ij} \neq d_{ji}$). Where n is the mode of individuals, m is the mode of a 'higher' entity, and k is a third mode of a 'lower' entity[31]. Such type of structure can be represented using a join matrix for the analysis of a multilevel network Ω_D, in which, in this case, the links between the higher and lower levels are restricted for simplicity considering the directed network, such as,

$$\Omega_D = \begin{pmatrix} 0 & A^T & 0 \\ A & D & B^T \\ 0 & B & 0 \end{pmatrix}$$

[31] Note that the levels are analytically delimited and potentially interchangeable as 'higher' or 'lower' (examples of entities could be organisations, departments, groups, topics, technologies, papers, and semantics), in which it is assumed that one of the entities has a relationship with the other if actors intermediate them.

Another distribution that uses the matrix of Ω_D is the mixed multilevel degree distribution (an alternative to this distribution could be the join distribution of indegree and outdegree). For this measure, the Borgatti and Everett (1997) extension of Freeman (1978) is used to normalise the degree centralities to analyse bipartite networks. Here, each mode can have a total of n or m degree according to the opposite sets for each mode. For this reason, it is considered the number of connections for A or B in relation to their opposite set,

$$2mNDeg = \frac{d(n_i)}{m}$$

Where $2mNDeg$ is the normalised degree of the actor level in comparison with a 'higher' or 'lower' level (i.e., $m = k$ for the equation). By expanding the measure of Borgatti and Everett (1997), in this case, one of the levels can freely interact within and between the levels. For these reasons, the 'higher' and 'lower' levels have the same properties as Borgatti and Everett (1997) measures for the bipartite network, in which each level is normalised using as the denominator of the degree the actor level (n) (i.e., the opposite level in a bipartite network that can be either outdegree or indegree). The actor level can interact with all the levels. For these reasons, an equivalent extension for a directed network is assuming normalisation for the degree of the actors ($D_D(n_i)$), considering the possible connections within its level and between the other levels[32].

$$\frac{D_D(n_i)}{2(n-1)+m+k}$$

Finally, the mixed multilevel quadrilateral census is used as a direct expansion from the triadic census of Hollway, Lomi, Pallotti and Stadtfeld (2017) (Figure 13). Rather than using the original terminology, this census has the form of a diamond

[32] The notation in Borgatti and Everett (1997) is n_o for nodes' own vertex set, and n_i is the size of the other set. Here, n is the actor level, k is the 'lower' level, and m is the 'higher' level.

(i.e., complete subgraph without one edge). It also has broader features because it has more connections between the actors in comparison with the triadic motif. In the quadrilateral census, there are situations in which different configurations tend to have the same nomination but are different (e.g., 01D1, 01U1, 11D1, 11U1, 01D1, 01U1, 12D1 and 12U1). The nomination of Hollway et al. (2017) will be maintained for comparison reasons. When the configuration has the same label, to distinguish between them, *P* is added in the label when there is one path between 'higher' and 'lower' levels. And *W* is considered if there is a walk between the nodes. For cases 101, *N* is added on the configuration that is not connected.

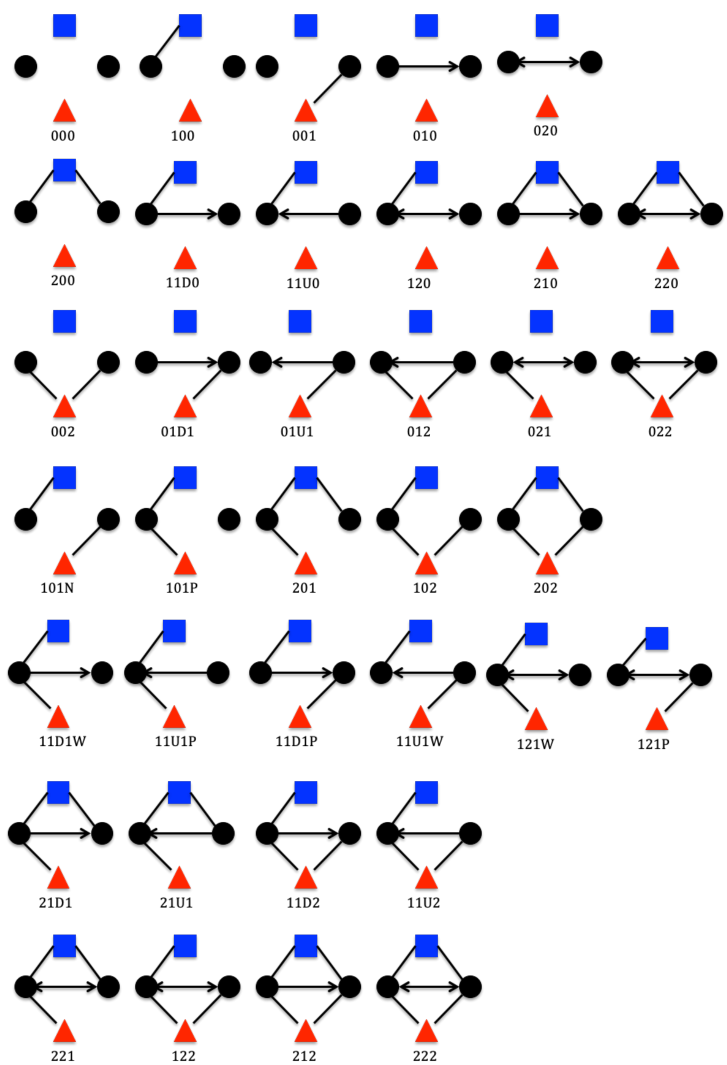


Figure 13 Quadrilateral Census in a Multilevel Network

There are many potential features for the goodness of fit, and theoretical reasons should motivate their selection. These features allow the combination of statistical network analysis with a descriptive analysis of social networks. As far as we know, most of the studies using dyadic similarity-based mechanisms in statistical network models take for granted the correct representation of homophilous processes at the macro level. This article offers a list of potential alternatives to be implemented in statistical models. Similarly, the mixed multilevel structure allows the connecting of the micro–macro levels for complex networks (e.g., bipartite, tripartite, multiplex, or multilevel), allowing the identification of potential features that are currently not well represented in the available goodness of fits.

3.3 Data

This study aims to understand how academics create interpersonal *intercitations* considering the co-evolution of a multilevel network. To conduct the exploration, use proximity-based mechanisms and the diagnostic method presented in the previous section in the context of a complex three-mode, multilevel and multiplex scientific network on a relatively small scientific community.

	Description
Networks	
Citation network	Direct network of scientists citing other scientists in this national discipline
Collaboration network	Undirected network of scientists collaborating with other scientists in this national discipline
Institutional affiliation network	Bipartite network of authors and their institutional affiliation
Publishing in the Web of Science network	Bipartite network of authors connected with journals from the Web of Science
Covariates	
Foreigner	Nationality of the astronomers
Age	First paper published in the Web of Science
Citations	Accumulated number of citations of the astronomer for each year in the Web of Science
Publications	Accumulated number of publications of the astronomer for each year in the Web of Science
Type of organisation	The organisations could be universities, research centres or astronomical observatories
Year of institutions	Year in which the institution consolidated a formal group of astronomers, department or equivalent held in universities
Impact factor	Impact factor of the journal citation reports in the Web of Science in the last five years
Astronomical journals	Journals categorised as 'astronomy and astrophysics' in the Web of Science
Interdisciplinary journals	Journals that are classified as interdisciplinary in the Web of Science
National-based journals	Journals that are regional based

Table 13 Summary of the Data Used in the Analysis

The astronomical and astrophysics community in Chile has a strong presence in local science policy (Espinosa-Rada et al., 2019). And the national dimension is relevant to have access to the telescopes in this country, in which national committees allocate the time observation (e.g., The Chilean Telescope Time Allocation Committee, CNTAC). Chilean astronomers tend to apply to the national resources for science (National Commission for Scientific and Technological Research, CONICYT) because universities require that scientists apply to those funding for institutional accreditation. The funding mix local and foreign resources for the local development of this community (e.g., ESO–Government of Chile Joint Committee funded projects, ALMA-ANID, Gemini-ANID, QUIMAL, among others).

Data collection took place between May and June 2014, and data was corrected, updated, and expanded until October 2019. The data corresponds to the local group formation (Mullins, 1972, 1973) period of astronomers and astrophysics a few years after the recent arrival of the Atacama Large Millimeter/submillimeter Array (ALMA) in 2011 full operative in 2013 – the biggest radio astronomical observatory worldwide of that time. For the data collection, a list was created of all relevant researchers and university departments, including research institutes that host astronomer and astrophysics academics in Chile during 2014 with access to 10 per cent of the astronomical facilities' observation time settle in the country. This percentage is not trivial because the Chilean astronomical community held in its territory some of the most relevant astronomical infrastructures (such as VLT, the Magellan Telescopes, ALMA and soon the E-ELT and GMA) and will have 10 per cent of the LSST computer cluster in 2023. That represents near 70 per cent of the entire infrastructure on earth.

Overall, 6,008 documents were recorded for the 87 astronomers in ten Chilean Institutions from 1971 to 2017. WOS has a well-developed database actively used in scientometrics studies. Table 13 summarises the data used in the following analysis – data collection, descriptive information and change statistics in the appendix (section C and D).

For the forthcoming analysis, data from 2013 until 2015 is used. This was the period in which the data was gathered initially for this cohort of astronomers and astrophysicists.

3.4 Method

The evolution of the intercitation network is modelled using the stochastic actor-oriented models (SAOMs) (Snijders, 2001, 2005, 2017; Snijders et al., 2010), implemented through the RSiena package (Ripley et al., 2021). This approach is suitable for the analysis because it considers the co-evolution of the undirected, directed, and bipartite networks. The SAOMs model is being used as a model that can analyse interdependent processes expressing a multilevel approach with coupled outcomes that have their timing rate. This approach is known in the literature as an 'analysis of multilevel network' (AMN) in comparison with the 'multilevel analysis of networks' (MAN) (Snijders, 2016). This study concentrates on the first case. Below, a brief description of the model is provided. More detail and technical explanation are elsewhere, for one-mode networks (Snijders 2001, 2005, 2017; Snijders et al., 2010) and two-mode networks (Koskinen & Edling, 2012; Snijders et al., 2013).

Four networks are analysed in this study. The first network corresponds to a directed network of actors citing other scientists. The second network is an undirected network of actors co-authoring papers as a proxy of collaboration within the actors. For these cases, all researchers have at least one publication at the beginning of the period. A third network is the institutional affiliation of the scientists (controlled using structural zeros for institutions that were not formed in the observed period; additional information in the Appendix, Section D). The last network is the journal network in which the scientists are publishing during the period under analysis. In this case, because all actors, institutions and journals were available in the time under study, there are no structural zeros for these cases.

The SAOMs analysis assumes that there is an unobservable continuous time measured in discrete observations. Through simulation, the continuous time is estimated as a Markov process in which actors optimise their 'micro-steps' in the short term as outgoing ties across the networks (creating, dropping, or maintaining their network ties). For this case, the models would assume that actors control their outgoing aggregated citation and collaboration ties and their affiliation, or the events of publishing in a scientific journal that can be changed at any point in time. The *rates* indicate the expected average number of change opportunities for each actor in each period as the difference in the network's speed between actors.

The interpretation should assume some regularity in the network evolution from time t_{m-1} to t_m, in which the actors are deciding whether they will create the aggregate ties to an actor (authors, institutions or journals), maintain the relationship, or drop the connection in the considered year of this relatively short period. For the case of the collaboration network, the model proposed by Snijders and Pickup (2017) is followed, in which the actors negotiate their collaboration following a 'one-sided initiative with reciprocal confirmation'. In this model, when a scholar has the opportunity to consider a change in the collaboration tie, they would prefer changing the tie that is most satisfactory to both sides, considering that when this tie is created, the other actor has to agree.

Different linear 'evaluation functions' are estimated to control for the utility of an actor when they decide to change their ties as the micro regularities in the network evolution. Defined below are the characteristics to which the actors seem to be attracted (Snijders et al., 2013).

The first evaluation function shows which change in the network an actor decides to realise concerning their institutional affiliation or in the journal that they published, and this is expressed as an 'objective function' for bipartite networks:

$$f_i^Y(x, y) = \sum_m \beta_m^Y s_{im}^Y(x, y) + U_i(t, x, j)$$

The second is defined as the 'objective function' for the citation and collaboration network in which the actor decides to change a tie as:

$$f_i^X(x, y) = \sum_m \beta_m^X s_{im}^X(x, y) + U_i(t, x, j)$$

The functions f_i^Y and f_i^X control the evaluation ('objective function') of actor i's and their current state according to the two-mode networks (Y) and the one-mode network (X). The statistical $S_{im}^{Y,X}$ captures the different effects. Each of these effects describes a process of the evolution of the network. Some of these effects are structurally endogenous (e.g., transitivity or out-degrees) or exogenous covariates (e.g., nationality or accumulation of citations), estimated by the weighted parameter β_m from the data estimated through the method of moments (Snijders, 2001). In each equation is also considered a random variable, indicating the part of the actor's preference that is not represented in the systematic component $f_i(\beta, x^{i\rightsquigarrow j})$ or $f_i(\beta, y^{i\rightsquigarrow j})$, respectively.

Each actor has a multinomial choice probability in the evaluation function where an actor i makes a particular tie change in the network. As in the notation of Stadtfeld et al. (2016), it is considered that $x^{i\rightsquigarrow j}$ as the network x after actor i changes the tie to j (created or dropped), and $x^{i\rightsquigarrow i} = x$ as the tie maintained. The same is applicable for $y^{i\rightsquigarrow k}$, with $\mathcal{R}_X$ and $\mathcal{R}_Y$ as the set of 'receiving' actors in the network (citation or collaboration network X, institutional affiliation, or publication in journals in Y). The probability of actor i making an outgoing tie change in the network x or y is defined as:

$$P(x^{i\rightsquigarrow j}\ in\ x) = \frac{exp(f_i^X(x^{i\rightsquigarrow j}, y))}{\sum_{h\in\mathcal{R}_X} \exp\left(f_i^X(x^{i\rightsquigarrow h}, y)\right)}$$

$$P(y^{i\rightsquigarrow k}\ in\ y) = \frac{exp(f_i^Y(x, y^{i\rightsquigarrow k}))}{\sum_{l\in\mathcal{R}_Y} \exp\left(f_i^Y(x, y^{i\rightsquigarrow l})\right)}$$

As an interdependent process, X depends on Y and vice versa. The SAOM model is a reasonable model to analyse the dynamic multilevel structure of this community.

There is a recent discussion about the accuracy of different dynamic networks (Block et al., 2018, 2019; Leifeld & Cranmer, 2019) in which the two main approaches of the controversy are TERGMs (Hanneke et al., 2010; Desmarais & Cranmer, 2012; Leifeld & Cranmer, 2019; Robins & Pattison, 2001) and SAOMs. For the 'ERGM family' (including LERGM [Koskinen & Lomi, 2013; Snijders & Koskinen, 2013; Koskinen et al., 2015] and StERGM [Krivitsky & Handcock, 2014] for dynamic alternatives), there are some extensions for 'multiplex', 'multi-layer', 'multi-relational', or 'multilevel' networks (e.g., Krivitsky, 2012; Wang et al., 2013; Krivitsky et al., 2020). These 'tie-based' models have not been extended yet in temporal networks for the co-evolution of different networks. Because this study is interested in how the network evolves based on a multilevel and multiplex network, the analysis is restricted to SAOMs.

Further details, explanations of the relational and covariates effects and modelling specifications SAOMs models can be found in Ripley et al. (2021).

3.5 Measurement

3.5.1 Dependent variables

The first network to explore is the citation network as an approximate measure of social processes that are considered as social and intellectual at the same time (Crane, 1972; Chubin, 1976; White, 2011). A citation is a proxy estimated from the 'formal channel' of communications used through bibliometric information (Schrum & Mullins, 1988; Zuccala, 2006) in which the act of citing is according to the *oeuvres* of the researchers (White, 2011). Citing can be measured as a directed network when one scientist decides to quote someone else assuming interpersonal *intercitation* contexts (Lievrouw, 1989; Schrum & Mullins, 1988; White et al., 2004; Milard, 2014). The matrix is dichotomised, and the diagonal is assumed to be zero (additional exploratory analysis using weighted network in the Appendix, Section G and H). Another network often used to analyse science is scientific collaboration, where co-authoring is used as a proxy (Mullins, 1972; Chubin & Studer, 1979; Moody, 2004) as an undirected network. A third network is the bipartite network of scientists affiliated in institutions, considering that the common social settings are relevant as a context to explore multilevel mechanisms (Mullins et al., 1977; Lazega et al., 2008). The bipartite network of scientists publishing in scientific journals is used[33], in which the visibility of the journal contributes to the prominence of the positions of the researchers within small groups (Burt & Doreian, 1982).

3.5.2 Explanatory variables

All the covariates in the analysis are centred, and Section E of the Appendix summarises the type of effects included in the model considering a graph representation and the mathematical expression. The micro-mechanisms used for the analysis are briefly explained using the three types of mechanisms reviewed before.

For relational-based mechanisms, some default effects are used, such as reciprocity, for the citation network, as a measure of 'awareness' between researchers (Breiger, 1976) and for the consideration of 'in-house' relationships (Chubin & Studer, 1979; Schrum & Mullins, 1988), and density for the fourth networks because of its relevance in social contexts. The triadic triplets and transitive ties effects are used to control local clustering processes (Mullins, 1972, 1973) in the citation and the collaboration network. As a local hierarchical measure in the citation network, the transitive reciprocal triplet effect is used (Block, 2015). Recall that reciprocity and transitivity are not directly applicable for bipartite networks because ties within actors are not possible. The four-cycle effect is incorporated as a structural equivalent measure for the journal network (interpreted as the disposition of a portfolio of journals and a social position measure [Burt & Doreian, 1982]). Scientists are rarely affiliated with two or more organisations simultaneously, and the four-cycle effect would not be used for this network.

Different effects based on the degree are used to control for the Matthew effects and peer recognition (Zuckerman, 1967; Merton, 1968a). These measurements are the indegree effects that capture the Matthew effect directly (Snijders, 2011) in the citation network and the journal network. Indegrees often adopted a skewed distribution (Price, 1965), and this effect is added to control for the speed of the

[33] Journals are used instead of other semantic networks due that venues are often more institutionally stable.

changes of citations. A degree measure is used in the collaboration network to control the researchers' tendency to have more co-authors (Newman, 2004). For the network of universities, the degree represents the size of the institution (can be interpreted as big or small ponds). Outdegree measures are used to identify the participation of the actors in each network. For the citation network, this represents the tendency to cite other colleagues of the national group as an 'in-house' tendency (Chubin & Studer, 1979). This can be interpreted as the tendency to publish in the journal network, and in the institutional affiliation network, this represents the number of institutional affiliation that each scientist has. For the citation network and the institutional affiliation, the assortativity mechanism is also added, describing the tendency of actors to send more ties to other actors that have higher indegrees, which is reasonable in scientific networks to represent the reinforcement of active actors on the Matthew effect (Merton, 1968a) and with more visible positions in small groups (Breiger, 1976; Mullins et al., 1977). In the institutional affiliation network, this tendency is also present at the second level. It is considered that actors affiliated with more institutions are also affiliated with organisations or research centres with comparatively more actors. Because of the tendency of the citation networks to be sparse, the truncated outdegree is controlled.

The dyadic similarity-based mechanisms are measured through homophilous (heterophilous) tendencies. For the ascribe measure, locals and foreigners are distinguished in the astronomical community. As a proxy of nationality, foreigners are considered scientists born or studied their bachelor's degree in a country different from Chile. The year of the first publication is also used as a proxy of age to control for seniority. A negative similarity effect might be an indicator of heterophily or 'status' tendencies between seniors and academics in early careers (younger researchers might cite more often consolidated academics) (Merton, 1988). This measure is subtracted for the wave under study, and if scientists have not published yet, it is set to zero. Acquired attributes use the accumulation of citations as a direct measure of recognition. Even when the citation is not necessarily a good indicator of quality (Mulkay, 1974; Gilbert, 1977; Nicolaisen, 2008), it is considered as a perceived measure of recognition or quality (Price, 1976; Barabási & Albert, 1999; Bianconi & Barabási, 2001), used for hiring, institutional promotion, or the estimation of aggregated indicators such as the H index. The citations are often skewed and not a normally distributed variable, the covariate is transformed using the $\log(v_i + 1)$, as many publications did not receive citations, increasing the number of zeros. The same treatment of the variable is used for the number of accumulated publications (Merton, 1988).

We also control for parallel network effects (*multiplex network*) as a particular type of relational-based mechanisms, referred to as relational multiplex mechanisms. First, we identify whether the citations arise due to the collaboration and if collaboration generates citations. These differences reflect the *inter-citation* dimension driven by social and cognitive networks as a mixed process (White, 2001; White, Wellman & Nazer, 2003; Milard, 2014). The effect of degree and transitivity between the networks is added to distinguish the relevance of cognitive networks driven by concrete interactions, such as collaboration networks. For the degree-based measure, the trend of having more co-authors in the tendency of citing other researchers is controlled. The effect of co-citation (White & Griffith, 1981) of collaborators (when actors are referred together by the focal researcher) is also considered. Other effects incorporated are the tendency to cite an author if two researchers share a joint co-author and the tendency of two actors collaborating to be cited by the same authors. The possibility that two different authors co-cite the same researchers in the citation network as a variation of the co-citation effect is included. These mechanisms are interpreted for the relationship between citation and collaboration networks as a tendency towards specialities (Mullins, 1972, 1973). It is not expected that sharing citations would lead to collaboration; neither is it expected that the collaborators of a cited scholar would lead to a potential co-authorship. For this case, it is believed that social relationships are more appropriate for social closure than intellectual connections.

There is less empirical research for proximity-based measures than the other types of mechanisms, and in this study, we use two different mechanisms to capture the interrelation between levels as a cross-level effect. A different approach is used in the original delimitation of Rivera et al. (2010). Here, it is assumed that rather than inferring the relevance of context in creating ties as a bipartite network, it can be studied directly through a multilevel structure.

The *closure by affiliation* and by *association* mechanisms are used (Lomi & Stadtfeld, 2014). These mechanisms are distinguished for social and cognitive networks, in which for the social network, it is expected that scientists will tend to cite and collaborate if they share an institutional affiliation because of spatial proximity (Chubin & Studer, 1979; Katz & Martin, 1997; Boschma, 2005). The cases in which the ties are also reciprocal for the citation network are also considered. A similar affiliation closure at the level of journals is expected, in which actors that publish in similar journals will tend to cite and collaborate because of cognitive proximity and reciprocate their citations. For the institutional affiliation, the tendency of association closure will not be controlled because these networks are stable, and extended periods will be needed to identify whether the tendency of actors to cite or collaborate with other actors leads to affiliation in the same institutions. The study has limitations for this effect because, by design, its network controls for academics already affiliated with institutions. This study will control the association closure in journals (Burt & Doreian, 1982), as actors' expected tendency to cite and collaborate with academics aspiring to publish in similar journals.

From the empirical research reviewed insofar, it is expected that *scientists prefer social and cognitive multilevel closure processes when they decide to send ties in their scientific networks*.

Some additional control variables are used in the estimated models. First, the impact factor of the journals in the Web of Science in the last five years is used as a measure of scientific recognition. Previous studies in dynamic actor-oriented models for scientific networks (Ferligoj et al., 2015; Kronegger et al., 2012) have used this measure as a dichotomy variable distinguishing between the 25 per cent top journals in the field (quartile 1) in comparison with the rest of the quartiles. This article treats the journal's quality on a different level without dichotomising. Some journals are from 'astronomy and astrophysics, which are interdisciplinary and regional-based (i.e., Slovakia, Mexico, Australia and Japan) that are characteristics controlled in the model. Astronomy is considered an endogenous and isolated area of research (Leydesdorff & Rafols, 2009; Jansen et al., 2010). It is expected to positively affect the category of astronomy and astrophysics and a negative effect on its interdisciplinary measure. It is also considered that the regionally based indicator should be negative since the discipline might be globally oriented. Finally, the type of institution in the affiliation network (i.e., universities, research centres and astronomical observatories) is also included, in which all actors are affiliated with at least one university.

3.6 Results

Model 1 is a baseline model that includes structural and covariates effects without considering the multiplex or multilevel effects. The networks can be interpreted as independent change models (Stadtfeld & Lomi, 2016). Model 2 includes multiplex effects and the parameters of model 1. The estimation indicates whether there is a tendency to increase the likelihood of occurrence of a tie in the other network. Model 3 includes multilevel effects and the parameters of model 1. The estimation also indicates whether there is a tendency to increase the likelihood of occurrence of a tie between networks of different levels. Model 4 is the full model that adds the baseline model, the multiplex effects, and multilevel effects together. The convergence *t-ratio* for all reported statistics in Table 14 are smaller than 0.1 in absolute values and has an overall maximum convergence < 0.25.

Overall, the four models (Table 14) have similar parameters, and they achieve good convergence in most of the measured features except for the multiplex goodness of fit for the overlapping triadic census (details of the *p*-values for all goodness of fit measured in the Appendix are available in Section F). Models 3 and 4 achieve a reasonable convergence on multiplex features. In the following sections, the parameters of model 4 are interpreted, and the parameters of other models when there are differences in significance according to conventional thresholds in empirical research are highlighted.

Effect	Full Model par.	(s.e.)	Multiplex Model par.	(s.e.)	Multilevel Model par.	(s.e.)	Baseline Model par.	(s.e.)
Citation Network (One-mode Network Dynamic)								
Rate (period 1)	18.13	(3.655)	17.778	(3.639)	16.820	(2.355)	17.059	(2.197)
Rate (period 2)	18.847	(4.361)	19.449	(4.199)	18.132	(2.689)	18.765	(2.671)
Rate indegree	0.048 **	(0.017)	0.048 **	(0.015)	0.020*	(0.009)	0.019 *	(0.008)
RM: Outdegree (density)	–3.344 ***	(0.328)	–3.233 ***	(0.336)	–3.533 ***	(0.379)	–3.313 ***	(0.369)
RM: Reciprocity	0.929 ***	(0.197)	0.936 ***	(0.122)	1.484 ***	(0.187)	1.516 ***	(0.114)
RM: Transitive triplets	0.142 ***	(0.020)	0.144 ***	(0.019)	0.191 ***	(0.029)	0.191 ***	(0.027)
RM: Transitivity reciprocated triplets	–0.127 ***	(0.021)	–0.128 ***	(0.021)	–0.186 ***	(0.027)	–0.186 ***	(0.024)
RM: Transitive ties	1.328 ***	(0.174)	1.363 ***	(0.174)	1.399 ***	(0.170)	1.440 ***	(0.167)
RM: $\sqrt{Indegree}$ (popularity)	0.271 *	(0.108)	0.252 *	(0.107)	0.533 ***	(0.126)	0.500 ***	(0.120)
RM: $\sqrt{Outdegree}$ (popularity)	–0.151 **	(0.048)	–0.142 **	(0.047)	–0.300 ***	(0.055)	–0.293 ***	(0.052)
RM: $\sqrt{Outdegree}$ (activity)	0.553 ***	(0.108)	0.542 ***	(0.108)	0.702 ***	(0.127)	0.667 ***	(0.124)
RM: Outdegree at least one	–2.452***	(0.343)	–2.390 ***	(0.331)	–2.715 ***	(0.383)	–2.572 ***	(0.340)
RM: Assortativity	–0.223***	(0.035)	–0.218 ***	(0.035)	–0.305 ***	(0.042)	–0.288 ***	(0.041)
RM: 4-cycles	0.004 †	(0.002)	0.003 †	(0.002)	0.015 ***	(0.002)	0.014 ***	(0.002)
C: Nationality alter (1=Chilean)	–0.092	(0.063)	–0.056	(0.058)	–0.010	(0.065)	0.033	(0.061)
C: Nationality ego (1=Chilean)	–0.239***	(0.055)	–0.199 ***	(0.052)	–0.176 **	(0.062)	–0.119 *	(0.056)
DSM: Nationality ego x Nationality alter	–0.082	(0.099)	–0.036	(0.095)	0.087	(0.102)	0.157	(0.099)
C: Citations alter	0.204 ***	(0.035)	0.201 ***	(0.034)	0.241 ***	(0.041)	0.234 ***	(0.038)
C: Citations ego	0.025	(0.030)	0.038	(0.027)	0.022	(0.031)	0.030	(0.028)
DSM: Citations accumulated similarity	0.054 *	(0.027)	0.063 *	(0.026)	0.081 **	(0.029)	0.097 ***	(0.028)
DSM: Age similarity (year first paper)	–0.030	(0.021)	–0.032	(0.021)	–0.020	(0.021)	–0.020	(0.021)
DSM: Papers accumulated similarity	–0.093 †	(0.056)	–0.096 †	(0.055)	–0.122 *	(0.061)	–0.132 *	(0.058)
C: Time	0.216 ***	(0.061)	0.202 ***	(0.057)	0.271 **	(0.099)	0.265 **	(0.083)
Iterations	*12,161*		*11,937*		*12,033*		*11,832*	

† p < 0.1; * p < 0.05; ** p < 0.01; *** p < 0.001;

Note: RM: Relational mechanisms; RMM: Relational multilevel mechanisms; DSM: Dyadic similarity mechanisms; C: Control; PM: Proximity mechanisms. To control for time heterogeneity, we add a linear time variable (Lospinoso & Snijders, 2019)

Continuation

Effect								
Citation Network (One-mode Network Dynamic)								
RMM: Collaboration network	1.329 ***	(0.109)	1.340 ***	(0.108)				
RMM: Degree collaboration	–0.066 *	(0.028)	–0.070 **	(0.026)				
RMM: Co-author closure (Collaboration)	–0.129 ***	(0.017)	–0.129 ***	(0.017)				
RMM: Association closure (Collaboration)	0.082 ***	(0.024)	0.082 ***	(0.023)				
RMM: Co-citation closure	0.084 ***	(0.010)	0.082 ***	(0.010)				
PM: Affiliation closure (Institutions)	0.426 ***	(0.123)			0.608 ***	(0.129)		
PM: Affiliation closure (Journals)	0.032	(0.049)			0.100 *	(0.050)		
PM: Reciprocity X Affiliation closure (Institutions)	–0.069	(0.247)			–0.027	(0.242)		
PM: Reciprocity X Affiliation closure (Journals)	0.000	(0.095)			–0.010	(0.095)		
Collaboration Network (One-mode Network Dynamic)								
Rate (period 1)	1.050	(0.168)	1.064	(0.170)	0.951	(0.146)	0.957	(0.149)
Rate (period 2)	1.261	(0.186)	1.292	(0.197)	1.148	(0.160)	1.172	(0.168)
RM: Outdegree (density)	–2.878 ***	(0.649)	–2.429 ***	(0.603)	–2.768 ***	(0.598)	–2.074 ***	(0.545)
RM: Transitive triads	0.362 ***	(0.056)	0.338 ***	(0.052)	0.334 ***	(0.050)	0.311 ***	(0.048)
RM: Transitivity ties	2.075 ***	(0.534)	2.138 ***	(0.515)	2.342 ***	(0.520)	2.426 ***	(0.512)
RM: Degree	–0.300 ***	(0.087)	–0.303 ***	(0.086)	–0.268 ***	(0.077)	–0.275 ***	(0.071)
C: Nationality (1=Chilean)	0.192	(0.197)	0.269	(0.179)	–0.082	(0.165)	0.003	(0.146)
DSM: Nationality ego x Nationality alter	–0.063	(0.445)	0.138	(0.427)	–0.071	(0.410)	0.103	(0.383)
C: Citations	0.107	(0.085)	0.139 †	(0.082)	0.189 *	(0.078)	0.219 **	(0.071)
DSM: Citations *accumulated similarity*[2]	–0.024	(0.030)	–0.034	(0.030	–0.018	(0.027)	–0.028	(0.026)
RMM: Citation network	2.539 ***	(0.724)	2.631 ***	(0.685)				
PM: Affiliation closure (Institutions)	1.274 **	(0.393)			1.596 ***	(0.378)		
PM: Affiliation closure (Journals)	0.075	(0.172)			0.168	(0.150)		
Iterations	*12,161*		*11,937*		*12,033*		*11,832*	

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;

Note: RM: Relational mechanisms; RMM: Relational multilevel mechanisms; DSM: Dyadic similarity mechanisms; C: Control; PM: Proximity mechanisms. To control for time heterogeneity, we add a linear time variable (Lospinoso & Snijders, 2019)

Continuation

Effect								
Institutional Affiliation in Institutions (Bipartite Network Dynamic)								
Rate (period 1)	0.243	(0.062)	0.243	(0.061)	0.242	(0.061)	0.242	(0.061)
Rate (period 2)	4.654	(0.628)	4.669	(0.632)	4.658	(0.658)	4.641	(0.664)
RM: Outdegree (density)	–0.988 *	(0.466)	–0.983 *	(0.489)	–0.989 *	(0.485)	–0.996 *	(0.469)
RM: Indegree (popularity)	0.113 **	(0.040)	0.114 **	(0.043)	0.113 **	(0.042)	0.113 **	(0.039)
RM: Outdegree (activity)	0.358 *	(0.158)	0.360 *	(0.161)	0.359 *	(0.161)	0.357 *	(0.156)
RM: Assortativity	–0.361 †	(0.191)	–0.364 †	(0.198)	–0.362 †	(0.196)	–0.359 †	(0.189)
C: Type of Organisation (1=University)	0.042	(0.185)	0.040	(0.182)	0.042	(0.186)	0.040	(0.180)
DSM: Citations accumulated similarity	2.103 *	(0.841)	2.110 *	(0.872)	2.111 *	(0.843)	2.110 *	(0.844)
Journals in the Web of Science (Bipartite Network Dynamics)								
Rate (period 1)	4.095	(0.520)	4.070	(0.494)	4.093	(0.503)	4.068	(0.497)
Rate (period 2)	3.896	(0.520)	3.879	(0.486)	3.889	(0.508)	3.870	(0.498)
RM: Outdegree (density)	–3.541 ***	(0.223)	–3.552 ***	(0.218)	–3.543***	(0.219)	–3.554***	(0.220)
RM: Cycle of fourth	0.015 **	(0.005)	0.015 **	(0.005)	0.015 **	(0.005)	0.015 **	(0.005)
RM: $\sqrt{Indegree}$ (popularity)	0.418 ***	(0.048)	0.424 ***	(0.046)	0.418 ***	(0.047)	0.424 ***	(0.046)
RM: Outdegree (activity)	0.084 **	(0.029)	0.084 **	(0.029)	0.084 **	(0.028)	0.084 **	(0.029)
C: Interdisciplinary Journal (1=Interdisciplinary)	0.015	(0.290)	0.009	(0.292)	0.013	(0.292)	0.014	(0.294)
C: National-based journal (1=National)	–0.309	(0.364)	–0.315	(0.359)	–0.314	(0.352)	–0.304	(0.357)
C: Astronomical journal (1=Astronomy)	0.263 *	(0.128)	0.263 *	(0.128)	0.264 *	(0.127)	0.264 *	(0.130)
C: Impact Factor	0.022	(0.094)	0.024	(0.093)	0.023	(0.094)	0.023	(0.095)
DSM: Citations accumulated similarity	–0.043	(0.446)	–0.032	(0.456)	–0.046	(0.458)	–0.043	(0.446)
RMM: Citation to journal agreement	0.026	(0.029)			0.027	(0.030)		
RMM: Co-authorship to journal agreement	–0.015	(0.024)			–0.017	(0.026)		
Iterations	*12,161*		*11,937*		*12,033*		*11,832*	
† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;								

Note: RM: Relational mechanisms; RMM: Relational multilevel mechanisms; DSM: Dyadic similarity mechanisms; C: Control; PM: Proximity mechanisms. To control for time heterogeneity, we add a linear time variable (Lospinoso & Snijders, 2019)

Table 14 SAOM Models for the Evolution of the Citation Network, Collaboration Network, Scientists Affiliated with Institutions and Scientists Publishing in Journals from the Web of Science

At the micro-level and for relational mechanisms, there is a tendency for scientists to prefer local transitivity processes when they decide to send ties in their scientific network in one-mode networks. For the relational-based mechanisms in the different types of models, as was expected, density is negative in the four networks, and reciprocity is positive for the citation network. The transitivity effect is positive in both versions (e.g., in the full model and for the citation network $\beta = 0.142, SE = 0.020$, and $\beta = 1.328, SE = 0.174$) indicating that this is an attractive effect for citing and collaborating with other scientists. Consistent with previous research (Block, 2015), the transitivity reciprocated triplet is negative ($\beta =- 0.127, SE = 0.021$ in the full model) as a tendency against reciprocation within transitive triplets indicating local hierarchies in the network expected in scientific networks. There is a similar tendency in the case of the collaboration network, where transitive triads and transitive ties are also attractive for closing co-authorship ($\beta = 0.362, SE = 0.056$ and $\beta = 2.075, SE = 0.534$ respectively), similar to findings in previous research (Kronegger et al., 2012; Ferligoj et al., 2015; Purwitasari et al., 2020). In the bipartite network of scientists publishing in Web of Science journals, the effect of generating four-cycle is also positive and more significant ($\beta = 0.015, SE = 0.005$).

The four models represent well the selected features of the relational-based mechanisms using the goodness of fit as a diagnostic in the citation and collaboration network. The goodness of fit test proposed by Lospinoso and Snijders (2012, 2019) is used to assess whether an estimated model fits the data well. The result shows similarities on average between the observed and simulated networks, using the macro features such as geodesic distribution, triad census and clique census (up to 10 levels).

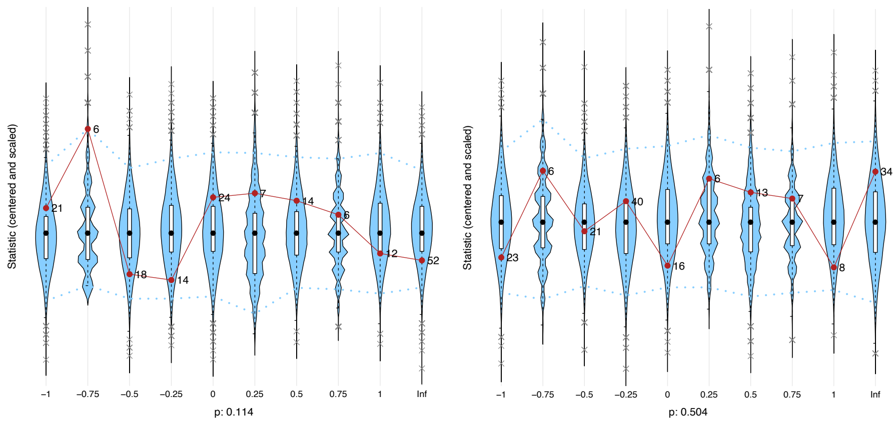
Scientists prefer degree-based processes when they decide to send ties in their scientific networks in the citation network, collaboration network and the bipartite network of scientists publishing in journals. For the indegree as a Matthew effect (Snijders, 2011), this is positive ($\beta = 0.271, SE = 0.108$ in the full model) as was expected. Scientists that receive more citations in this national disciplinary network tend to be more attractive among the academics. Researchers that cite more are also more attractive ($\beta = 0.553, SE = 0.108$ in the full model). Authors that are referring to someone who also cites other researchers more are less attractive ($\beta =- 0.151, SE = 0.048$ in the full model), which we interpreted as the tendency to cite the source actor rather than the intermediary. There is a negative tendency to cite actors that send more ties and that are citing other researchers that have higher indegree citations within this group ($\beta =- 0.223, SE = 0.343$ in the full model). The assortativity effect could represent a social niche effect within this scientific group, avoiding citing many popular researchers and as a tendency in favour of specialities (Chubin, 1976). The outdegree of at least one is negative ($\beta = -2.452, SE = 0.343$), which inversely represents a positive tendency of not sending ties to the group (that could be because of citations outside this group).

For the collaboration network, the tendency of having more co-authors is less attractive for further collaborations ($\beta = -0.300, SE = 0.087$). According to the descriptive analysis, this network is highly stable. Therefore, we interpreted this effect as a tendency to maintain a local stock of collaborators that is also confirmed by the positive transitive, and according to previous interpretations (Purwitasari et al., 2020). For the bipartite network of scientists affiliated with institutions, the institution's size makes them more attractive. As was previously mentioned regarding the relevance of big ponds (Lazega et al., 2008), at least the attractiveness of being affiliated with big institutions affects scientists' decision in this network is more significant ($\beta = 0.113, SE = 0.040$). It is attractive to be affiliated with more than one institution (the effect is positive and more significant) ($\beta = 0.358, SE = 0.158$), which in this community tends to represent participation in astronomical observatories or research centres with public funding. Being affiliated with more institutions where these organisations are also with comparatively more actors has a negative effect ($\beta = -0.361, SE = 0.191$). As an assortativity effect, we interpret this effect as the tendency of positioning and differentiation within the community. For the journal network, publishing in popular journals ($\beta = 0.418, SE = 0.048$) or being productive ($\beta = 0.084, SE = 0.029$) increases the attractiveness in being accepted by Web of Science journals.

The structures are well represented in the goodness of fit using indegree and outdegree distributions for the citation network and degree distribution for the other networks, considering the macro-level.

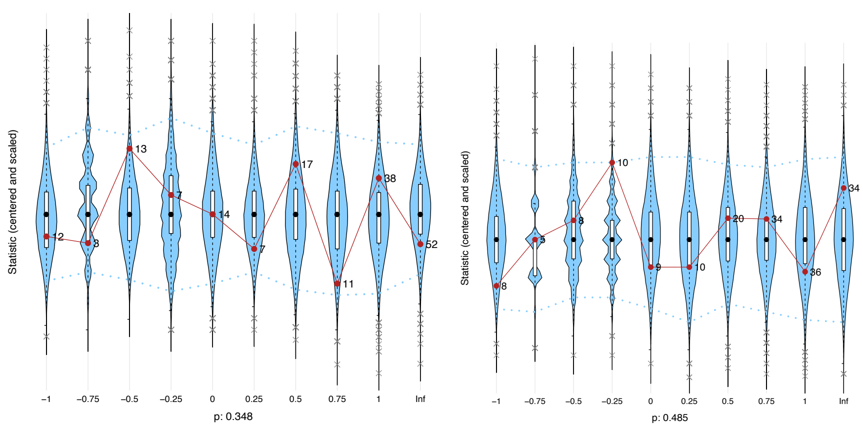
From the perspective of dyadic similarity mechanisms, there is less evidence that scientists prefer homophilous processes when they decide to send ties within their scientific networks for a citation network and the other networks. This mechanism is less significant in the case of the tendency of the scientists to cite other actors that share the same nationality (not too significant) ($\beta = -0.082, SE = 0.099$). Citation similarity is positive ($\beta = 0.054, SE = 0.027$) in which there is a tendency for mutual reinforcement of well-established scientists in this community. The effect of age similarity is negative and less significant even when the expected heterophily is present ($\beta = -0.030, SE = 0.021$), contrary to what we were expecting. The heterophily in the accumulative number of publications is more significant and negative ($\beta = -0.093, SE = 0.056$). Considering citation and publications, we do find a heterophilous reinforcement of accumulative advantages in acquired attributes, but not in ascribed characteristics (Merton, 1988) such as age and nationality for the citation network. We did not appreciate homophily for the collaboration network between researchers that share the same nationality ($\beta =- 0.024, SE = 0.445$) or a similar number of citations ($\beta = -0.063, SE = 0.030$). The similarity citation effect is positive in the tendency of being affiliated with institutions ($\beta = 2.103, SE = 0.841$). Finally, regarding publishing in Web of Science journals, the tendency that two scientists share the same or similar attributes in the case of citations does not show significance ($\beta = -0.043, SE = 0.446$). There is a heterophily tendency to have similar accumulated publications ($\beta =- 0.093, SE = 0.056$) that increase its significance level when the multiplex effects are not considered (such as in the multilevel or baseline model). The heterophily, in this case, represents the tendency of less productive researchers to collaborate with more productive scholars.

Figure 14 Goodness of Fit E-I index Distribution



Left: Citation network. Right: Collaboration network

Figure 15 Goodness of Fit Yule-Q Distribution

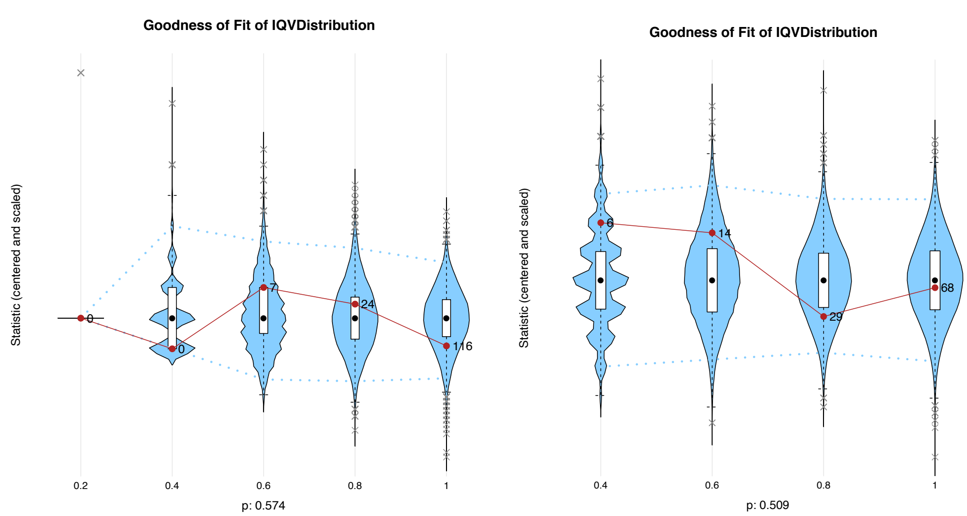


Left: Citation network. Right: Collaboration network

In the macro features for the proposed dyadic similarity-based mechanisms, the E-I index, Yule-Q, IQV, the similarity distribution and Burt's measures performed well in all cases. The observed citation and collaboration networks tend to have more homophily considering the nationality of the actors (E-I index and Yule-Q in Figure 14 and 15) for reciprocal ties, which is reasonable in the context of a local community (62% are Chileans). The homophily in the model is not significant considering the micro-mechanisms. There should be caution in making conclusions about the characteristic of the macro-level deduced from the micro-level.

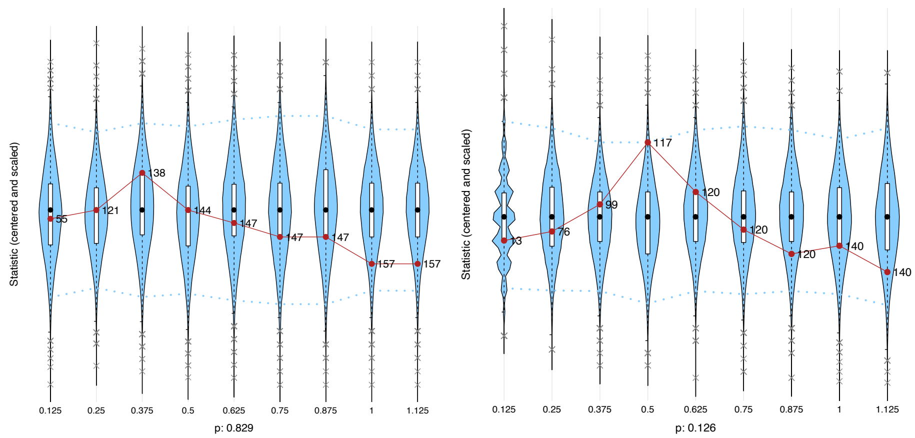
According to the IQV index the researchers have more heterogeneity in the number of contacts of a different nationality in the citation and collaboration network (Figure 16). The tendency in the micro-mechanisms of sending ties if the scholars are Chilean, compared to being from abroad, seems to be negative and more significant ($\beta =-0.239, SE = 0.055$). We interpret this as a distinction between the condition of the researchers and their attractions at the micro-level. The network is more homophilous and heterogeneous at the macro-level, considering the nationality of the researchers, but they are attracted to the international researchers at the micro-level for the cognitive network. The results are less significant considering the other control variables and homophily effect of nationality covariate in the citation and collaboration networks, with a mixed tendency between models that requires further scrutiny.

Figure 16 Goodness of Fit IQV Distribution



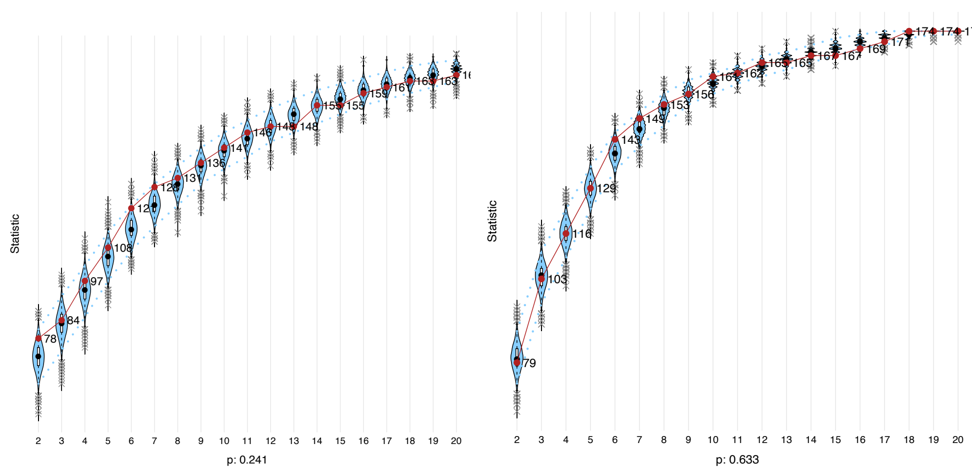
Left: Citation network. Right: Collaboration network

Figure 17 Goodness of Fit Effective Size



Left: Citation network. Right: Collaboration network

Figure 18 Goodness of Fit Constraint

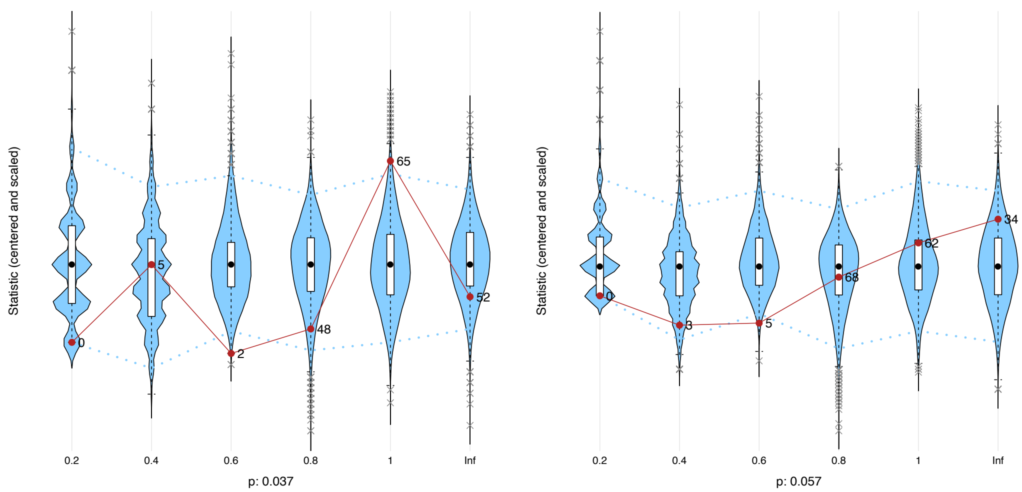


Left: Citation network. Right: Collaboration network

We consider Burt's measures to identify structural holes as the potentiality of brokerage when there is a lack of connection between the alters of an ego (Figure 17 and 18). The observed networks at the actor level represent increasingly effective size where actors tend to be more disconnected from each other. Because there is more constraint, there are also less structural holes. Degree-based measures might be some of the potential micro-mechanisms responsible for these tendencies, and a certain type of cohesive subgroups might exist in this network, and further exploration could be made in this direction.

There is also a heterophily in the macro-structure in the observed network of citation and collaboration (Figure 19). This heterophily measures a particular distribution of similarity in which researchers show a greater difference than their reciprocal ties, which displays mixed evidence in the micro-mechanisms on the estimated models.

Figure 19 Goodness of Fit Similarity Distribution



Left: Citation network. Right: Collaboration network

From the perspective of relational multiplex mechanisms, there is reasonable evidence that scientists prefer multiple relationships when they decide to send ties for the citation and collaboration network. There is a tendency of actors that are previously collaborating with others to send citation ties, and vice versa ($\beta = 1.329, SE = 0.109$ and $\beta = 2.539, SE = 0.724$, respectively). In terms of mixing degrees, the tendency of citing researchers that already have more collaborators is more significant and negative ($\beta = -0.066, SE = 0.028$). The co-citation effect, expressed as a four-cycle, is also positive ($\beta = 0.004, SE = 0.002$ in the multilevel and full model), but when multilevel effects are not considered, the significance of the coefficients increases in the multiplex and baseline models ($\beta = 0.015, SE =$

0.002 and $\beta = 0.014, SE = 0.002$ respectively). The co-citation, expressed as the tendency of the focal author that is citing one author to cite their collaborator, is positive and more significant ($\beta = 0.084, SE = 0.010$). Having the same collaborators has a negative effect on the tendency of the actors to close the triadic relationship with a citation ($\beta = -0.129, SE = 0.017$), contrary to what we were expecting, and that need further investigation. One possible interpretation is that authors might be promoting the works of others. There is an association closure when two authors previously collaborated, and one of them cites a third author, which triggers the focal author's interest to be attracted to referring the other author ($\beta = 0.82, SE = 0.024$). And, the tendency to publish in a Web of Science journal is not significantly attractive, in terms of association closure, for actors that were previously collaborating or citing ($\beta = 0.026, SE = 0.029$ and $\beta = -0.015, SE = 0.024$ respectively).

Differentiating between the network as a whole and the simulated network, we conducted the diagnostics using the overlapping triad census (Figures 20 and 21). Two of the models estimated were not able to achieve reasonable goodness of fit (Model 1 and Model 2). We explored other multiplex effects that were added into Model 3 and the full model achieving a p-value different to zero, as a threshold suggested in Lospinoso and Snijders (2019). For the extended overlapping triadic census, we reached the expected boundary, and, in the simpler overlapping triadic census, the model performs better. The exploration of the multiplex network was expected to be challenging because of the complexity of the model.

Figure 20 Goodness of Fit Overlapping Triadic Census

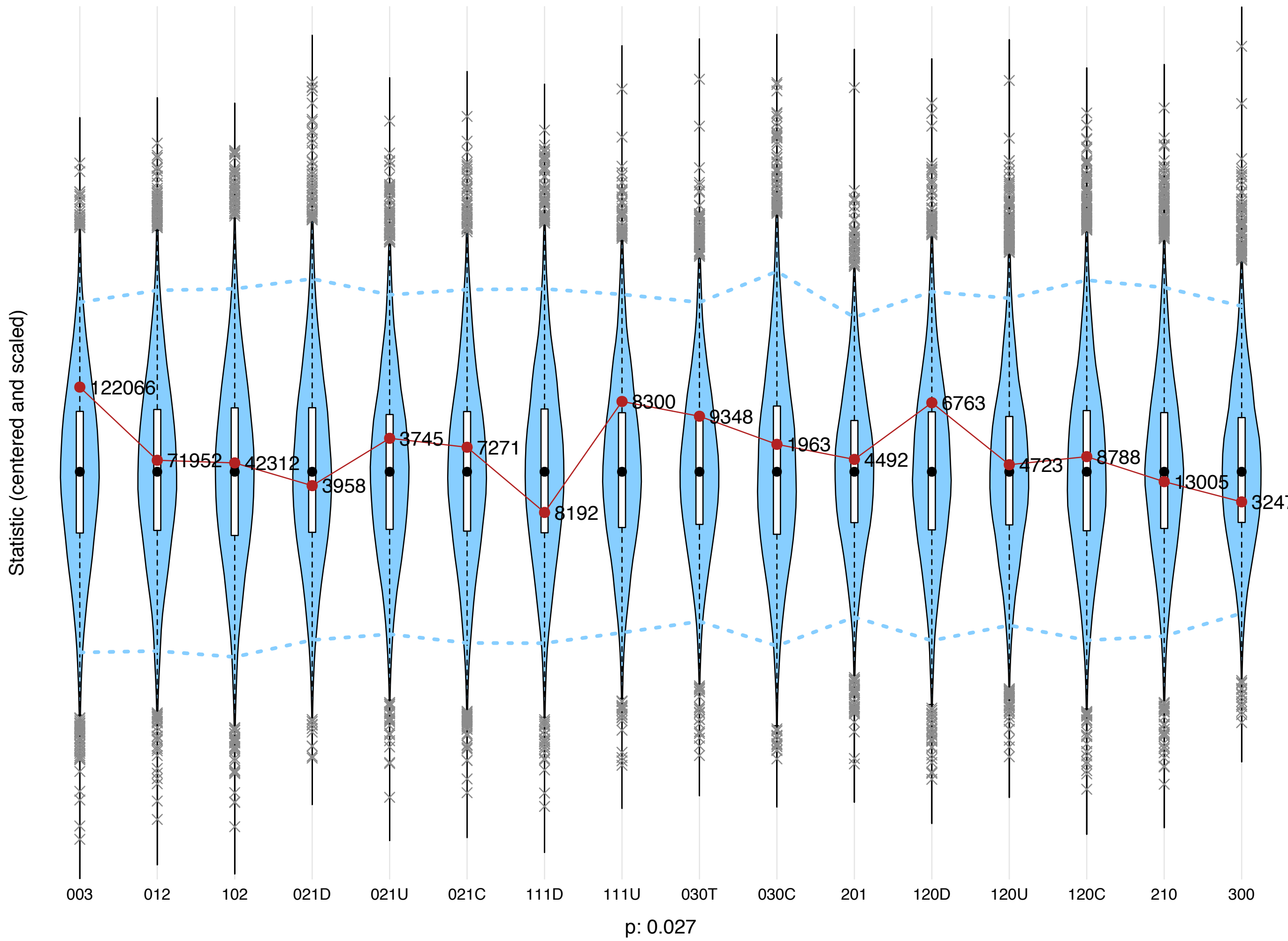
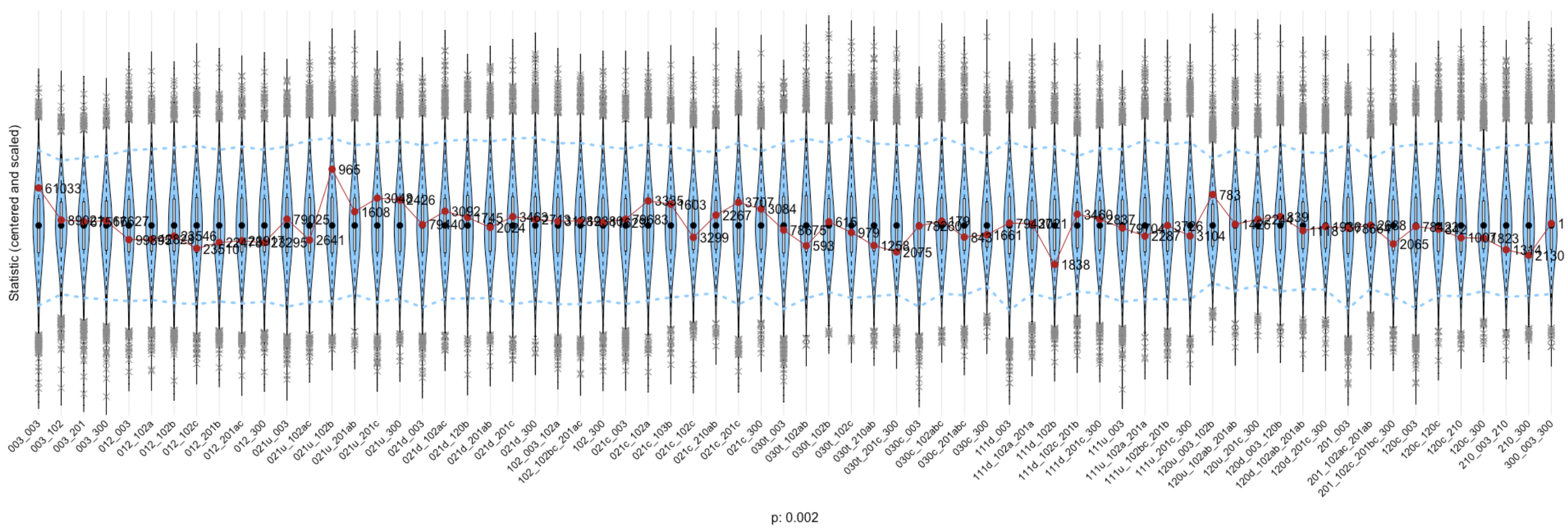


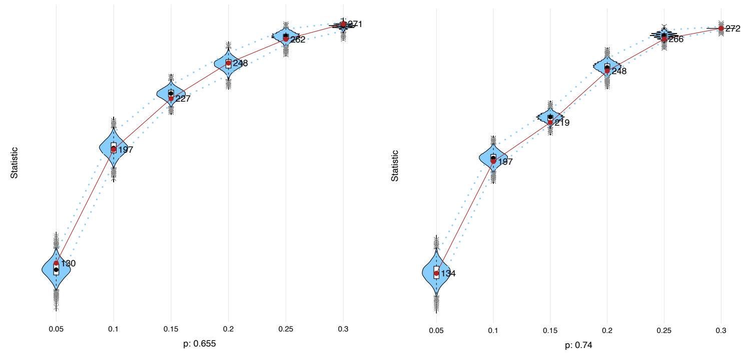
Figure 21 Goodness of Fit Mixed Layer Triadic Census



Finally, considering the main hypothesis that *scientists might prefer multilevel closure processes when they decide to send ties in their scientific networks*, there is some support for these processes at the micro-level. For the proximity mechanisms, there are some meso-level social forces across different networks in the outgoing ties for the institutional affiliation, but not for the journals from the Web of Science. In the case of affiliated closure, scientists that share institutional affiliation increase the attractiveness for citing other colleagues from the same institution ($\beta = 0.426, SE = 0.123$). There is a less significant tendency to reciprocate the ties within the same universities ($\beta = -0.069, SE = 0.247$). Therefore, results might indicate that within the departments, there are also hierarchies within cited scientists and others that mention them in proximate places. The closure by affiliation is more significant and positive in the case of the collaboration network ($\beta = 1.274, SE = 0.393$).

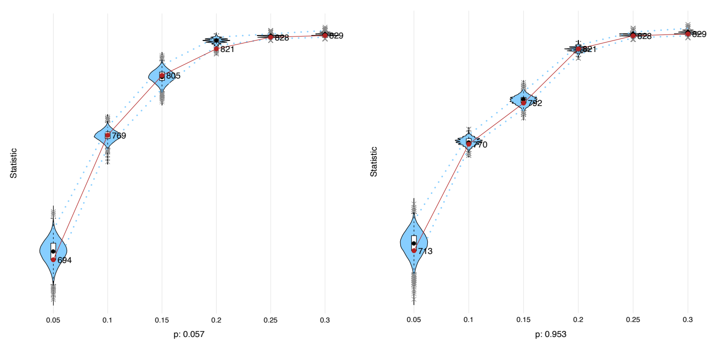
Actors that publish in the same journals are less significantly attracted to citing a researcher already publishing in the same journal ($\beta = 0.032, SE = 0.049$) and are even less likely to reciprocate a citation ($\beta = 0.000, SE = 0.247$). The same tendency for the collaboration network can be appreciated ($\beta = 0.075, SE = 0.172$). For this group, it is less significant to publishing in journals in the Web of Science than researchers that they are citing ($\beta = 0.026, SE = 0.029$), and it is negative and less significant in the collaboration network ($\beta = -0.015, SE = 0.024$) as an associated closure. The significance of the coefficients increases when the multiplex effects are not considered (multilevel model). Previous research using co-authorship and scientific topics has a non-significant but positive effect on association closure (Purwitasari et al., 2020). In comparison with topics, our interpretation is that publishing in journals requires combining different skills and perspectives to create a joint oeuvre (Moody, 2004). This might explain the difference in the direction of the parameters as two sides of a similar issue. We interpret this as a tendency in favour of concrete social relations compared with cognitive closure (Mullins, 1973; White et al., 2004) for expanding invisible colleges because they are bound by physical proximity and considering a proxy of concrete relationships through co-authorship.

Figure 22 Goodness of Fit Three-mode Mixed Outdegree Distribution



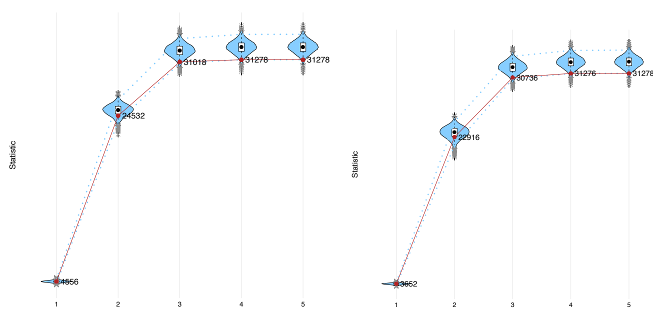
Left: Citation Network. Right: Collaboration Network

Figure 23 Goodness of Fit Three-mode Mixed Indegree Distribution (Normalised)



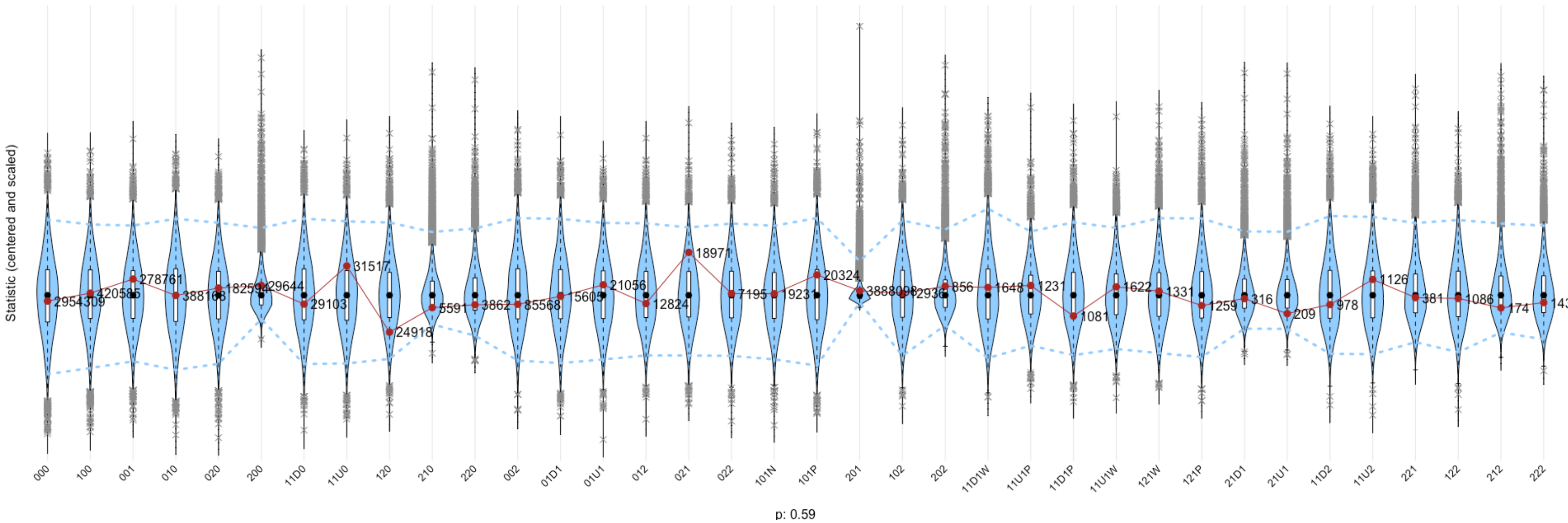
Left: Citation Network. Right: Collaboration Network

Figure 24 Goodness of Fit Three-mode Mixed Geodesic Distance Distribution

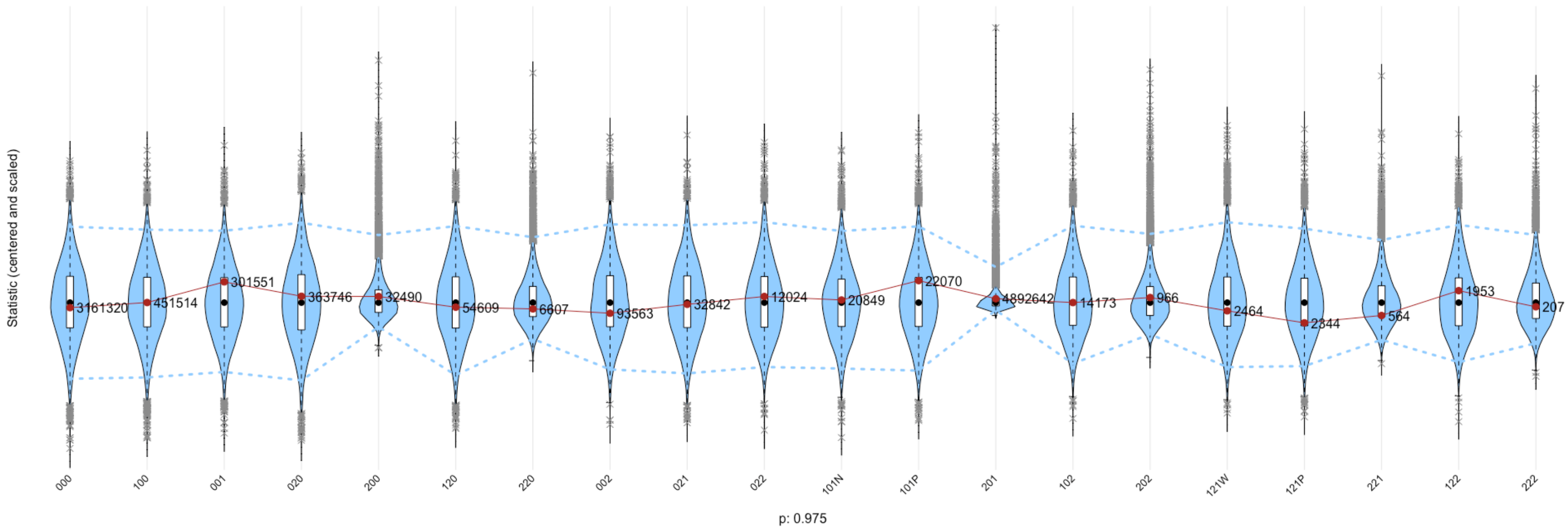


Left: Citation Network. Right: Collaboration Network

Figure 25 Goodness of Fit Mixed Multilevel Quadratic Census



a) Citation network, institutional affiliation network and journals in the Web of Science



b) Collaboration network, institutional affiliation network and journals in the Web of Science

The representation of the multilevel network at the macro level considers the three-mode normalised mixed outdegree and indegree distribution, the mixed geodesic distribution and the quadrilateral census that has reasonable goodness of fit (Figures 22 to 25). The normalisation of the three-mode mixed outdegree distribution (Figure 22) and indegree distribution (Figure 23) allows us to describe the most central actors comparing researchers, institutions and journals in a standard measure for this group (e.g., Borgatti & Everett, 1997), which could be helpful for further analysis. We also checked the non-normalised version for the two-mode and three-mode degree distributions for inferential purposes that achieved a reasonable convergence (Appendix, Section F).

Comparing the model without proximity mechanisms (Models 1 and 3), we can contrast how well some of the multilevel features are represented using micro-mechanisms to explore the multilevel features between the citation network, the scientists affiliated with institutions and the scientists publishing in journals. The inclusion of these effects (associative and affiliation closure) achieves a better fit of the model. This is the case for the three-mode mixed outdegree distribution and the mixed quadratic census for the citation and collaboration networks. For the three-mode mixed indegree distribution in the citation network, the fit performs poorly in Model 4 compared to Model 2, considering that there are consequences when the multiplex effects are incorporated. The geodesic distributions perform slightly better in the model that does not add the proximity-based mechanisms.

The main hypothesis is supported for proximity-based micro-mechanisms that consider concrete relationships and physical proximity through institutional affiliation. The results are mixed for cognitive proximity for the participation in the Web of Science journals. Further research should be done exploring other cognitive measures such as topics or other semantic networks (e.g., Roth & Cointet, 2010; Purwitasari et al., 2020; Stark et al., 2020). And, distinguishing between the levels of the citation and collaboration networks within scientists, and the co-evolutionary interdependency with the same scientists affiliated with institutions and publishing in the Web of Science, allows us to explore the presence of the different types of mechanisms considered in this study at the micro and macro levels of this network. At the micro-level of actors deciding whom to cite, whom to collaborate with, where to be affiliated or where to publish in the Web of

Science, there are no apparent differences in how the interdependence of other levels affects the multilevel network. The effect seems stable across networks, rarely affected by the presence of different levels. In some cases, significant effects diminish, which could be produced by power issues in incorporating more parameters. At the network level, the difference starts to trigger, enriching the prevalence of certain features conditioned to the effect at the micro-level under consideration.

3.7 Discussion

The results support the presence and relevance of some of the general types of mechanisms, in which scientists prefer local transitivity, degree-based and relational multiplex mechanisms. There is less support for ascribed homophily and some support for acquired homophily as a process of accumulative advantages. For the main hypothesis, the social relationships based on scientific collaboration and space proximity, and based on institutional affiliation, are more accurate in understanding the co-evolution of the networks in a scientific network when it is considered closure by affiliation, in comparison with cognitive-based networks measured through the journal network. There should be some caution in the interpretation of the co-evolutionary process and the micro-macro linkages. The mechanisms are different between the networks, and some of the features are better represented than others. These complex structures are well represented with few proximity-based micro-mechanisms.

Some of the limitations of this approach in the context of scientific networks are that the SAOM is less suitable for longer relationships unless extended periods of times are aggregated (e.g., Ferligoj et al., 2015; Purwitasari et al., 2020) or considered. We envisage that this particular issue could be explored further using novel extensions of the SAOMs for hierarchical multilevel networks of separate waves (e.g., Koskinen et al., 2015 using LERGMs). Further exploration should be done to consider the weighted aspect of the network often present in scientific networks. An alternative could be adding the weight of the network as a dyadic covariable. One limitation is that the weighted network is a function of the dependent variable requiring further assumptions. A second option might be to identify a cut point distinguishing between weak or strong ties, as is currently available in SAOMs (Elmer et al., 2017). Assigning a threshold would require further exploration and adding more networks would greatly increase the complexity of the model. From an empirical perspective, we expect to advance in some of these directions in further analysis.

Most studies that use inferential models to analyse networks take for granted the correct representation of homophily or heterophily as a macro-structure through the estimation of micro-mechanisms. In this article, we extended and applied already available measures for the goodness of fit. These are the similarity–distance distribution, the E-I index distribution, Yule-Q distribution, and the IQV for the goodness of fit, approaches that can be expanded using other measures. Using already available statistics allows the combination of descriptive and inferential analysis, which can be further explored with a backward inductive strategy to identify which micro-mechanisms were responsible for the correct representation of macro-structures.

We propose different goodness of fit for multilevel and multiplex networks in this paper. These are the two-mode and three-mode mixed geodesic distance distributions, a two-mode and three-mode mixed degree distribution (for indegree and outdegree at the actor level), two overlapping triadic censuses for multiplex networks, and the mixed quadratic census for multilevel and three-mode networks. These features are complex and achieving a good convergence using a structural perspective might allow exploring the dynamics of multilevel networks and the connections between the levels. These diagnostics can also be strategies for identifying misspecifications or potential unobserved effects. In this paper, we have shown that even by achieving the standard goodness of fit currently available in SAOM, we might be confounding other macro-structures that otherwise would be indistinguishable, which we might believe are substantively relevant to representing multilevel networks.

Acknowledgments

We thank members of the Mitchell Centre for Social Network Analysis at the University of Manchester, the Social Networks Lab at ETH Zürich and Tom A.B.

Snijders from the University of Groningen for helpful comments and suggestions that considerably improved this work.

Financial Support

The study was supported by the Chilean *National Commission for Scientific and Technological Research* (ANID PFCHA/DOCTORADO BECAS CHILE/2017, grant number 72180242). This project has also received funding from PGR Mobility Grant Scheme of the Humanities Strategic Investment Fund (HSIF) of the University of Manchester.

Conflict of interest

The authors have nothing to disclose.

Chapter 4

Meso-level social forces in an inter-organisational scientific field

Alejandro Espinosa-Rada*

**The Mitchell Centre for Social Network Analysis/Sociology, The University of Manchester*

Abstract

Scientific networks of researchers and their institutional affiliations are interdependent processes that affect each other to create scientific relationships. However, how do the regular join patterns of intercitations among researchers in organisations – as meso-level social forces – vary within scientific communities? Do researchers in core organisations have similar patterns compared to institutions on the periphery? To explore this, I analysed an inter-organisational field within an astronomical discipline. I used the Microsoft Academic database to collect data and a novel methodological strategy using a meta-analysis stationary stochastic actor-oriented model to deal with multilevel networks. I used a sample of personal networks within organisations (astronomical observatories, research centres and universities) to compare different case studies. The results indicate that researchers in this community do not preserve endogamic intercitation within their organisations. Moreover, researchers affiliated with the same external research centres tend to create intercitation in specific research niches. This tendency push toward diversity and multi-connectivity.

Keywords

Multilevel Scientific Networks; Scientific Communities; Social Network Analysis; Sociology of Organisations; Stochastic Actor-oriented Models

4.1 Introduction

Researchers have a dual position according to their peers and their affiliated institutions in scientific communities (Bellotti et al., 2016a; Lazega et al., 2008; Lazega & Jourdá, 2016). However, how do the regular join patterns of intercitations among researchers in organisations – as meso-level social forces – vary within scientific communities? Do researchers in core organisations have similar patterns compared to other institutions on the periphery?

Organisations engaging in shared activities are said to have similar pressures in terms of reputation and regulation (Meyer & Rowan, 1977; DiMaggio & Powell, 1983), which can vary according to their relative position in the inter-organisational field (Galaskiewicz & Wasserman, 1989; Powell et al., 2005). Prior research emphasises that organisations can be more important for scientific outcomes than the relational capital of individuals (Lazega et al., 2006). The dual position of actors in organisations allows the identification of social status as a meso-level perspective (Lazega et al., 2008) because they acquire social position according to peer recognition (Merton, 1968a) and the institution to which they are affiliated (Merton, 1988; Lazega & Jourdá, 2016).

Individuals and collective actors tend to create relational infrastructures as, for example, social niches or forms of social status leading to a collegial oligarchy responsible for the establishment, harmonisation, and creation of new norms, its interpretation, and priorities (Lazega, 2018). The dual position in scientific networks can be reinforced in the process of accumulating peer-recognition (Merton, 1968a) by members of their own organisation as an endogamic process (Chubin & Studer, 1979), or can be promoted as a process of stabilisation among core organisations in inter-organisational fields as an 'in-group' tendency of authors who know each other, creating interpersonal intercitation (Schrum & Mullins, 1988; White et al., 2004; Milard, 2014).

Understanding meso-level social forces – as the regular join patterns among researchers in organisations – provides insights into the mechanisms that remain stable across different levels (individuals and collective) according to their social position in the inter-organisational field. From a theoretical perspective, the actors in science are considered to have objective relations between positions already won in previous struggles. The position gives access to the monopoly of scientific authority (prestige, recognition, and fame, among others) depending on the actor's position in the institutional hierarchy and the recognition of others (Bourdieu, 1975), in which the fields that become established push towards homogenisation (DiMaggio & Powell, 1983). One way of representing these social positions in fields is through social networks (e.g., Bottero & Crossley, 2011; de Nooy, 2003; Powell et al., 2005; Ramos-Zincke, 2014). These allowed some community morphological properties in terms of relationships between researchers and organisations to be mapped, identifying the distribution and heterogeneity of their relative positions. By mapping the morphological relational structure, social network exploratory analysis can identify organisations in the core and other institutions on the periphery (Borgatti & Everett, 2000).

A wide variety of regularities are used in empirical analysis from a network perspective (Rivera et al., 2010). These regularities are specific mechanisms that operate at the micro-level, responsible for the emergence of social networks as macrostructures (Robins et al., 2005; Snijders & Steglich, 2015; Stadtfeld, 2018) formalised in features of simultaneously operating effects. There is scarce empirical research exploring the *meso-level social forces* operationalised as *cross-level effects* as specific micro-mechanisms in scientific networks (Lazega et al., 2008; Gondal, 2018; Purwitasari et al., 2020) or the interdependency between two different networks.

Most of the empirical analysis of scientific networks describes the micro-mechanisms in case studies. It does not compare organisations according to their social positions in the context of scientific fields. There are current statistical developments for social network analysis (Snijders & Baerveldt, 2003; An, 2015) that compare micro-mechanisms in different cases to understand the similarities and differences in broader populations. To identify the similarities and differences in the micro-mechanisms of the different organisations, stationary stochastic actor-oriented models (Snijders & Stiglich, 2015; Block, Stadtfeld & Snijders, 2019) were used.

The institution members are considered the core set for each organisation, and the outsiders or marginal actors relevant for network formation processes are identified (Crane, 1969; Mullins, 1972, 1973; Chubin, 1976). A similar strategy is used to reduce the size of large networks and then merge the resulting *cohesive subgroups* to estimate the common micro-mechanisms of the network when it becomes larger (Stivala et al., 2016). The suggested strategies offer an extension that uses information from one level combined with the other, characterised as multilevel networks, known as a *second-zone multilevel sampling from a second-mode focal actor*. Which, in this case represent extended opportunity structures (Lazega et al., 2013) of organisations.

The astronomical and astrophysics discipline in Chile is used as an example for the empirical investigation of this strategy. Chile has geographical conditions that enable it to have nearly 70 per cent of the global astronomy infrastructure. Scholars in universities and research centres in this country have access to ten per cent of the observation time of all these telescopes, generating an organisational field within the observatories. The year analysed corresponds to a particular moment when the Chilean government became interested in the development of astronomy to spur economic activity to national advantage (Guridi et al., 2020). This was because of the impending arrival of the Vera C. Rubin Observatory (a.k.a. the Large Synoptic Survey Telescope [LSST]), considered to be at the research forefront of the discipline (Espinosa-Rada et al., 2019). With the increasing interest in multi-disciplinary research, various research centres, observatories and local universities started preparing for this joint enterprise (Espinosa-Rada et al., 2019; Arancibia et al., 2020).

First, empirical studies that identify common micro-mechanisms in the study of scientific networks are reviewed. Then, the case study exploring the multilevel properties of a scientific community is presented. The subsequent section analyses the results, comparing different organisations within the inter-organisational field to identify the common cross-level effects responsible for the stability of this field. Finally, a discussion of the main findings is presented.

4.2 Micro-mechanisms in multilevel scientific networks

Recent studies have investigated the interdependency of the different levels to study scientific networks that identify the cross-effects between levels (Lazega et al., 2008; Bellotti et al., 2016a). The interdependency of systems can be reconstructed at least within two or more partially interlocked levels that can have different synchronisations in terms of their co-evolution (Brailly et al., 2016). These levels have different relations between entities. There is a flow of resources, positions in the system and strategies that individuals must use to appropriate, accumulate and manage their resources and the organisations' resources (Lazega et al., 2008). Academic institutions are considered corporate actors that institutionally represent the cluster of academics working there (Bellotti, 2012). These actors provide the infrastructure, organisational and intellectual environment facilitating intellectual cultures (e.g., research topics)[34]. Here, scholars shape their possibilities and are constrained by the places where they participate (Chubin, 1976).

Researchers create relationships with people they meet at the micro-level that facilitate knowledge production (e.g., advice exchange, collaboration or mentorship) (Mullins, 1972, 1973). Networks assume that actors (individual and aggregate) are simultaneously in everyday micro-interactions and affiliations at the higher macro-level (Bellotti, 2015). They are considered to create stable patterns (Merton, 1968b: 339), contrasting with more contingent interactions[35], and assuming the duality of social life (Breiger, 1974) as an interdependent process. Researchers have investigated various approaches to identify common network effects in case studies within scientific communities in recent years. These studies often analyse scientific networks considering one level. When the juxtaposition of other levels is considered, a common practice is to treat this extra network level as an attribute.

Micro-mechanisms allow an understanding of the network sub-structures responsible for the emergence of the network as a social system (Robins et al., 2005) and knowledge of how institutions are sustained and modified by individuals (Powell & Colyvas, 2008; Powell & Rerup, 2017). Different types of scientific

[34] For a more comprehensive understanding of the practices and the processes in which researchers create different types of interactions and relationship, mixed methods has been used in combination with social networks (e.g., Mitchell, 1969; Mulkay et al., 1975; Lievrouw et al., 1987; Bellotti, 2015)

[35] Stable patterns are different from *events* or *contingent interactions* that are more situational, not stable, and can become a pattern. The *continuum* between *events* and *states* has a long tradition in the network perspective (e.g., Homans, 1950; Boissevain, 1968; Borgatti & Halgin, 2011b; Crossley, 2011).

networks analysed the relevance of some of these network micro-mechanisms and studied stable relationships without considering the interdependency between levels. Some considered scientific networks from the level of authors studying collaboration networks (Fagan et al., 2018; Ferligoj et al., 2015; Akbaritabar et al., 2020), grant research (Zinilli, 2016), combining different scientific networks into a similar measure (Sciabolazza et al., 2017) or exploring different measures separately (Cimenler et al., 2015; Luke et al., 2016; Harris et al., 2017). Other approaches used emergent or collective entities as the unit of analysis, investigating similar citation mechanisms between disciplines (McLevey et al., 2018) or journals (Peng, 2015).

In line with identifying the relevance of micro-mechanisms, creating scientific relationships is also described in the sociology of science and knowledge literature identifying different mechanisms. From some of these perspectives, scientists tend to reciprocate (Hagstrom, 1965; Breiger, 1976). The assortative tendency is another type of mechanism (Rivera et al., 2010) in terms of popularity or activities (Price, 1965; Barabási & Albert, 1999; Newman, 2004) or similarity-based patterns sharing different acquired or ascribed social attributes that differentiate scholars (Merton, 1968a, 1988; Cole & Cole, 1973). Actors tend to create dyad relationships or transitivity (Mullins, 1972, 1973), or different type of closure within the scientific networks, creating invisible colleges of thought (Crane, 1972; Lievrouw et al., 1987; Zuccala, 2006). Researchers may share activities based on the subject matter, school of thought or university affiliation that often create personal and professional interactions among scholars as a multilevel closure (Mullins, 1972; Chubin & Studer, 1979; Feld, 1981). Collective actors may also create inter-organisational networks that allow them to generate more innovations, comparative advantages, or social niches through their tendency for multi-connectivity (Powell et al., 2005).

More specifically, studies that explore different micro-mechanisms often investigate how scientists share different attributes in science as a *homophily process*, in which actors sharing similar social attributes tend to interact more often in comparison with others that have different attributes (Lazarsfeld & Merton, 1954; McPherson et al., 2001). Some of the common indicators are the same gender or race, joint affiliation within college or departments, spatial proximity or similarity between topics or discipline (Cimenler, Reeves & Skvoretz, 2015; Dhand et al., 2016; Fagan et al., 2018; Harris et al., 2017; Kronegger et al., 2012; Luke et al., 2016; McLevey et al., 2018; Peng, 2015; Wang, Bu & Xu, 2018; Zinilli, 2016). These studies control for attributes such as age, gender or race, academic position, academic degree, research interest or discipline, the author's position in the papers, the number of authors or publications in papers or accumulative citations. Some of these attributes are ascribed and others are acquired (Merton, 1988). There are also similarities in terms of belonging to entities as collective actors. The similarity between these attributes are treated as homophily tendencies, mixing at the same level researcher's attribute and focus of activities (Feld, 1981).

Other types of micro-mechanisms often measured in the study of scientific networks are *dyadic* or *triadic closures* as local configurations for group formation at the level of the researchers (Mullins, 1972, 1973; Chubin, 1976) often aim to identify transitivity processes in science. A less explored effect are tie-based relationships often present in scientific networks, such as weak or strong ties among researchers. As far as I am aware, there is one study that uses this effect, adding a dyadic effect into the model, controlling for other micro-mechanisms, and considering the frequency in communication such as personal contacts (strong) and professional networks (weak) as a relevant aspect to form ties (Cimenler et al., 2015).

A transitivity triad involves actors i, j, and k, which is transitive if whenever $i \rightarrow j$ and $j \rightarrow k$, then $i \rightarrow k$ (Wasserman & Faust, 1994). The study of transitivity in scientific networks identifies the probability that two scholars would collaborate if they shared collaborators (Newman, 2001b, 2004) by using different types of triadic isomorphism classes (Davis & Leinhardt, 1972) and further specifications (Hunter & Handcock, 2006; Snijders et al., 2006). This local structure has been further studied as micro-mechanisms in scientific networks (Dhand et al., 2016; Fagan et al., 2018; Harris et al., 2015; Kronegger et al., 2012; Luke et al., 2016; McLevey et al., 2018; Peng, 2015; Sciabolazza et al., 2017; Zhang et al., 2017; Zinilli, 2016).

Another common type of mechanism is the peer recognition tendency of actors to send or receive more scientific ties than others. In the social network perspective, this was identified as the tendency of actors to receive more nominations as a friend than others as a 'social dynamic effect' (Moreno & Jennings, 1938), and as a relevant feature of macro-level networks represented in the skewed distribution of different scientific networks (Price, 1965; Barabási & Albert, 1999; Newman, 2004). From the perspective of micro-mechanisms in scientific networks, this is often one of the features considered in the analysis (Dhand et al., 2016; Harris et al., 2015; McLevey et al., 2018; Peng, 2015; Zhang et al., 2017).

Consequently, it would be expected to identify different general types of mechanisms for the first level of scientific networks. These mechanisms are classified as accumulative advantages (Merton, 1988; McPherson et al., 2001), assortativity-based mechanisms as a local peer recognition effect (Merton, 1968a, 1988), and dyadic or triadic processes corresponding to group formation (Mullins, 1972, 1973; Chubin, 1976).

A different perspective uses an *analysis of multilevel networks* (Snijders, 2016) in which the cross-effects are directly explored, maintaining the network levels separately – within and between levels – and assuming that a second level has a certain level of compactness in which a type of collective agency could be deduced. This second level allows the study of scientific networks to identify the relevance of intermediary entities between scholars (Gondal, 2011) or to explore cross-effects such as three-path, triadic closure incorporating both entities and shared attributes in the intermediary levels (Wang et al., 2013; Gondal, 2018; Purwitasari et al., 2020).

In these applications, triadic closure considering two different levels tends to be a common mechanism in scientific networks (e.g., Wang et al., 2013; Gondal, 2018; Purwitasari et al., 2020) and has been explored in other applications as the tendency to create closure by affiliation or association (Lomi & Stadtfeld, 2014; Fujimoto, Snijders & Valente, 2018). In the case of closure by affiliation, actors tend to create ties with other actors if they belong to a similar second level entity (e.g., if two actors are affiliated with the same scientific institutions, they tend to create a tie between them). Closure by association assumes that actors would tend to create ties (i.e., participate or become a member) to the second level entity if associated with an actor that already belongs to that level. For example, if an actor creates a tie with another actor affiliated with an institution, the first actor would be attracted to the same institution.

From a more abstract and general viewpoint, the context or global perspective can be considered (Snijders, 2016) by identifying the position of the actors in the organisational fields. Some particularities of the organisational field are the increase in interactions among organisations, the identification of the inter-organisational structure of domination and patterns of coalitions, the information flowing in the organisation and the mutual awareness among participants in a joint enterprise (DiMaggio & Powell, 1983). From this perspective, there is an emphasis on the organisational environment's relevance affecting the organisations through their inter-dependency as a vertical or horizontal process.

Following the work of Warren in its understanding of the community structures, Scott and Meyer (1991) distinguish the tendency of organisations to have vertical and horizontal patterns of relations linking the social units within and among communities, leading to a prevalence of extra-community relations and declination of the autonomy and cohesiveness of the community. Lazega et al. (2016) also identify a similar pattern in researchers that seek advice. In which, actors seek advice from competitors more often if they share the same social niche as an endogamic process. *Multi-connectivity* is the tendency of organisations to have multiple links (direct and through chains of intermediaries) and a preference for diversity (Powell et al., 2015). From this research's perspective, and in stationary scientific networks, the *closure by affiliation* and *closure by association* triadic multilevel closure differentiates a tendency of actors to maintain the endogamy as a vertical pattern or multi-connectivity as a horizontal pattern among organisations, respectively.

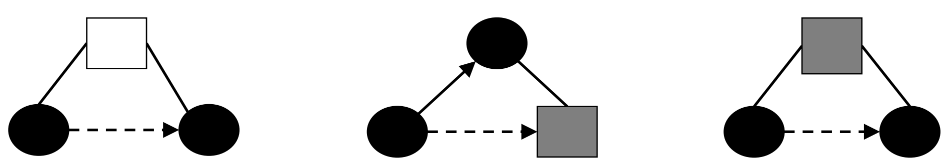


Figure 26 Closure by Affiliation and Closure by Association

a) Left: Closure by affiliation (ego); b) Centre: Closure by association; c) Right: Closure by affiliation. Black circles represent researchers, white squares represent the core organisation of the personal network, and grey squares represent external organisations. The dashed line corresponds to tie formation, and the solid lines are simulated as previous ties.

Figure 26 represents different micro-mechanisms controlling for vertical and multi-connectivity in the different patterns of creating interpersonal intercitations. From a variation of a type of *closure by affiliation*, we can identify whether researchers prefer creating interpersonal intercitations – 'the record of who has cited whom within a fixed set of authors' (White, 2011, p. 275) – within their organisations as a vertical or endogamic process (Figure 26a). Previous research identifies that in intra-laboratories, there is often a self-citation tendency when actors are part of the same organisations (Chubin & Studer, 1979). *Closure by affiliation* can represent the multi-connectivity pattern as the tendency of actors that share more external affiliations to create more interpersonal intercitation ties (Figure 26b) from a broader 'in-group' (Schrum & Mullins, 1988; White et al., 2004). *Closure by association* identifies multi-connectivity as the tendency of actors already creating intercitation ties to other actors to be attracted to being affiliated to the same external organisations (Figure 26c). These micro-mechanisms allow exploration of the standardisation and the context in which the researchers create interpersonal intercitations. This scientific community is expected to have

H1a: A *closure by affiliation tendency* within members of the same institution.
H1b: A *closure by affiliation* and *closure by association* of actors that tends to be affiliated to external organisations.

Even when similar forces could be prevalent across different organisations, variations and specificities are often expected from a field perspective (Wooten & Hoffman, 2017), leading incumbents' actors in more stable fields to reproduce themselves over more extended periods. They can maintain rules and cultural norms specific to their domain (Fligstein & McAdam, 2012), which can be considered as the tendency of core organisations to be more endogamic and generate more multi-connectivity to maintain the field. Recent developments that focus on the analysis of micro-mechanisms to compare overall tendencies across cases studies or different networks use *multilevel network analysis* (Snijders, 2016).

From this perspective, cohesive subgroups[36] (Fortunato, 2010; Everett & Borgatti, 2019) are distinguished in scientific networks. The prevalence of micro-mechanisms among these cohesive subgroups is explored within disciplines and then compared with other disciplines or similar substantive research focus (Ferligoj et al., 2015; Kronegger et al., 2012; Sciabolazza et al., 2017; Akbaritabar et al., 2020). In these studies, the aim is to explore the stability of these micro-mechanisms and identify specific effects and their change in significance according to the disciplines or similar substantive research focus. When effects are similar between these cohesive subgroups, similar micro-mechanisms are present. Common constraint could be recognised as a process toward homogenisation and stability.

The variation of these micro-mechanisms could be compared in a two-step approach (estimating the model, then comparing the cohesive subgroups between them) (Ferligoj et al., 2015; Kronegger et al., 2012). Another option is through a one-step analysis (Sciabolazza et al., 2017; Akbaritabar et al., 2020), where the variation between the cohesive subgroups is included as cross-effects to control for the internal variation (in this case, it often uses the second level as an attribute of the first level). A second hypothesis is an expectation that

H2: There will be variations in the strategies of the cohesive subgroups, in which organisations in the core will tend to be more endogamic (*closure by affiliation [ego]*) and have more multi-connectivity among organisations (*closure by association and affiliation*) than institutions on the periphery.

The different hypothesis uses the interdependency of the two levels to analyse multilevel networks to explore *closure by affiliation* and *closure by association* mechanisms. From a multilevel network analysis perspective, the

[36] In the sociology of science and knowledge, the cohesive subgroups are often considered as an *invisible college* (Crane, 1972; Lievrouw et al., 1987; Zuccala, 2006), which had been referred to other different terminologies such as 'research clusters', 'specialities', 'research network', 'collectivities', 'scientific community', among others (Hagstrom, 1976; Schrum & Mullins, 1988; Morris & der Veer, 2009).

differences among effects between cohesive subgroups in a scientific community can be analysed. To explore the multilevel property of scientific networks, considering the micro-mechanisms responsible for the emergency of global patterns, the interdependency of different entities and the variation of these micro-mechanisms between groups in a scientific community, I would study a national scientific field.

4.3 Data and Methodology

4.3.1 Data

The astronomical and astrophysical community in Chile currently has access to some of the most relevant astronomy infrastructure worldwide. There is an inter-organisational complex network of international research groups, local research centres, both public and privately funded, and different universities in this scientific community. Chile has geographical conditions that enable it to have nearly 70 per cent of the astronomy infrastructure. Scholars in universities and research centres in this country have access to ten per cent of the observation time of all these telescopes.

For the analysis, the complete record of researchers affiliated in organisations settled in Chile in 2017 and published under 'Astronomy and Astrophysics' in the Microsoft Academic database was extracted. This year overlaps with the period in which the ASTROdata program was created, funded by the Strategic Investment Fund of the Economic Ministry through the Digital Transformation Agency (CTD) of the Chilean Economic Development Agency (CORFO). This program aimed to identify and initiate measures and investments to diversify and grow the Chilean economy using natural advantages in astronomy and astrophysics (Espinosa-Rada et al., 2019; Arancibia et al., 2020) and was considered as having a spill-over capacity to help in the development of the country (Guridi et al., 2020). The organisations became a principal focus of inquiry because of the upcoming arrival of the Vera C. Rubin Observatory (a.k.a. the Large Synoptic Survey Telescope [LSST]) – one of the biggest telescopes worldwide that would create an impressive amount of information, leading to the so-called data turn in astronomy (McCray, 2017).

After extracting the data, the institutions and the authors of the database were manually disambiguated. Microsoft Academic was used as it can extract complete records of all references to each paper covering a significant amount of citation (Martín-Martín et al., 2021; Visser, Jan van Eck & Waltman, 2020) rather than other well-established databases (such as Google Scholar, Web of Science, Astrophysics Data System and Scopus-Elsevier) which cannot. By using this information and the references, it is possible to distinguish between directed citations (from paper p to paper q) with complete information on each paper referred to, allowing the extraction of all authors indexed in the references. The analysis is limited to references within authors affiliated with institutions located in Chile to analyse the local environment. The total number of authors considered in the references is based on the number of Chilean authors (i.e., affiliated to institutions located in this country) participating in the paper and not the total number of authors (including researchers from abroad).

The total numbers of researchers (cited and citing) were $a = 440$ within 2017, and the number of different institutions of the researchers in 2017 was $u = 29$. As a proxy of accumulative citations, the number of citations accumulated was extracted. Then the average citation from all the papers in which they were co-authors was considered. The type of organisation (i.e., university, research centre or astronomical observatory) and the size of each institution were identified for organisational purposes.

There are a variety of organisations (acronyms in the Appendix, Section I). Some are research centres dependent on public funding (MAD and MAS), universities (AIUC and CMM), international partnerships (UMI-FCA), private research centres of astronomy (HARLINGTEN and INEWTON), national universities (PUC, UA, UAUTONOMA, UCH, UCN, UDA, UdeC, UDP, ULS, UNAB, USACH, UTFSM, UV), and international observatories (CTIO, GEMINI, JAO, LCO, NAOJ, NOAO, NRAO, and SOAR). In some cases, the organisations are sub-units of broader institutions. For example, JAO is the Atacama Large Millimeter/Submillimeter Array (ALMA) scientific team, currently one of the largest radio astronomical observatories, which is also part of the partnership between ESO, NAOJ and NRAO. The authors' institutional affiliation on the papers was stated and checked according to each author's official web page to disambiguate their institutional affiliation and to identify, for example, if they were part of the same department or scientific group.

The citation network was analysed for the first level, in which two citing works, p and q, have a relationship in which $p\ Ci\ q \equiv$ *work p cites work q*. From the works, the citation measure was extracted from the authors' perspective, where A_{ik} is an incident matrix in which an actor i participated in a work k, and B_{kj} corresponds to the work k that was authored by j. With this information, the author citation network could be derived as $Ci = A_{ik}B_{kj}$, as the aggregation of *oeuvres* (White & Griffith, 1981), and the diagonal was set to zero. From matrix Ci, only the authors cited in 2017 were analysed (creating a square matrix of dimension $Ci = a \times a$). Further investigation was carried out for authors cited before 2017 in the Appendix (Section K).

Differentiating between both networks enabled the comparison of actors simultaneously publishing in the same year[37] as an *intercitation* network (White, 2011), which can potentially reciprocate a tie in the aggregated network compared with authors[38] that were not present in the considered year. The study also distinguished between strong ($> median(Ci_{ij})$) and weak ties ($< median(Ci_{ij})$) for the network of 2017 in which strong ties represent the direct relationships between the actors and weak ties as a controlling dyadic covariate in the model (Cimenler et al., 2015). The citations are considered an approximate measure of social debts that are social and intellectual at the same time (Crane, 1972: 20; Chubin, 1976: 451–452), but citations are difficult to understand (Gilbert, 1977; Nicolaisen, 2008) when the interpersonal intercitations are not considered. The interdependency with a second level entity is represented through the institutional affiliation network expressed as an incident matrix X_{iv}, in which researcher i is affiliated to an organisation v situated in Chile.

---

[37] On average, each author published 10.48 papers with $median = 5$ and $s.d. = 14.20$ in 2017.
[38] The researchers in 2017 cite other authors on average from 1.88 years ago, $median = 1$ year and $s.d. =$ 2.25.

4.3.2 Stochastic Actor-Oriented Model

There is increasing interest in analysing multilevel networks (Lazega & Snijders, 2016), often distinguished as the analysis of a network with nodes and ties of several types (*analysis of multilevel networks*) or a multilevel strategy in which a sample of social networks (*multilevel network analysis*) are combined (Snijders, 2016). For the first perspective, the focus is often the interdependency between actors and different entities in specific settings. The variation within and between cohesive subgroups in a community can be identified by comparing different case studies.

The stochastic actor-oriented model ([SAOMs] Snijders, 2001; 2017; Snijders, Van de Bunt & Steglich, 2010) is a versatile approach convenient for understanding multilevel networks (Snijders et al., 2013; Snijders, 2016). One feature is the emergent properties that allow researchers to identify the connections between micro-macro levels (Snijders & Steglich, 2015; Stadtfeld, 2018). They explore the micro-mechanisms responsible for the features that arise from actors to understand the emergence and stability of cohesive subgroups (i.e., structural groups) in social networks (Stadtfeld et al., 2020). Other studies have explored the interdependency of different entities (Snijders et al., 2013), the identification of similar meso-level social forces in a community and their variation within and between groups using a two-steps approach (Snijders & Baerveldt, 2003; An, 2015) and more recently a one-step approach (Snijders et al., 2020).

For the analysis, SAOMs were used, which is estimated through RSiena software (Ripley et al., 2021) and analysed as cross-sectional data (Snijders & Stiglich, 2015; Block et al., 2019). SAOMs for cross-sectional data are an actor-based alternative (Block et al., 2019). Another option is the exponential random graph model, which is a tie-based approach, and currently the statistical model most used to analyse social networks (ERGM; Lusher et al., 2012; Schweinberger et al., 2020). For each cohesive subgroup, 10,000 iterations were specified in phase 3 to calculate standard errors, and the estimation was made using the Method of

Moments[39]. A meta-analysis of different social networks and iterated weighted least squares estimator was used without assuming a normal distribution (Snijders & Baerveldt, 2003).

To compare different organisations, the social environment or personal network was considered for each organisation, treating the organisation as an ego-network (Crossley et al., 2015). The personal network is a specific type of cohesive subgroup of an actor (organisation in this case) considered as role-relationships (Bott, 1968: 3), quasi-groups (Meyer, 1966: 115–116), or stars and zones relationships (Barnes, 1969: 60–61) – and further referred to as personal networks (Boissevain, 1974: 26–27). From a different perspective, Collins (1974: 177–178) suggested in his ethnographical study about a gas laser (known as TEA) that a way to shape a network could start from contact with a laboratory, in his case, the Canadian defence research laboratory as an ego-network (the core set), and then trace other actors involved in the diffusion of knowledge[40]

In this case, the cohesive subgroup would be based on the institutional affiliations of the scholars as an extended opportunity structure (Lazega et al., 2013) of the organisation. It is considered that the actors that belong to these organisations are the core set of this institution, in which they are co-workers and where there is informal communication and colleagueship (Mullins, 1972, 1973; Chubin, 1976). Furthermore, the actors are citing or have been cited by external researchers that are 'outsiders' or ‘marginal’ (Crane, 1969; Chubin, 1976) and affiliated to other institutions in a circuit of broader communication, referred to as an invisible college (Crane, 1973; Lievrouw et al., 1987; Zuccala, 2006). Similar strategies from a methodological approach select seeds of actors (ego) and a subset of the nodes that are at a distance of two (or more) to estimate the size of the hidden population (Frank & Snijders, 1994). Others identify hard-to-reach or hidden populations (Giles & Handcock, 2010) and reduce the size of big networks merging the resulting personal networks to estimate the common micro-mechanisms of the

[39] The *rate functions* (Snijders, 2001) were fixed to $\lambda = 100$ for the citation network and $\lambda = 30$ for the institutional affiliation since in the first case there are often more opportunities to create ties in comparison with the second network.
[40] In the following years, Collins (1981, 1988) identify the *core set* through controversies and the social contingencies of those involved in experimentations and observations, but in which it is not possible to know who is inside or outside the core set.

network (Stivala et al., 2016). From the proposed strategy, multilevel networks were included in the creation of the seeds.

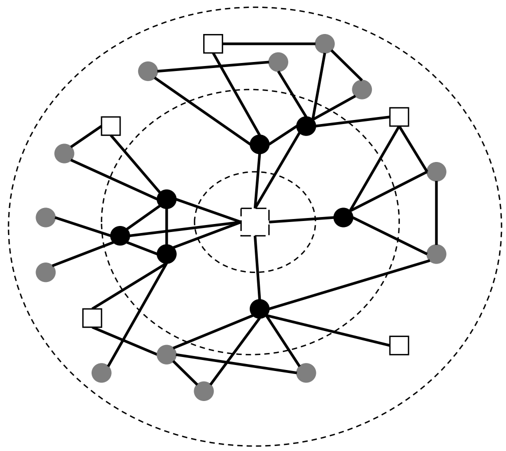


Figure 27 Personal Network of an Organisation (ego)

In Figure 27, a cohesive subgroup is represented from the perspective of an organisation (ego) and its members as the core set, considering its social environment or personal network. In Figure 27, squares represent organisations and circles researchers. The black nodes represent members of the focal institution (white dash square that is further extracted for the analysis) while grey nodes represent researchers outside the focal institution cited or citing the organisation's members. The methodological strategy used here is a second-zone multilevel sampling from a second-mode focal actor. After identifying the core set, other actors (researchers and organisations) at a distance of two from the organisation are included. The underlying aggregated strong citation network and the institutional affiliation are used together as a multilevel network. The exploration is restricted to a distance of two; otherwise, the sample of networks increase their size (Feld, 1991) significantly. A further examination should explore the consequences of using this strategy for larger distances in SAOMs. Following this approach, the interdependency of institutional affiliations and actors can be explored to identify cross-level effects as the presence of meso-level social forces and the variation of the effects between cohesive subgroups (i.e., personal networks of a sample of organisations) within a local scientific community.

4.3.3 Explanatory Variables

Different mechanisms of group formation were controlled for the citation and organisational network. *Density* was considered a similar measure to the intercepts in standard linear regression models, and *reciprocity* was considered as the tendency to create intercitation (White, 2011) for the citation network. For peer recognition processes or assortativity-based mechanisms, control was for indegree and outdegree. In some of these measures, the square root version was used when there were reasonable differences between low and high degrees.

Actors that received more citations from other researchers in the personal network will be more attractive to be cited on average (*indegree popularity*). According to their outdegree, actors that cite more would be attracted to send even more citations (*square outdegree activity*), and actors at a distance of two would be more attractive to be cited (*square outdegree popularity*). It was also controlled for whether citing more often than another researcher leads to more reciprocity (*reciprocity degree activity*). For the affiliation network, the effect of the assortativity-based mechanisms will be similar, but the interpretation, in this case, is different. For the indegree, it was assumed that if more actors are affiliated with the organisation, it will be more attractive for affiliation (*square indegree popularity*). Actors in more institutions will have even more affiliations (*outdegree activity*), as organisations tend to have direct multi-connectivity and a preference for exploration at specific moments (Powell et al., 2005).

Different transitivity measures were considered to explore triadic closures within scholars in the citation network. These transitivities were differentiated between direct and indirect transitivity. For bipartite networks, triadic closure was not possible. The differences between these effects are that *transitive triples* assume that if an actor sends a tie to another actor and the second actor sends a tie to a third researcher, the first actor would be attracted to send a tie to the third actor. *Transitive ties* generalise transitive triplets because they consider the direct path between any pair of actors and count the indirect paths between them. Variations of triadic closure are often used to analyse citation networks. A dyadic covariate effect was also added using the weak weighted ties citation network to control for co-citation of authors (*co-citation from weak ties*), in which authors are perceived as cognitively similar by a third party (White, 2003). Co-citation tends to represent authors that are close intellectually or reflect conflict or oppositions between them (White, 2011).

The cross-level effects *closure by affiliation* and *closure by association* were used to understand the *meso-level social forces* hypothesis. The first case distinguishes between an endogamic dyadic covariate effect (*closure by affiliation (ego)*) to identify whether being affiliated to the core set organisation creates a tendency to be attracted to cite researchers from the same institution. *Closure by affiliation* identifies if a researcher will cite another author if they are affiliated to the same external organisations. *Closure by association* identifies whether the attractiveness to an institution is because the researchers were citing others already affiliated.

4.4 Results

4.4.1 Descriptive and Explorative Analysis

The 29 personal networks within organisations (described in the Appendix, Section I) range from 2 to 132 actors in the first level. These organisations have between 0 and 18 ties with external organisations. The range of the median degree is 0 to 10.94 in the citation network and 0 to 13.27 in the institutional affiliation network. Removing the personal networks smaller than ten actors, the group mean is 3.89, median 3.36, and the standard deviation is 2.91 for the citation network. The average degree of the number of institutional affiliations is 5.75, with a median of 5.25 and a standard deviation of 3.10. In most personal networks, there are fewer inside members than external actors. The organisations where there are more inside members than external are, for example, the *European Southern Observatory* (ESO), the *University of Chile* (UCH), and *the Pontifical Catholic University of Chile* (PUC). The last two organisations are two of the most prestigious universities in this country.

From the sample of organisations, 11 cases achieve convergence with overall maximum convergence < 0.25 and with t values smaller than 0.1 in absolute value. These cases have reasonable goodness of fit in key features not explicitly included in the model specification (Lospinoso & Snijders, 2019) (p-values of each case that achieve convergence in the Appendix, Section J). The converge models are some of the core personal networks of organisations that have a well-established astronomical department (PUC, UCH, UdeC, UDP, ULS, UV), two research centres (MAS, CMM) and certain observatories (CTIO, ESO, LCO). Most of the universities with a consolidated astronomy department (CONICYT, 2012) are in the analysed group, except UA, UAUTONOMA, UCN, UDA, UNAB, UTFSM, and USACH that have small personal networks for SAOMs (< ten researchers). In contrast, most of the personal networks of the international observatories did not achieve convergence (GEMINI, JAO, NAOJ, NOAO, NRAO, SOAR), which might require further scrutiny.

Some context is provided by describing personal networks that achieves convergence. ESO contributes and operates some of the most important astronomy observation sites in this country (e.g., La Silla, Paranal and Chajnantor). LCO was established in 1969 in the Atacama Desert of Chile and is owned by the American private research centre Carnegie Institution of Science. This observatory holds various telescopes (i.e., the Magellan telescopes) operated in collaboration with the University of Michigan, the University of Arizona, Harvard University, and the Massachusetts Institute of Technology. In the case of CTIO, this observatory contributes to gathering data for the discovery of the Universe's accelerating expansion, leading to the presentation of the Nobel Prize in Physics in 2011 to Saul Perlmutter, Brian Schmidt, and Adam Riess. This observatory was vital to the local history of this community due to the contribution of the Chilean team, Mario Hamuy, Mark Phillips, Nicholas Suntzeff and Jose Maza, working on these telescopes in the discovery and achievement of the scientific contribution and the further recognition of the laureate Nobel prize. There was some controversy due to the allocation of scientific recognition in the discovery (Heidler, 2017).

Some of the astronomical observatories are also small (NOAO, NRAO and SAOR $\leq$ 10 nodes) or (almost) exclusively cite researchers from their own observatories in the considered year (GEMINI and NAOJ). The exception is JAO which has inside/outside citations of its network sufficient for SAOM but did not achieve convergence with the suggested specification. These organisations have scientific teams who settle in this country, working for international observatories. Some of these organisations are more or less embedded in the local community.

The Millennium Institute of Astrophysics (MAS) is a Millenium Institute, a program funded by the National Agency of Research and Development (ANID) of the Chilean Ministry of Science, Technology, Knowledge and Innovation. These research centres are funded through public competitions considering their scientific merits since 1999 and are created to study specific areas of knowledge that contribute to this country. MAS is currently a research centre collaboration created by researchers from PUC, UCH, UNAB, UdeC, UV, and UAI to develop and prepare the new generation of researchers in big data as a niche in astronomy. One of the central scientific policies of this country is to become a worldwide leader in the subfield of astro-informatics during the next decade (Espinosa-Rada et al., 2019).

Universities and professional institutions create organisations that become similar through professionalisation where central organisations act as models for other organisations driving status competition (DiMaggio & Powell, 1983). The organisations involved tend to hire individuals from the same field, establishing a cognitive base occupation that rests on formal educational institutions and a professional network. Following Scott's definition, a field is 'a community of organisations that partakes of a common meaning system and whose participants interact more frequently and fatefully with one another than with actors outside the field' (1995: 56).

Figure 28 Discrete Core-periphery Structure of the Chilean Astronomical and Astrophysical Inter-organisational Network in 2017

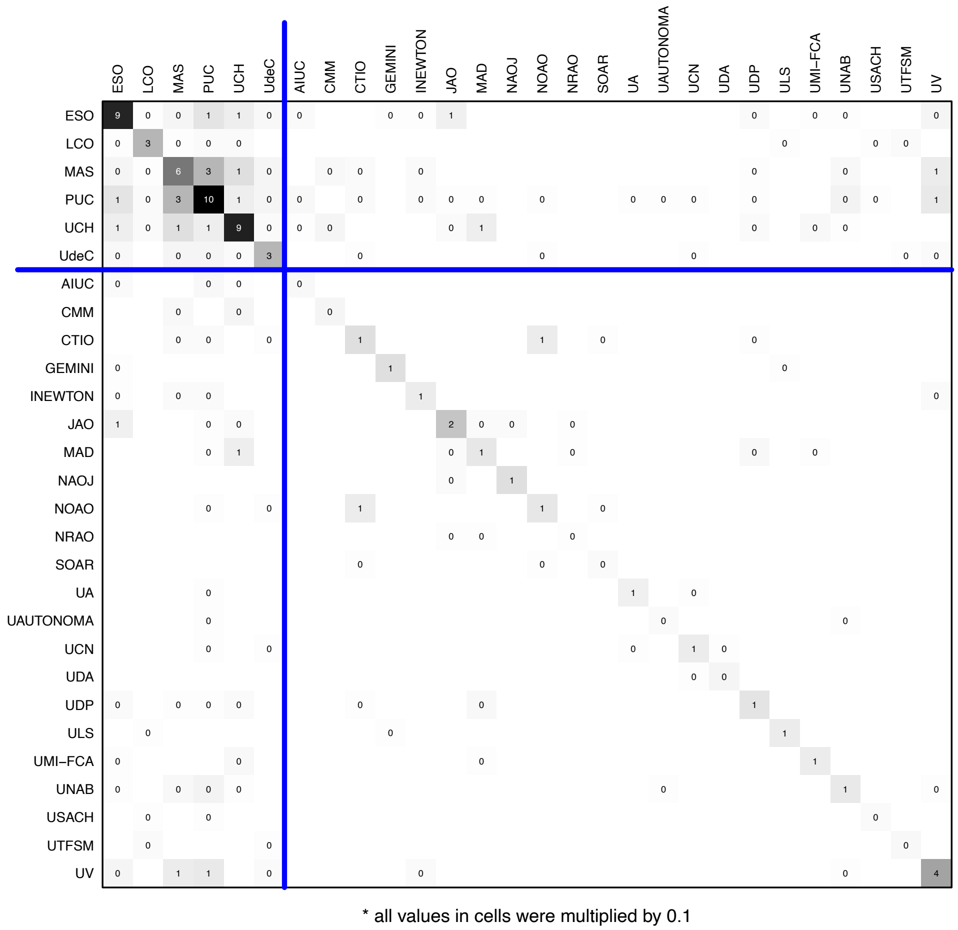
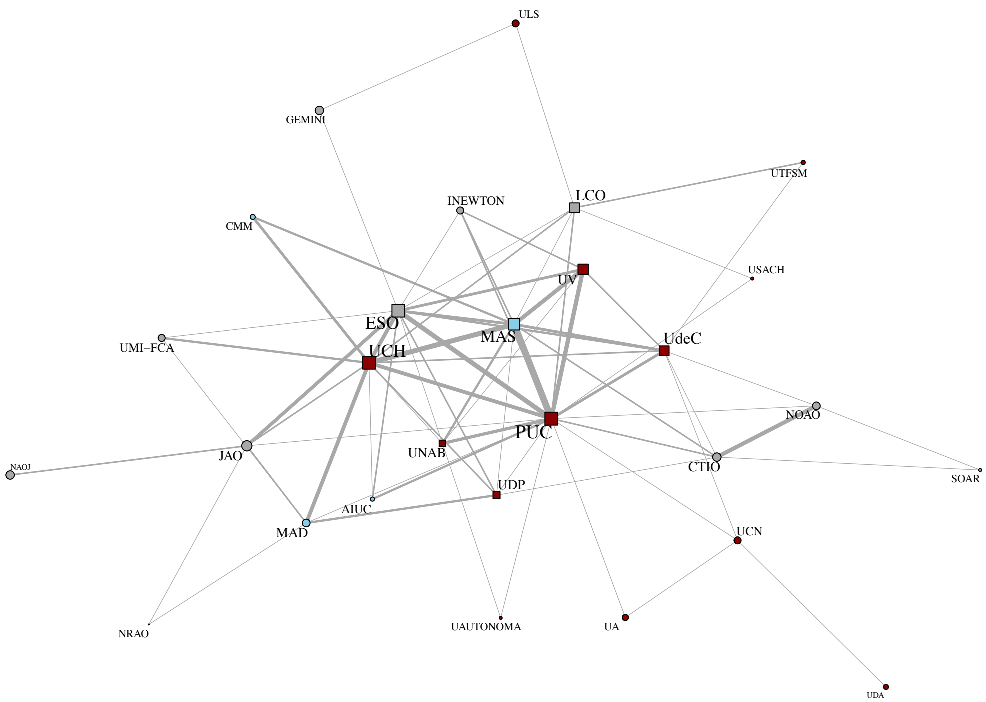


Figure 29 Continuous Core-periphery Structure of the Chilean Astronomical and Astrophysical Inter-organisational Network in 2017



The representation uses stress majorisation layout (Gansner et al., 2004). Blue nodes are research centres, red nodes are universities and grey nodes are astronomical observatories. The size of the nodes is according to the Freeman degree (1978), and the ties are the shared number of actors affiliated in both institutions.

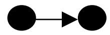
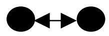
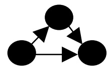
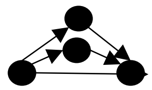
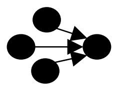
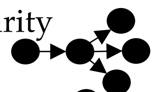
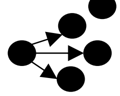
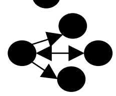
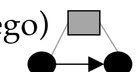
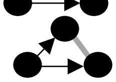
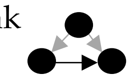
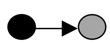
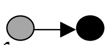
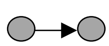
Exploring the context through the inter-organisational network, most organisations that achieve convergence in SAOMs models are also on the core side of the national core-periphery structure (Borgatti & Everett, 2000) in the discrete (Figure 28) and continuous (Figure 29) version. Alternative approaches can be used, such as multilevel block modelling (Žiberna, 2014) or dual projections (Everett & Borgatti, 2013). Citations can be considered as a non-stable relationship in comparison with institutional affiliation. Hence, the exploration is restricted to institutional affiliations as a proxy of social relationships, but further analyses should explore potential limitations. From the exploratory analysis, a first representation of the inter-organisational allowed representation of the connectedness of organisations (Laumann et al., 1978) between their co-members and consideration of their relational equivalence (White et al., 1976; Borgatti et al., 2018) using the core-periphery approach for organisations, as a structure previously identified by researchers in scientific networks (e.g., Breiger, 1976; Mullins et al., 1977).

4.4.2 Stationary Stochastic Actor-oriented Model

Table 15 presents the results to identify how stable micro-mechanisms and meso-level social forces operate cross-level mechanisms between actors and institutional affiliations. Additional model specifications are in the Appendix (Section K) that explores further specification that considers retrospective time and the usage of similarity measures in one of the effects controlled in the model. The stability of the patterns is considered as the overall tendencies towards stabilisations in this inter-organisational field.

The results show that different relational mechanisms simultaneously operate in the stationary and aggregated strong citation network. For instance, the *density* effect is negative ($\beta = -5.307, \mathrm{s.e.} = 0.246$), in which authors tend to cite fewer researchers than the available scholars in their personal networks. The researchers also reciprocated citations in the estimated year ($\beta = 2.353, \mathrm{s.e.} = 0.170$), which is interpreted as a tendency towards interpersonal intercitation (White et al., 2004). For peer recognition based on degree mechanisms, there is, on average, a more significant tendency between the personal networks against *indegree effect* ($\beta = -0.007, \mathrm{s.e.} = 0.016$) in the national bounded astronomical and astrophysical community. This tendency is often interpreted as the *Matthew effect* (Merton, 1968a), as researchers tend to be attracted to actors with more peer recognition. However, in this particular case, the personal networks lead to the inverse tendency when intercitation between local actors is considered. A different process is the outdegree popularity ($\beta = -0.616, \mathrm{s.e.} = 0.063$) as researchers tend to feel less attracted to cite researchers at a distance of two.

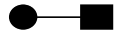
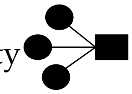
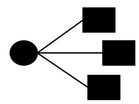
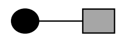
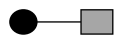
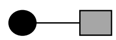
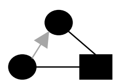
Table 15 Meta-analysis Results from Stationary Multilevel Models for Citation within Researchers and their Institutional Affiliation Networks.

		Est	SE	Σ	*Q*
Citation Network					
Outdegree (density)		-5.307**	0.246	0.000	7.680
Reciprocity		2.353**	0.170	0.000	7.519
Transitive triplets		0.187**	0.019	0.000	9.953
Transitive ties		3.414**	0.286	0.668	23.995*
Indegree popularity		-0.007	0.063	0.007	8.239
√Outdegree popularity		-0.616**	0.078	0.001	10.987
√Outdegree activity		0.199**	0.045	0.000	9.416
Reciprocity degree activity		-0.110 **	0.018	0.000	7.036
Closure by affiliation (ego)		-0.009	0.046	0.001	10.446
Closure by affiliation		0.019*	0.007	0.000	4.509
Co-citation from weak ties		0.009**	0.003	0.003	20.986
Accumulative citations (alter)		0.108**	0.019	0.036	15.631
Accumulative citations (ego)		-0.045*	0.015	0.000	4.314
Absolute difference of the accumulated number of citations		-0.120**	0.015	0.002	10.233

Note: For the figures representing the effects specified in the model, the circles are researchers, square are organisations, black ties are the dependent relations, and grey ties are the controlled ties.

Σ standard deviation, Q chi-squared test statistic. $^{*}p < .05$; $^{**}p < .001$;

Continuation

Institutional Affiliation					
Outdegree (density)		-1.184**	0.217	0.000	8.805
√Indegree popularity		0.261**	0.037	0.083	18.801*
Outdegree activity		-0.341**	0.055	0.000	3.167
Observatory (ref. University)		-0.069	0.043	0.000	7.725
Research Centre (ref. University)		0.147*	0.048	0.000	4.053
Size		0.134**	0.029	0.067	20.214*
Closure by association		0.310**	0.058	0.000	5.034

Note: For the figures representing the effects specified in the model, the circles are researchers, square are organisations, black ties are the dependent relations, and grey ties are the controlled ties.

Σ standard deviation, Q chi-squared test statistic. $^{*}p < .05$; $^{**}p < .001$;

A possible interpretation is that researchers might prefer to directly cite other researchers without referring to a potential intermediary author. The *outdegree activity effect* is positive ($\beta = 0.199, \text{s.e.} = 0.045$). Authors tend to cite others that are likewise citing other researchers from the same personal networks. Contrary to expectation, the *reciprocity degree activity* is negative ($\beta = -0.110, \text{s.e.} = 0.018$), as the tendency not to be attracted to reciprocating other researchers citing comparatively more than other researchers.

The peer effect or degree-based relational mechanisms in the institutional affiliation network should be interpreted differently. The density is also negative ($\beta = -1.184, \text{s.e.} = 0.217$), as authors tend to be less affiliated in the pool of organisations available. Institutions with more researchers are more attractive to researchers to participate ($\beta = 0.261, \text{s.e.} = 0.037$), but actors are less attracted to more institutional affiliations ($\beta = -0.341, \text{s.e.} = 0.055$).

A possible interpretation is that this might arise because of the connectedness with organisations representing different scientific niches in research centres ($\beta = 0.147, \mathrm{s.e.} = 0.048$), where researchers are selective in the tendency of multi-connectivity (Powell et al., 2005). Specifically, when the organisation is a research centre compared to universities, an author will have a 15.8% ($e^{0.147} - 1$) chance to be affiliated with these organisations if all other variables remain fixed[41]. The astronomical observatories are less significant and attractive for creating a tie than universities ($\beta = -0.069, \mathrm{s.e.} = 0.043$). This tendency might be expected because research centres act as intermediary hybrid entities (Etzkowitz & Leydesdorff, 2000) between universities and private institutions. Similarly, the centres are predominantly directed by researchers from the universities, and scholars infrequently have many affiliations.

Regarding *accumulative advantage* mechanisms in the citation network, there are differences in the tendency to cite other researchers from this community. From these results, the relative accumulative advantages (Rigney, 2010) in citations can be distinguished. From the perspective of the researchers, it is more attractive to cite other authors that have, on average, more accumulated citations ($\beta = 0.108, \mathrm{s.e.} = 0.019$) in the considered year. Recalling that the *indegree effect* in the personal network is negative, this tendency was interpreted as a predominance of external recognition instead of a local one. If the authors have comparatively more accumulated citations, on average, the tendency is the contrary ($\beta = -0.045, \mathrm{s.e.} = 0.015$) with less attraction for citing other researchers in the local community. If the absolute difference of the average accumulative number of citations between the focal actor and the other researchers is large, the tendency is negative ($\beta = -0.120, \mathrm{s.e.} = 0.015$). This heterogeneity gives insights into the inequality of recognition within these personal networks in the considered year. The results should be considered with caution because only the average citation of

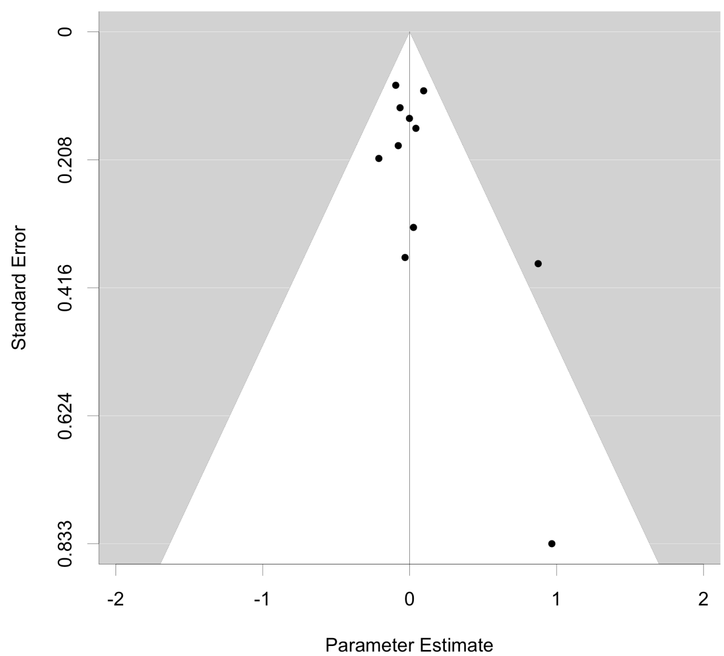
[41] The size of the coefficient should be read with caution as they are unstandardised and rarely independent because of the inherent interdependency of the networks. In the following, I am using the strategy suggested by Snijders, van de Bunt and Steglich (2010). A more suitable option is the relative importance of effects (Indlekofer & Brandes, 2013), which is currently not implemented for bipartite networks in RSiena.

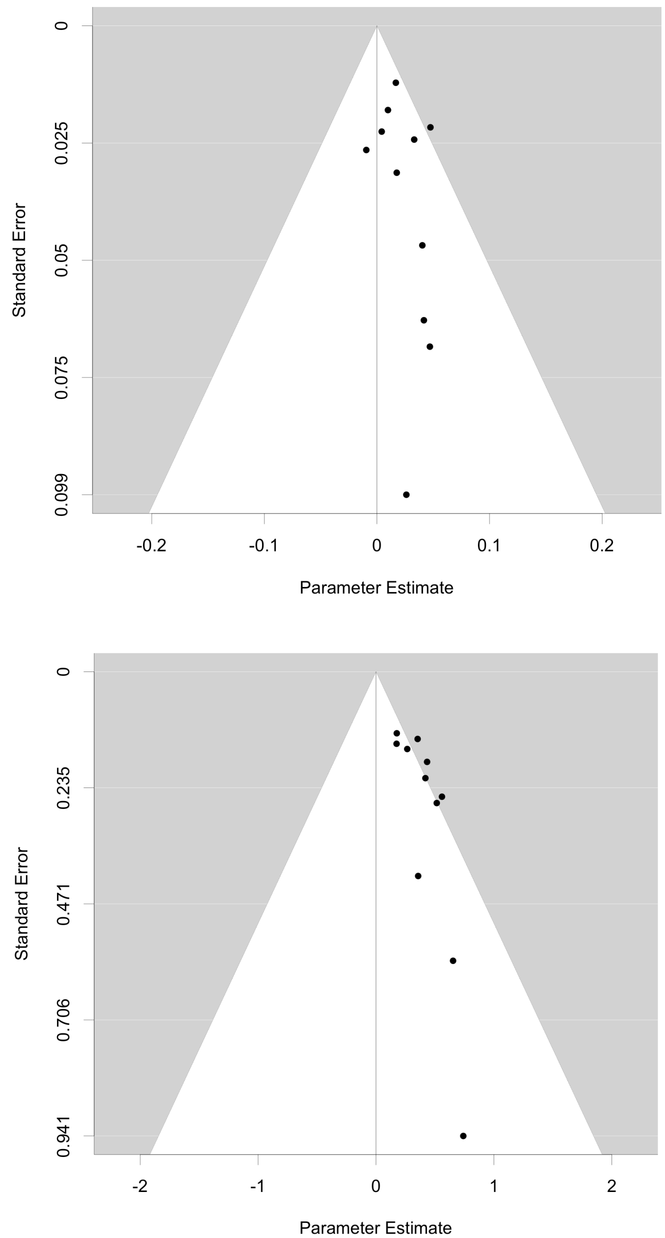
the authors' references from the analysed year and not the accumulated citations of the authors was considered in its academic trajectory.

There are different closure processes between researchers regarding *dyadic and triadic mechanisms* as local structures for group formation. The positive tendency of different effects together can be considered to indicate the tendency of scholars to form schools of thoughts in the community. In this perspective, 'Members of a social circle come together on the basis of their interests more often than on the basis of propinquity or ascribed status' (Crane, 1972: 14) in which indirect interactions are mediated through intervening parties. The influence of publications from authors they have never met is an important element to create the social closure that stabilises scientific communities. Consistent with previous research, if a researcher cites another author that is likewise citing a third author, the first researchers will cite the third author (β=0.187, s.e.=0.019), which is considered an interpersonal intercitation tendency. Considering the *effect of transitive ties* ($\beta = 3.414, \mathrm{s.e.} = 0.286$), when there are more direct and indirect ties between two researchers, there is a tendency to create a tie between these two researchers. The results should be interpreted cautiously because these parameters are more significantly different across personal networks ($Q = 23.995, \mathrm{Qp} = 0.008$). The dyadic covariate effect of *co-citation from weak ties* identifies a similar pattern. If two researchers are similarly and weakly co-cited more often ($\beta = 0.009, \mathrm{s.e.} = 0.003$), the researchers tend to cite the other author more often. Following the interpretation of White (2011), co-citation tends to characterise authors that are close 'intellectually' or reflect conflict or oppositions between them.

There is mixed support for the hypothesis under consideration for the meso-level social forces as a cross-level effect. A tendency for *closure by affiliation* (ego) as a propinquity tendency within the own organisation, reinforcing their inside ties leading to a positive feedback loop to promote internal recognition was expected. Based on the estimated specification, the parameter is less significant and negative ($\beta = -0.009, s.e. = 0.046$and $e^{-0.009}$~0.99 keeping all other variables fixed). For *closure by affiliation*, when an author citing another author shares more institutional affiliation with a third author, the results indicate that the first author tends to cite the third researcher ($\beta = 0.019, \text{s.e.} = 0.007$) between the different personal networks. The significance of this effect diminishes ($\beta = 0.016, \text{s.e.} = 0.009$) if a retrospective citation network is controlled as an additional rate function (Appendix, Section K). From the results, there is more support for *closure by association* in the tendency of actors citing other actors affiliated to an external organisation to be attracted to be affiliated to the same institution ($\beta = 0.310$, $s.e. = 0.058$). Shared cognitive interest as interpersonal *intercitation* ties corresponds to the researchers' embeddedness in similar external organisations, allowing multi-connectivity motivated by proximity tendencies.

Figure 30 Variations of the Personal Networks of the Organisations





a) Variation of *closure by affiliation (ego).* b) variation of *closure by affiliation*. c) variation of *closure by association.* Parameter estimated plotted against their standard errors, with line demarcating the region of significance ($\alpha = 0.05$).

More variations were expected in the personal networks' strategies from the funnel plots in Figure 30 in considering the second hypothesis. In these organisations, it was envisaged that the core will tend to be more endogamic and have more multi-connectivity among organisations than the actors on the periphery. There is little support for the core organisation having more endogamic ties. The only group (Figure 30a) in which the effect is more significant ($\beta = 0.87$, $s.e. = 0.38$) (separate models presented in the Appendix Section L) is the Center for Mathematical Modeling (CMM) at the University of Chile, an interdisciplinary centre focusing on basic and applied mathematics research for sciences, industry, and public policies. Members of CMM are not exclusively astronomers and astrophysicists. This tendency might be more prevalent in CMM because of this distinction, which corresponds with the increasing interest in astro-informatics as the combination of astronomy with other disciplines, such as statistics and the computational sciences (Espinosa-Rada et al., 2019).

Figure 30b on the funnel plot in the centre, shows UDP as the only organisation with a more significant tendency to *closure by affiliation* across different personal networks – an author that shares more institutional affiliation with a third author will be attracted to cite the third researcher ($\beta = 0.05$, $s.e. = 0.02$). *The Astronomy Nucleus at Universidad Diego Portales* (UDP) is a group of astronomers and astrophysics formed in 2013 focused on observational astronomy and observations at X-ray, optical, infrared, and submillimetre/radio wavelengths. UDP is not one of the organisations in the core of the core-periphery structure. However, it has connections with other organisations in the centre, according to Figure 29 (e.g., UCH, ESO, MAS, PUC, CTIO), and in this case, benefits from these connections.

There are some cases of *multi-connectivity* tendency, operationalised as *closure by association* (Figure 30c). Most of the organisations that were in the core of the exploratory analysis have a more significant effect in multi-connectivity, except for UdeC ($\beta = 0.36$, $s.e. = 0.41$) and PUC ($\beta = 0.65$, $s.e. = 0.59$). Some of these organisations benefits from their own extended opportunity structure. The effect is more significant considering the personal networks of ESO ($\beta = 0.35$, $s.e. = 0.14$) and UCH[42] ($\beta = 0.43$, $s.e. = 0.18$), and less significant in LCO ($\beta = 0.26$, $s.e. = 0.16$) and MAS ($\beta = 0.52$, $s.e. = 0.27$). Other cases from the periphery

[42] For the case of UCH, Guridi et al. mentioned that it ‘is the only university that can legally negotiate agreements with international observatories […] perpetuating Chile’s scientific hierarchy and generating inequalities within the local astronomy community’ (2020: 8) giving it a particular position that can be explored further.

in which the effect was prevalent were UDP ($\beta = 0.56,\ s.e. = 0.25$) and less significant, ULS ($\beta = 0.42,\ s.e. = 0.22$). The last two organisations have consolidated astronomical groups (CONICYT, 2012). ULS also held the La Serena Data Science Winter School, supported by the National Science Foundation and ANID for trained astronomers and data-driven sciences to prepare for the challenges of big data in astronomy due to the arrival of LSST. In other organisations on the periphery, the effect did not seem to be prevalent. In CMM ($\beta = 0.17,\ s.e. = 0.15$), members of this organisation do not exclusively work in astronomy and astrophysics because they are interdisciplinary communities focused on mathematics, engineering, and physics. The other cases are CTIO ($\beta = 0.18,\ s.e. = 0.12$), the *Cerro Tololo Inter-American Observatory*, and the UV ($\beta = 0.74,\ s.e. = 0.94$) university that is also on the periphery of the exploratory analysis.

4.5 Discussion

This research explores two mechanisms to identify the join patterns of *intercitations* among researchers in organisations – as meso level social forces – from a sample of personal networks from the organisations' perspective. These mechanisms enabled exploration of whether this community tended towards vertical endogamy or horizontal multi-connectivity of *intercitation* in this scientific community. The results indicate less evidence for endogamy in personal networks to reinforce their inside ties. Overall, there is more support in this community to share cognitive interest through *intercitation* when actors are connected to the same organisations. The connection through the same organisations allowed the study of multi-connectivity with these external organisations that are often research centres (i.e., social niches).

The diversity between organisations was the second focus of the analysis, considering the differences between incumbents in the core of the inter-organisational field and the peripheral actors. While endogamy was identified in an interdisciplinary organisation, the tendency for diversity and multi-connectivity among core actors, in comparison to peripheral actors, was more prevalent in the scientific community. This creates an inter-organisational core group that can be considered to push toward homogenisation through the mechanism of closure by association – the tendency of actors to cite other actors affiliated to an external organisation to be attracted to affiliation into the same institution. Some cases also in the core do not follow this pattern creating an internal differentiation. Other peripheral actors tend to have similar tendencies, which can be explored in further studies using longitudinal perspectives to identify whether the peripheral actors following these strategies converge with the core group of the inter-organisational structure or other actors in the core can create parallel core groups. This research focused the analysis on selection processes instead of influence (Steglich et al., 2010). Further research is needed to explore this possibility to identify, for example, if core or periphery organisations benefit from their own extended opportunity structure, and if so, under what conditions is detectable.

Most studies of scientific networks use a case study to emphasise the homogenisation of the entire structure. This research argued that combining different micro-mechanisms gives a broader perspective to understand the different scientific network processes. There is a variation in the different types of mechanisms, classified in different tendencies toward group formations between researchers creating invisible colleges, maintaining the Matthew effect and accumulative advantages. Less explored is the relevance of the multilevel perspective, which emphasises the meso-level social forces – as cross-level effects – that make identifying scientists' positions in the inter-organisational field possible. The multilevel perspective identified the dual position of researchers in organisations as a stable process within a scientific community, from a sample of cases to understand a specific population. Also, it enabled identification of the variation within these effects in different personal networks contrasting their disparity with the core-periphery inter-organisational context at the macro-level. This approach is suitable for identifying prominent incumbents or particularities that give a deeper comprehension of the environment of a scientific discipline in a local community that can be explored further and in more detail.

Finally, a *second-zone multilevel sampling was suggested from a second-mode focal actor* (i.e., the organisation in this case) to identify how stable the meso-level social forces between actors and institutional affiliations are in the researchers' tendency to create ties in a scientific community. This strategy proved suitable for this analysis, but more research should identify the consequence of potential effects at further distances. This strategy can orientate empirical research when more extensive networks are considered, facilitating statistical networks models that tend to have difficulties in their estimation, convergence, and reasonable goodness of fit when the network increases its size to more than a few hundred cases.

Using the stochastic actor-oriented model to analyse cross-sectional data is rare, giving an alternative to other methodological strategies already available, allowing multilevel networks and a sample of networks to be explored. A further expansion of this study should point to potential changes in the community and understand the development of the scientific fields considering newcomers. Consequently, because the study is a stationary network, it will benefit from an expansion with longitudinal data to understand the changes in the scientific community.

Acknowledgments

The author would like to thank members of the Mitchell Centre for Social Network Analysis at the University of Manchester for helpful comments and suggestions that considerably improved this work. Additionally, he acknowledges the assistance given by Research IT and the use of the Computational Shared Facility at The University of Manchester, and the assistance given by Daniel Perez from Epistemonikos Foundation in the data collection.

Financial Support

The study was supported by the Chilean *National Commission for Scientific and Technological Research* (ANID PFCHA/DOCTORADO BECAS CHILE/2017, grant number 72180242).

Conflict of interest

The authors have nothing to disclose.

Chapter 5.

Conclusion

In this chapter, I present some concluding remarks of the thesis. This thesis begins with the problem of exploring the 'true' underlying structure that was said to require a *multigraph* to describe the scientific networks (Chubin, 1976), and which often involves socio-cognitive networks (Merton, 2000; White et al., 2004; White, 2011) that conflate social, cognitive, and situational dimension in science. In this chapter, I first present the summary of the empirical findings of the three articles. Then, I emphasise the overarching contributions and some limitations of the thesis. Finally, I will provide some outlooks for further research.

5.1 Summary of the empirical findings

In this section, I present the summary of the empirical findings of the articles of the thesis. These articles explore the usage of *multigraphs* to explore *intercitations* – as the analysis of citations among a fixed population of authors that share a similar context (White, 2011) – in the Chilean astronomy and astrophysics community.

5.1.1 Study 1

In Chapter 1, it was mentioned that citation is a theoretical and empirical puzzling endeavour which requires further scrutiny. In Chapter 2, a methodological perspective was adopted to disentangle the main citation components when used as a *direct citation*, *bibliographic coupling*, or *co-citation*. The assumption to explore their differences was to enquire whether the usages of citation itself can derive in a *multigraph* in which the 'true' underlying structure can be discovered (Holland & Leinhardt, 1974). From a methodological perspective and in concordance with the interpretation that citation can be interpreted as a cognitive and a social element (Crane, 1972; Small, 1977), it is suggested that depending on the type of citation-based and the similarity measures, different dimensions are highlighted. The similarity measures that combine the different alternatives have a relevant role in the merged measure. When the merged measure used different similarities for the normalisation it will give more emphasis to the shared cognitive dimension according to the community's perception (i.e., bibliographic coupling or co-citation when *Jaccard* or *cosine* are used) or the authors' communicational trace (i.e., direct citation when *association strength* is considered) as a social element. This distinction might allow recovering the socio or cognitive dimension of the citation measure that are conflated.

5.1.2 Study 2

Following the discussion of Chapter 1, the citation can be further investigated and analysed considering the co-evolution of other networks (i.e., co-authorship, institutional affiliation, and publishing in scientific journals). In this article, the period after the arrival of the ALMA observatory was analysed. In the analysis, I concentrate on one specific type of mechanism that involves two different levels for *group formation* – closure by affiliation and closure by association as proximity mechanisms – to understand how a group of academics create interpersonal *intercitations* studying the co-evolution of a multilevel network. I investigate how well the micro-level represents macro features in a three-mode multilevel and multiplex network using the goodness of fit used in statistical models for social network analysis. The results indicate that social relationships grounded on scientific collaboration and space proximity based on institutional affiliation are more accurately suited to understand the co-evolution of the networks and the intercitation among astronomers than cognitive-based networks such as the journal network. Also, from a methodological perspective, the diagnostic allows identifying misspecifications or potential unobserved effects. The unobserved effects might be confounding other macro-structures that otherwise would be indistinguishable and that we might believe that they are substantively relevant to representing multilevel networks.

5.1.3 Study 3

Similar to the previous study, the third article (Chapter 4) continues the exploration of group processes (Chapter 1), identifying the patterns that are consistent in different personal networks (as a more ill-defined boundary) distinguishing between organisations that are in the core or periphery of the inter-organisational field of Chilean astronomy. This study explores how the regular join patterns of intercitations among researchers in organisations – as meso level social forces - vary within scientific communities and explore if the core organisations tend to have similar patterns compared to other institutions in the periphery. The data analysed relates to the state of the astronomical and astrophysics community during the same year in which the government of Chile shows interest in the development of astronomy to spur economic activity (Guridi et al., 2020) because of the upcoming arrival of the Vera C. Rubin Observatory (a.k.a., the Large Synoptic Survey Telescope [LSST]) (Espinosa-Rada et al., 2019). From the analysis, the results indicate that researchers in this community are not preserving endogamic intercitation - the tendency of citing researchers from the same organisation. However, there is a tendency upon intercitation among researchers affiliated in the same external research centres creating closure in scientific niches (i.e., research centres) as a community's tendency towards diversity and multi-connectivity. The closure by association is the tendency of actors citing other researchers to be attracted to share institutional affiliation in similar research centres. I also suggested a *second-zone multilevel sampling from a second-mode focal actor* – or extended opportunity structure of the organisations - as a strategy that allows identifying how stable the meso-level social forces between actors and institutional affiliations are in the researchers' tendency to create ties scientific community.

5.2 Overarching, contributions of the thesis and limitations

This thesis offers several contributions advancing our knowledge in using multigraphs to understand *intercitation* (theoretical background presented in

Chapter 1). The multigraphs used are multiplex networks based on citations (i.e., direct citation, bibliographic coupling, and co-citation) and co-authorship. And the multigraphs are also studied through multilevel networks (i.e., researchers, institutional affiliations, and journals). These different networks are used jointly to identify *intercitations* as the tendency of citations between a fixed group of researchers that share a similar context (White et al., 2004; White, 2011).

A first contribution is the exploration of the joint usage of these networks to understand *intercitation* through multiplex relationships of networks based only on citations, which combination was not sufficient to understand if the citation network is social or cognitive (Chapter 2). These networks based on citations are the co-citation and bibliographic coupling that rely on the perception of third authors and the direct citation that is considered as the author's communicational trace. As a limitation, the combination of these three measures (i.e., direct citation, bibliographic coupling, and co-citation) is not sufficient to understand whether the citation network is social or cognitive. The analysis reveals that the normalisation selection will emphasise the perception of third authors (i.e., *Jaccard* or *cosine*) or the direct communicational trace between researchers (i.e., *association strength*). In the first paper, three different strategies to normalise the weighted network are presented, which are the dichotomisation of the network (Breiger, 1974; Neal, 2014), other normalisation processes called the 'fractional approach' (Batagelj & Cerinšek, 2013; Perianes-Rodriguez et al., 2016; Leydesdorff & Park, 2016; Batagelj, 2020), and the simultaneous usage of two projections (Everett & Borgatti, 2013, 2018). The analysis only explores the most popular alternative often compared in scientometrics (Ahlgren et al., 2003; van Eck & Waltman, 2009; Egghe & Leydesdorff, 2009), which in practice can incorporate much more alternatives (e.g., Wasserman & Faust, 1994; Borgatti & Halgin, 2011a; Borgatti et al., 2018). Further work is needed to explore the merged measure with different databases, use other normalisations, and combine with other dimensions and/or methodologies.

The contribution of the second article expands our knowledge of *intercitation* with the analysis of cross-level mechanisms using multilevel and multiplex networks (Chapter 3). The cross-level effect that combines citation with social networks (i.e., co-authorship and institutional affiliation) shows to be more relevant to understand citations than publishing in the same journals (as a cognitive network). Two cross-level mechanisms explored are the closure by association and closure by affiliation as mechanisms that explain *group formation* from a multilevel perspective. Closure by association is the tendency of researchers citing another researcher who participates in an organisation or published journal to participate in the same organisation or journal. And the closure by affiliation is the tendency of a researcher that is participating in an organisation or publishing in a journal to cite other researchers that share the same organisation or publications. Combining the *multigraphs* (i.e., citation network with social networks) allows understanding the social dimension of citation when actors share a similar context or share similar ties. The astronomical community in this period was reasonably small (< 100 established scientists), in which the instances of evaluation of their works (through committees, funding available and access to the observational time), the newsletter or the 'white list' of SOCHIAS that give access to some of the most relevant telescopes worldwide allowed assuming that the actors are maintaining fluid communication. In smaller groups in science (~100), researchers are capable of absorbing technical information from a limited number of sources, when there is high specialisation, have similar problems, they tend to share costly component for their projects (Price, 1963: 83; Mullins & Mullins, 1973: 38; Mulkay et al., 1975: 188-189; Kuhn, 2012: 177) and were the stability and fluidity of information on time is often based on propinquity, this might be a plausible assumption, but that requires further analysis.

The *intercitation* is also explored in the third article by analysing the relative position of the organisations and authors in the inter-organisational field as a multilevel network (Chapter 4). The results expand our understanding of scientific fields in revealing that authors in core organisations, in comparison with peripheral organisations, have different patterns of citations. The consideration of organisations gives some evidence of the relevance of social dimensions (i.e., institutional affiliation) on citing in the context of this community. Most of the organisations that are in the core tend to create a *closure by association*. The *closure by association* is considered as the tendency of authors citing other authors affiliated to an external organisation to be attracted to become part of the same external organisations. These external organisations are often research centres created with public funding that focus on specific areas of knowledge. The results are consistent with the Chilean context because the local astronomical and astrophysical community, with the active interest of the Chilean government (Espinosa-Rada et al., 2019; Arancibia et al., 2020; Guridi et al., 2020), began the preparation for the arrival of a new class of telescopes (i.e., Vera C. Rubin Observatory) that will accumulate thousands of *petabytes* and will make available the astronomical information immediately. This Chilean astronomical community started the creation of different research centres to overcome the era of big data in astronomy (McCray, 2017; Hoeppe, 2014). One of the limitations of this analysis is that the data is cross-sectional. For further research, it is expected to trace the longitudinal aspect to reconstruct the history of the discipline due that the size of its members still allowed to have more detailed information.

Methodologically, this thesis contributes to some specific developments. Chapter 3 uses goodness of fit often used in the context of the stochastic actor-oriented model (Lospinoso & Snijders, 2019) for multiplex (i.e., overlapping multiplex triadic census and mixed layer triadic census) and three-mode multilevel networks (i.e., mixed degree distributions, mixed geodesic distance distribution and mixed quadrilateral census). The goodness of fit allows identifying miss specifications in complex networks, and further research should be done to expand these features. Some alternatives to expand the goodness of fit are multilevel networks with connections within and between two or more levels and incorporate two or more directed networks. Another contribution is presented in Chapter 4, using a methodological strategy named a *second-zone multilevel sampling from a second-mode focal actor* analysed through *meta-analysis* of *stationary stochastic actor-oriented models*. This strategy allows using simultaneously two levels to create samples. Similar strategies were used before considering one-level to estimate large networks (Stivala et al., 2016 for a review). In this regard, there is extended literature for samples of networks that should be explored further (e.g., Frank & Snijders, 1994; Giles & Handcock, 2010; Stivala et al., 2016). Hence, the specification of the *second-zone multilevel sampling from a second-mode focal actor* might help to identify the 'hard-to-reach' population, estimate the size of these networks, among others, as has been done in these areas of development.

One of the limitations of this thesis is the estimation of weighted ties (i.e., citation and co-authorship) when the stochastic actor-oriented model is used.

Previous studies that use a stochastic actor-oriented model to analyse scientific networks do not consider the weighted dimension (Ferligoj et al., 2015; Kronegger, 2012; Zinilli, 2016; Stark et al., 2020). A recent alternative is the network's dichotomisation analysed as a separate co-evolving network (Elmer et al., 2017) to overcome this limitation. As explored in the first study (Chapter 2), the decision to create the distinction requires a strong assumption for citation networks, and some sensitivity analysis explores this alternative in the second study adding the weak ties as a dyadic covariate in the model (Appendix, Section G of Chapter 3). In Chapter 4, the weighted networks are distinguished, separating weak and strong ties, in which only the strong ties are analysed. The incorporation of weights is a limitation of the model, and as far as I am aware, other available alternatives (e.g., Krivitsky, 2012; Krivitsky et al., 2020) do not implement the analysis of co-evolving multilevel networks yet. In this research, all the networks showed stability on time, but some models are worthy to explore further to disentangle the granularity of time (Butts, 2008; Stadtfeld & Block, 2017), which can be reasonable to capture the weight of the network but have the additional limitation of assuming that the ties are sequential. As far as I know, there is no ideal model to explore the multilevel co-evolutionary dimension of this network, and the *stochastic actor-oriented model* is one of the alternatives. Some of these issues should be explored further for modelling scientific networks.

5.3 Outlooks for future research

The studies presented in this thesis highlights some areas that might be worthy of exploring further in future research. In the previous section I concentrated on issues that appear in the empirical results that require further examination. In the following, some general directions are briefly presented.

5.3.1 Multi-methods

The first chapter presents some critics of the delimitations of boundaries investigated in scientific networks according to different methodological preferences. Previous work has combined different methodological strategies to explore scientific networks (e.g., Mullins et al., 1977; Lievrouw et al., 1987; Zuccala, 2006; Lazega et al., 2008; Milard, 2014; Raimbault & Joly, 2021). Further research can point in this direction to gain knowledge of the phenomenon, in which as one of the alternatives, bibliometric information can be mixed with surveys, interviews or ethnographic work that can enrich the analysis. A mixed perspective might contribute to understanding questions such as, what is at stake in scientific research? What are the controversies and struggles between researchers in different invisible colleges? How do collegiality and relational management of cooperation dilemmas among rivals shape scientific growths? Among others. For example, the perception of belongingness to groups or research areas from third authors can be identified through co-citation analysis or interviews, and other types of relationships can be further explored through surveys or interviews.

Specific scientific contexts and smaller networks are reasonable for a more detailed analysis of the evolution of scientific networks that have been less studied in recent years. Previous research mixed bibliometric indicators with questionaries to explore their connection in small groups, in which Breiger (1974) considered that at least within the core actors, they tend to be aware of each other and recognise their visibility. Mullins et al. (1977) also investigate the overlapping between co-citation associated with interpersonal networks. At least for the actors in the centre of the groups analysed, they were a reasonable level of awareness and social contact. Lievrouw et al. (1987), using a triangulation of data, also identifies that scientist working in the same research grant cluster appear as a part of relevant literature (co-citations). In their analysis, researchers exhibit a certain degree of colleagueship through their institutional affiliation and co-authorship patterns, enhanced through regular and more personal communication patterns that are different from the content of the work in which they engage (i.e., different *specialities* that are complimentary). Another interesting result from Lievrouw et al. is that even when researchers work are from different specialities, they ‘have to compete for financial, human, and clinical resources from a limited pool of funding sources, graduate programs, lipid clinics for patients, etc.’ (1987: 245), which is often a relevant aspect for the delimitation and consideration of national boundaries. Lazega et al. (2006) also select a group of highly productive actors of cancer researchers in France that were further interviewed using a 'linked-design' approach, which is one of the few studies that allowed tracing the intra-organisation and inter-organisational relationships among researchers at the same time. The combination of multi-methods seems promising, capable of capturing the formal and informal communication between the researchers and the content of the relationships, gaining a deeper understanding of their social relationships.

5.3.2 Multilevel and Multi-temporality

Another area of development is the study of multilevel and the multi-temporality involved in scientific networks. This thesis uses second 'level' entities such as organisations and journals to explore the scientific networks that tend to be highly institutionalised, and their stability can be assumed and measured with some models that assumed this stability (e.g., Snijders, 2001; Lusher et al., 2012). For example, the department in universities might take some time to vanish. The granularity of time from other second 'levels' requires further research instead, in which even when two or more levels can be nested does not imply that their co-evolution is symmetrical and in synchrony (Brailly et al., 2016). In certain circumstances, the social settings are less stable (e.g., conferences, attending meetings), and the iteration of interactions in this focus of activities can create stable relationships on time. For example, the activity itself can be restricted to short time-laps (e.g., one day workshop). Hence, the consideration of events in contraposition to stable relationships can be treated differently from a statistical perspective (e.g., Butts, 2008; Stadtfeld & Block, 2017). When the complexity of the entities involved in various social settings is considered, further exploration can address a sociological-theoretical framework that can be potentially developed and expressed into statistical models that can combine stable relationships and events.

The exploration and understanding of the multilevel and multi-temporal dimensions underpinning scientific networks require a deeper understanding. The argument in this thesis followed the exploration of multigraphs as the juxtaposition of multilevel and multiplex ties to overcome the 'structural confusion' of scientific networks. This thesis gives some insights into the interdependency of cognitive, social, and situational dimensions to identify the processes that allowed the development of science. A further inquiry is to consider other types of relationships and identify those interactions and events, which in combination can give a more detailed appreciation of the phenomenon under study. Further research might examine if unscrambling the granularity of the temporal aspect of social ties – as signals of communications, fortuitous interactions, or accidental encounters – can give a more precise understanding of the strands that allowed the connections in science.

Blank page

Appendix.

0. Observatories in Chile

Table 16 Observatories in Chile

District	Physical Location	Telescope	Type of Telescope	Commissioning telescope
Antofagasta	*Chajnantor Plateau*	Atacama Large Millimter/submillimeter Array (ALMA)	Radio Telescope	2011
		Cosmic Background Imager (CBI)	Radio Telescope	1999
		Atacama Pathfinder Experiment (APEX)	Radio Telescope	2005
		miniTAO Telescope	Optical / Infrared Telescopes	2010
		Cerro Chajnantor Atacama Telescope (CCAT)	Radio Telescope	**2023 (Projection)**
	Cerro Toco	Atacama Cosmology Telescope (ACT)	Radio Telescope	2007
		The POLARBEAR experiment	Radio Telescope	2011
	Pampa la Bola	Atacama Submillimeter Telescope Experiment (ASTE)	Radio Telescope	2002
		The NANTEN2 Submillimeter Observatory	Radio Telescope	2004

	Cerro Paranal	Very Large Telescope (VLT)	Optical / Infrared Telescopes	2002
		Visible and Infrared Survey Telescope for Astronomy (VISTA)	Optical / Infrared Telescopes	2009
		The VLT Survey Telescope (VST)	Optical / Infrared Telescopes	2011
		Extremely Large Telescope (E-ELT)	Optical / Infrared Telescopes	**2025 (Projection)**
Atacama	*Cerro Las Campanas*	Magellan Telescope	Optical / Infrared Telescopes	2000
		Du Pont Telescope	Optical / Infrared Telescopes	1977
		Swope Telescope	Optical / Infrared Telescopes	1971
		Giant Magellan Telescope	Optical / Infrared Telescopes	**2023 (Projection)**

Coquimbo	*Cerro Tololo*	Blanco 4.0 m Telescope	Optical / Infrared Telescopes	1974
		Small & Moderate Aperture Research Telescope System (SMARTS)		2003
		Las Cumbres Observatory (LCOGT)	Optical / Infrared Telescopes	2007
		PROMPT Telescope	Optical / Infrared Telescopes	2004
	Cerro Pachón	Gemini South Telescope	Optical / Infrared Telescopes	1983
		Southern Astrophysical Research Telescope (SOAR)	Optical / Infrared Telescopes	2004
		Large Synoptic Survey Telescope (LSST)	Optical / Infrared Telescopes	**2023 (projection)**

	Cerro La Silla	The 3.6m telescope	Optical / Infrared Telescopes	1977
		The New Technology Telescope (NTT)		1989
		2.2-m Max-Planck Telescope	Optical / Infrared Telescopes	1984
		1.2 Swiss Telescope	Optical / Infrared Telescopes	1998
		1.54-m Danish Telescope	Optical / Infrared Telescopes	1979
		1-m Schmidt Telescope	Optical / Infrared Telescopes	1971

B. Quadratic Assignment Procedure

Table 17 Regression Based on Quadratic Assignment Procedure Without Log Transformation

Dependent Network	Predictor	Weighted Jaccard		Cosine/ Ochiai		Association Strength	
		B ($SE\ B$)	β	B ($SE\ B$)	β	B ($SE\ B$)	β
	Intercept	0.031*** (0.000)	0.000	0.549*** (0.000)	0.000	0.011*** (0.000)	0.000
	Direct Citation	1.366*** (0.009)	0.395	1.718*** (0.015)	0.533	1.160*** (0.002)	0.938
	Bibliographic Coupling	1.015*** (0.007)	0.425	0.847*** (0.024)	0.500	1.227*** (0.010)	0.208
		R^2 =0.616		R^2 =0.566		R^2 =0.936	
	Intercept	0.031*** (0.000)	0.000	0.549*** (0.000)	0.000	0.010*** (0.000)	0.000
	Direct Citation	2.238*** (0.005)	0.648	1.184*** (0.017)	0.367	1.006*** (0.002)	0.815
	Co-citation	0.995*** (0.002)	0.629	0.878*** (0.026)	0.616	1.077*** (0.006)	0.318
		R^2 =0.951		R^2 =0.659		R^2 =0.977	
	Intercept	0.031*** (0.000)	0.000	0.549*** (0.000)	0.000	0.011*** (0.000)	0.000
	Bibliographic Coupling	1.593*** (0.004)	0.667	1.056*** (0.017)	0.623	0.946*** (0.015)	0.160
	Co-citation	1.017*** (0.003)	0.643	1.148*** (0.021)	0.804	2.144*** (0.009)	0.634
		R^2 =0.978		R^2 =0.917		R^2 =0.457	

Note: B for the unstandardised beta, $(SE\ B)$ for the standard error of the unstandardised beta, β for standardised beta, and $***\ p < 0.001$. The numbers of draws to use for the quantile estimation are 5,000.

C. Data Chilean Astronomers 2013-2015

Between May and June 2014, data collection took place and corrected, updated, and expanded until October 2019 (summarised in table 1). A list of all relevant researchers in university departments was created to gather the data. The list was created including members from research institutes that host astronomer and astrophysics academics in Chile during 2014 that have access to 10 per cent of the observation time of the astronomical facilities in the country. This percentage is not trivial because the Chilean astronomical community held in its territory some of the most relevant astronomical infrastructures of the world (such as VLT, ALMA, and soon the E-ELT and GMA) and will have 10 per cent of the LSST computer cluster in 2021. That represents near 70 per cent of the entire infrastructure on the earth.

Previous works have estimated the total size of the community (see Table 1). In this research, we identify the information of the cohort of 2013 corresponding to 87 astronomers in 10 institutions that were also an intermediating period of consolidation of the discipline because of the construction of the largest radio observatory at that time (the Atacama Large Millimetre/submillimeter Array [ALMA]). Then, we compare the data with a list created by the European Southern Observatory (ESO) to cross-validate the information that was compared with the 'white list' of the Chilean Astronomical Association (SOCHIAS) from 2016 to evaluate the consistency of the information. The 'white list' specifies who can apply for the observational time in the country.

Previous literature indicates that astronomy and astrophysics are isolated from other disciplines (Leydesdorff and Rafols, 2009; Jansen et al., 2010) and confirmed in the case of Chile (Cárdenas et al., 2015). Astrophysics and astronomy have very few citations outside their network of collaborators considering the citation pattern (Wallace et al., 2011).

Year	Estimation of the size of the community	Source
2000	21	Gibert, 2011
2005	40	López et al., 2005
2009	52	SOCHIAS* census
2011	80	Gibert, 2011
2012	75	SOCHIAS* census
2013	87	Espinosa-Rada, 2015
2016	114	SOCHIAS* census
2017	131	SOCHIAS* census
** SOCHIAS: Society of the Chilean Astronomical Association.*		

Table 18 Estimation of the Total Size of the Community of Astronomers and Astrophysics from Chile (Academics)

The websites of these institutes were reviewed, and all researchers, along with additional information about them, were entered into a database, using the available CV, the institutional site, their private homepages, among others. With this information, it was possible to track their academic trajectory (bachelor, master, PhD, post-doc, research visiting, academic position) and their institutional affiliation in each year of their academic path. According to the academic trajectory, the data has too much missing data corresponding to their specific academic position in each year (adjunct, associate, or full professor), and this data was not used in this analysis. Not all documents have information about the date of submission, revision, and correction of the papers. We ask a group of six astronomers of different universities for the time-lapse of this process. There was a consensus that each paper takes on average near two to six months to be published.

Another methodological choice was how to extract the core bibliometric information of the community. There is some consensus that SAO/NASA ADS (NASA Astrophysics Data System Abstract), the Web of Science (called Institute for Scientific Information before), and SCOPUS-Elsevier are the central databases for the astronomical community. It is unclear which one is the best option (Gómez and Mérida, 2007; Marra, 2014). ADS is probably the database that has the highest coverage of documents. Nonetheless, HLWIKI- CADA mention that 'Scopus covers mostly scientific fields; relatively weak in sociology, physics and astronomy', and ADS has not the same amount of information extraction than Web of Science and is tend to be recognised as a platform of 'minor products' (such as Google Scholar), even when some of their contributions might be highly recognised. The Web of Science was selected as the main database of the data gathering of the citations between the scientists.

For each academic enlisted before, we delimited the query used to search in Web of Science with the initial of the first name and the surname. To cover all possible outcomes. We also use different combinations of the names, for example, ‘M Hamuy’, ‘Mario Hamuy’, ‘Hamuy, M’, among others. Then, we export all the data for further analysis. Each paper collected was manually reviewed, contrasting the previous information collected for each author. We check the scientific discipline (we consider first the odds categories), the authors' full name, their institutional affiliation, the countries of the authors, and the co-author and citation references (looking for coherence in each case).

After gathering the information from Web of Science, and extra information of the trajectory of each of the astronomers (i.e., full names, year of PhD, academic degrees, institutional affiliations, gender and nationality), we extract the entire record of the publications in the SAO/NASA Astrophysics Data System (ADS) for each of the astronomers and astrophysicist. To extract the information from ADS, we use the data of the trajectory of each astronomer and manually contrast to identify if the output was coherent. In some cases, the astronomers and astrophysics provide their own list of publications using ADS. Some of them only enlist the referee publication. In those cases, we expand their list, updating the information to add the non-refereed publication and all the information until 2018. To query the data in the cases in which there was no list already available, we used the list of institutions in which the astronomers were affiliated and the scientists' full name.

With the information from WOS and ADS, we compare each of the paper in both databases and personal information from each of the astronomers to check for coherence. With the before mentioned information, we were able to create a list of the different types of combination of names of each scientist for disambiguation to re-run the query until saturation. In case of differences in the databases, each paper was manually reviewed. The results in both databases are different. For example, the documents on ADS record arXiv files, and in WOS presentations in conferences and publications of books are added more often than ADS. They are often the same data in both databases for referee documents, but they have some differences within conferences and not peer-reviewed publications.

Finally, for each paper, we manually gather and extract all the references that are indexed within the Web of Science for three years (from 2013 to 2015) and the complete information provided from the database. We do not have information on previous citations within the authors before 2013. We did an iterative process re-running the analysis until we had saturation and not new papers in which authors of this cohort were involved.

It was recorded 6,008 documents for the 87 astronomers in 10 Chilean Institutions from 1971 to 2017. WOS has a well-developed database actively used in scientometric studies and has relevant information about interdisciplinary fields comparable using the Journal Factor Impact (JIF). We also collected the disciplines of the journals if the journal is nationally based and the impact factor of the journals in the last five years. We gathered extra information about the institutions considering the type of organizations in which the astronomers were affiliated.

D. Descriptive and Change Statistics for Each Level

Citation Network			
Observation time	1	2	3
N of scientists	87	87	87
Density	0.103	0.100	0.117
Average degree	8.862	8.609	10.092
Number of ties	771	749	878
Collaboration Network			
Density	0.098	0.101	0.107
Average degree	8.437	8.713	9.218
Number of ties	367	379	401
Institutional Affiliation			
N of Organizations	13	15	17
Density	0.066	0.074	0.084
Average degree	1.195	1.333	1.517
Number of ties	104	116	132
Journals in the Web of Science			
N of Journals	25	25	25
Density	0.072	0.069	0.069
Average degree	2.437	2.345	2.345
Number of ties	212	204	204

Table 19 Descriptive Analysis of Each Network (2013-2015)

Table 20 Descriptive of Covariates of the Three Actors (Scientists, Universities and Journals)

Variable name	Mean	SD	Min.	Max.
Scientists				
Foreigner	62% (Chileans)		0	1
Citations accumulated (Wave 1)	2,848	3,364	9	18,171
Citations accumulated (Wave 2)	2,961	3,456	15	18,321
Citations accumulated (Wave 3)	3,045	3,521	15	18,667
Papers accumulated (Wave 1)	64.954	58.463	3	351
Papers accumulated (Wave 2)	70.931	62.680	5	378
Papers accumulated (Wave 3)	77.253	66.821	5	402
Age (First publication)	1995	9.660	1971	2009
Alphabetic Papers (Wave 1)	8.575	9.659	0	46
Alphabetic Papers (Wave 2)	8.839	9.814	0	46
Alphabetic Papers (Wave 3)	9.022	9.929	0	46
Nº First or Second Author (Wave 1)	13.724	11.267	0	53
Nº First or Second Author (Wave 2)	14.149	11.462	0	54
Nº First or Second Author (Wave 3)	14.586	11.662	0	54
Last Author (Wave 1)	15.046	17.381	0	74
Last Author (Wave 2)	16.276	18.373	0	85
Last Author (Wave 3)	17.793	19.636	0	99
Universities				
Type of Organization	66% (Universities)		0	1
Journals				
Astronomical Journal	62% (Astronomy)		0	1
Interdisciplinary Journal	9% (Interdisciplinary)		0	1
National based Journal	12% (Regional)		0	1
Impact factor (Wave 1)	3.681	6.566	0	40.783
Impact factor (Wave 2)	3.742	6.754	0	41.296
Impact factor (Wave 3)	3.782	6.594	0	41.458

In Table 20 the covariates are presented without centring, logarithmic modification (first or second authors, papers, citations, alphabetic papers, last author, impact factor journals) or temporal adjustments (age).

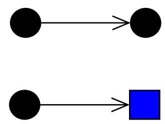
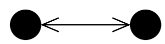
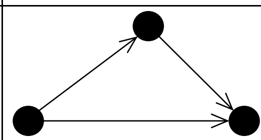
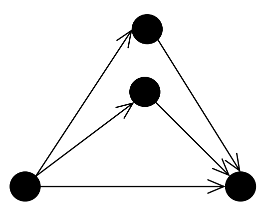
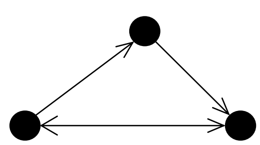
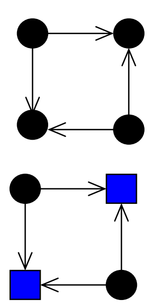
Citations (Network Dynamics)							
	Periods				Distance	Jaccard	Proportion between ties
	$0 \rightarrow 0$ (remain absent)	$0 \rightarrow 1$ (creation)	$1 \rightarrow 0$ (expiration)	$1 \rightarrow 1$ (maintained)			
$1 \rightarrow 2$	6,435	276	298	473	574	0.452	0.613
$2 \rightarrow 3$	6,316	417	288	461	705	0.395	0.615
Collaboration (Network Dynamics)							
$1 \rightarrow 2$	3,339	35	23	344	116	0.856	0.937
$2 \rightarrow 3$	3,316	46	24	355	140	0.835	0.937
Institutional Affiliation in Universities (Bipartite Network Dynamics)							
$1 \rightarrow 2$	1,442	20	8	96	17	0.774	0.923
$2 \rightarrow 3$	1,353	97	81	35	164	0.164	0.302
Journals in the Web of Science (Bipartite Network Dynamics)							
$1 \rightarrow 2$	2,667	79	87	125	166	0.430	0.590
$2 \rightarrow 3$	2,676	78	78	126	156	0.447	0.618

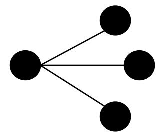
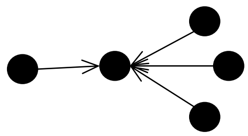
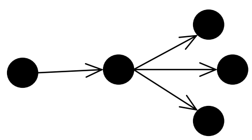
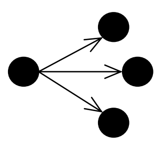
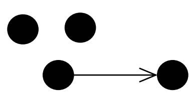
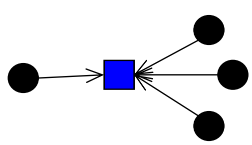
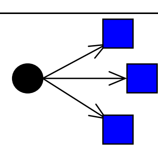
Table 21 Change Statistics for the Three Waves (2013-2015) of the Four Networks

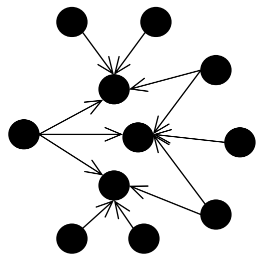
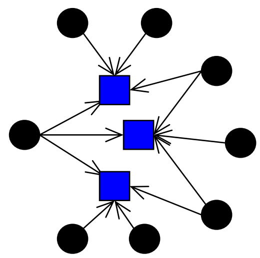
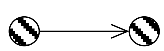
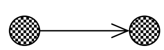
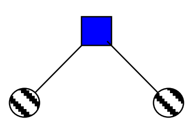
This network is under formation and growing (considering the number of ties for the institutional affiliation). As a complement of the Jaccard index as a measure of stability ($>$ 0.3), we also present the proportion between the ties presents at a given observation and the ties that remain in existence at the following observation ($\frac{N_{11}}{(N_{10}+N_{11})}$) in which values between 0.3 and 0.6 are still low but may still be acceptable (Snijders et al., 2010). For this particular network, the changes are due to the turnover of astronomers and astrophysics into four research centres (i.e., Millennium Institute of Astrophysics [MAS], The Center for Astroengineering of the Catholic University [AIUC], EMBIGGEN Anillo, the Center for Scientific Studies [CECS]), and an astronomical observatory (i.e., NRAO).

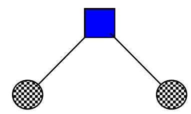
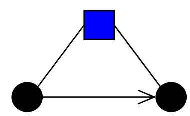
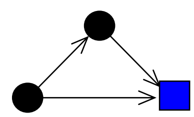
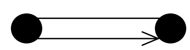
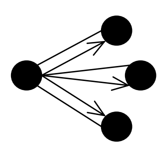
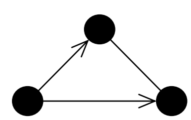
E. Effects

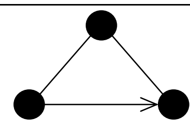
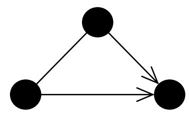
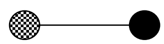
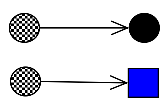
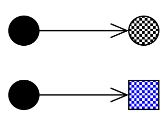
Table 22 Effects Incorporated in the Model Specifications

Name	Figure	Equation
Relational Mechanisms		
Density (Out-degree)		$s_{i1}^{X}(x,y) = \sum_{j} x_{ij}$ $s_{i1}^{Y}(x,y) = \sum_{k} y_{ik}$
Reciprocity		$s_{i2}^{X}(x,y) = \sum_{j} x_{ij}x_{ji}$
Transitive triad		$s_{i3}^{X}(x,y) = \sum_{j<h} x_{ij}x_{ih}x_{hj}$
Transitive ties		$s_{i4}^{X}(x,y) = \sum_{j} x_{ij}\max\,(x_{ih}x_{hj})$
Transitive reciprocated triad		$s_{i5}^{X}(x,y) = \sum_{j<h} x_{ij}x_{ji}x_{ih}x_{hj}$
4-cycle		$s_{i6}^{X}(x,y) = \frac{1}{4}\sum_{l\neq k\neq o} x_{il}\, x_{io}x_{kl}x_{ko}$ $s_{i2}^{Y}(x,y) = \frac{1}{4}\sum_{l\neq k\neq o} y_{il}\, y_{io}y_{kl}y_{ko}$

Degree		$s_{i7}^{X}(x,y) = \sum_{j} x_{ij}(\sqrt{x_{j+}} + \sqrt{x_{i+}})$
Indegree popularity		$s_{i8}^{X}(x,y) = \sum_{j} x_{ij} \sum_{h} x_{hj}$
Outdegree popularity		$s_{i9}^{X}(x,y) = \sum_{j} x_{ij} \sum_{h} x_{jh}$
Outdegree activity		$s_{i10}^{X}(x,y) = x_{i+}^{2}$
Outdegree at least one		$s_{i11}^{X}(x,y) = \min(x_{i+},1)$ In which, $\min(x_{i+},1) = \begin{cases} \min(x_{i+},1) = 0 \; if \; x_{i+} = 0 \\ \min(x_{i+},1) = 1 \; if \; x_{i+} \geq 1 \end{cases}$
Indegree popularity (square)		$s_{i3}^{Y}(x,y) = \sum_{k} y_{ik} \sqrt{\sum_{l} y_{lk}}$
Outdegree activity		$s_{i4}^{Y}(x,y) = y_{i+}^{2}$

Outdegree-indegree assortativity	 	$s_{i12}^{x}(x,y) = \sum_j x_{ij} x_{i+}^{1/2} x_{+j}^{1/2}$ $s_{i6}^{Y}(x,y) = \sum_j y_{ij} y_{i+}^{1/2} y_{+j}^{1/2}$
Dyadic similarity mechanisms		
Interaction of same covariate		$s_{i13}^{X}(x,y) = v_i \sum_j x_{ij} v_j$
Covariate ego x alter		$s_{i14}^{X}(x,y) = v_i \sum_j x_{ij} v_j$
Covariate alter at Z-distance of two		$s_{i7}^{Y}(x,y) = \sum_k y_{ik} \breve{v}_k^{Z}$

Similar covariate at Z-distance of two		$s^{Y}_{i8}(x,y) = \sum_{k} y_{ik}(sim(\breve{v}^{Z})_{ik} - \widehat{sim^{v}})$ Where $sim(\breve{v}^{W})_{ik}$ are defined as $sim(\breve{v}^{Z})_{ik} = \frac{\Delta - \lvert\breve{v}^{Z}_{i} - \breve{v}^{Z}_{k}\rvert}{\Delta}$ with $\Delta = max_{ij}\lvert v_i - v_k\rvert$ being the range of the covariate v observed.
Proximity mechanisms		
Closure by affiliation		$s^{XY}_{i1}(x,y) = \sum_{j \neq h} x_{ij} w_{ih} w_{jh}$
Closure by association		$s^{XY}_{i2}(x,y) = \sum_{k \neq l} x_{ij} w_{ih} x_{lj}$
Parallel networks		
Dyadic entrainment effect		$s^{X}_{i15}(x,y) = \sum_{k \neq l} x_{ij} w_{ij}$
Out-degree in W on X activity		$s^{X}_{i16}(x,y) = \sum_{j} x_{i+}(\sqrt{w_{i+}} - \sqrt{\overline{w}})$
Closure of mixed X-W two paths		$s^{X}_{i17}(x,y) = \sum_{k \neq l} x_{ij} x_{ih} w_{hj}$

Closure by affiliation		$s_{i18}^{X}(x,y) = \sum_{j \neq h} x_{ij} w_{ih} w_{jh}$
Closure by association		$s_{i19}^{X}(x,y) = \sum_{k \neq l} x_{ij} w_{ih} x_{lj}$
Control		
Covariate		$s_{i10}^{Y}(x,y) = \sum_{j} x_{ij}(v_i + v_j)$ $s_{i11}^{Y}(x,y) = \sum_{j} x_{ij}(v_i + v_j)^2$
Ego covariate		$s_{i20}^{X}(x,y) = v_i x_{i+}$ $s_{i12}^{Y}(x,y) = v_i y_{i+}$
Alter covariate		$s_{i21}^{X}(x,y) = \sum_{j} x_{ij} v_j$ $s_{i13}^{Y}(x,y) = \sum_{k} y_{ik} v_k$
Note: Considering w as a tie to another network of actor i in the X network, Z is the distances between actors, and V are the attributes of the actors. Blue represents node of a different network, and square are nodes of a different mode. Arrow lines are for directed networks, and straight lines are for undirected networks. More details of these effects in Ripley et al. (2021).		

F. Goodness of Fit

Table 23 Goodness of Fit of the Models

Goodness of fit	Full Model	Multiplex Model	Multilevel Model	Baseline Model
Relational Mechanisms Citation Network				
Indegree distribution	0.006	0.001	0.192	0.095
Outdegree distribution	0.132	0.105	0.029	0.012
Triad census	0.810	0.641	0.062	0.029
Geodesic distribution	0.102	0.096	0.123	0.102
Clique distribution	0.656	0.593	0.636	0.535
Eigenvalue distribution	0.476	0.454	0.259	0.209
Relational Mechanisms Collaboration Network				
Degree distribution	0.300	0.309	0.320	0.401
Triad census	0.997	1.000	0.992	0.998
Geodesic distribution	0.330	0.260	0.474	0.431
Clique distribution	0.144	0.142	0.094	0.105
Eigenvalue distribution	0.666	0.649	0.715	0.710
Similarity Based Mechanisms Citation Network				
E-I index	0.114	0.092	0.106	0.05
Yule-Q	0.348	0.314	0.459	0.368
Similarity Distribution	0.037	0.035	0.074	0.079
Average Euclidian	0.387	0.518	0.476	0.636
Constraint	0.829	0.807	0.934	0.908
Effective size	0.241	0.199	0.231	0.227
Similarity Based Mechanisms Collaboration Network				
E-I index	0.504	0.497	0.412	0.437
Yule-Q	0.485	0.516	0.361	0.382
Similarity Distribution	0.057	0.06	0.079	0.095
Average Euclidian	0.015	0.016	0.024	0.021
Constraint	0.126	0.169	0.116	0.166
Effective size	0.633	0.532	0.616	0.533

Continuation

Relational Based Mechanisms Institutional Affiliation Network				
Outdegree distribution	0.097	0.111	0.107	0.090
Indegree distribution	0.425	0.407	0.393	0.394
Triad census	1.000	1.000	1.000	1.000
Clique	1.000	1.000	1.000	1.000
Relational Based Mechanisms Journals of the Web of Science				
Outdegree distribution	0.237	0.223	0.234	0.236
Indegree distribution	0.173	0.149	0.164	0.158
Triad census	0.907	0.923	0.902	0.913
Clique	0.030	0.025	0.030	0.037
Relational Based Mechanisms for Multiplex Networks (Citation and Collaboration)				
Mixed Indegree distribution	0.042	0.045	0.323	0.466
Mixed outdegree distribution	0.128	0.17	0.136	0.212
Mixed triad census	0.815	0.707	0.498	0.477
Mixed geodesic distribution	0.058	0.052	0.047	0.053
Relational Based Mechanisms for Multiplex Networks (Collaboration and Citation)				
Mixed triad census	0.756	0.67	0.981	0.951
Mixed Indegree distribution	0.096	0.048	0.161	0.077
Mixed outdegree distribution	0.022	0.006	0.100	0.088
Mixed geodesic distribution	0.066	0.060	0.401	0.245
Relational Based Mechanisms for Multiplex Networks (Citation and Collaboration)				
Overlapping triadic census	0.027	0.013	0	0
Overlapping triadic census (extended)	0.002	0.001	0	0
Proximity Based Mechanisms for Multilevel Networks (Citation and Institutional Affiliation)				
Mixed Indegree distribution	0.009	0.002	0.181	0.100
Mixed outdegree distribution	0.176	0.093	0.031	0.011
Mixed triad census	0.412	0.203	0.302	0.078
Mixed geodesic distribution	0.169	0.065	0.213	0.083

Continuation

Proximity Based Mechanisms for Multilevel Networks (Collaboration and Institutional Affiliation)				
Mixed Indegree distribution	0.474	0.713	0.584	0.819
Mixed outdegree distribution	0.926	0.950	0.919	0.960
Mixed triad census	0.977	0.974	0.931	0.972
Mixed geodesic distribution	0.288	0.236	0.290	0.182
Proximity Based Mechanisms for Multilevel Networks (Citation and Journals of the Web of Science)				
Mixed Indegree distribution	0.058	0.03	0.41	0.263
Mixed outdegree distribution	0.951	0.853	0.914	0.721
Mixed triad census	0.856	0.664	0.811	0.568
Mixed geodesic distribution	0.166	0.121	0.173	0.143
Proximity Based Mechanisms for Multilevel Networks (Collaboration and Journals of the Web of Science)				
Mixed Indegree distribution	0.519	0.514	0.555	0.538
Mixed outdegree distribution	0.881	0.777	0.892	0.734
Mixed triad census	0.873	0.819	0.875	0.797
Mixed geodesic distribution	0.002	0.003	0.003	0.001
Proximity Based Mechanisms for Multilevel Networks (Citations, Institutional Affiliation and Journals of the Web of Science)				
Mixed Indegree distribution	0.057	0.017	0.379	0.196
Mixed outdegree distribution	0.655	0.475	0.510	0.314
Mixed quadratic census	0.590	0.339	0.449	0.147
Mixed geodesic distribution	0.282	0.272	0.300	0.308
Proximity Based Mechanisms for Multilevel Networks (Collaboration, Institutional Affiliation and Journals of the Web of Science)				
Mixed Indegree distribution	0.953	0.981	0.967	0.984
Mixed outdegree distribution	0.740	0.697	0.775	0.703
Mixed quadratic census	0.975	0.953	0.960	0.944
Mixed geodesic distribution	0.298	0.312	0.305	0.297
Relational Based Mechanisms for Multiplex Networks (Citation and Collaboration)				
Mixed indegree distribution (without normalization)	0.058	0.009	0.131	0.053
Mixed outdegree distribution (without normalization)	0.760	0.768	0.792	0.709

Continuation

Proximity Based Mechanisms for Multilevel Networks (Citation and Institutional Affiliation)				
Mixed indegree distribution (without normalization)	0.030	0.008	0.298	0.186
Mixed outdegree distribution (without normalization)	0.724	0.653	0.393	0.205
Proximity Based Mechanisms for Multilevel Networks (Citation and Journals of the Web of Science				
Mixed indegree distribution (without normalization)	0.311	0.301	0.335	0.346
Mixed outdegree distribution (without normalization)	0.399	0.392	0.383	0.382
Proximity Based Mechanisms for Multilevel Networks (Citations, Institutional Affiliation and Journals of the Web of Science)				
Mixed indegree distribution (without normalization)	0.217	0.222	0.247	0.260
Relational Based Mechanisms for Multiplex Networks (Collaboration and Citation)				
Mixed indegree distribution (without normalization)	0.025	0.014	0.074	0.065
Mixed outdegree distribution (without normalization)	0.276	0.230	0.753	0.768
Proximity Based Mechanisms for Multilevel Networks (Collaboration and Institutional Affiliation)				
Mixed indegree distribution (without normalization)	0.279	0.298	0.341	0.366
Mixed outdegree distribution (without normalization)	0.680	0.678	0.714	0.769
Proximity Based Mechanisms for Multilevel Networks (Collaboration and Journals of the Web of Science)				
Mixed indegree distribution (without normalization)	0.384	0.421	0.425	0.526
Mixed outdegree distribution (without normalization)	0.724	0.700	0.687	0.567
Proximity Based Mechanisms for Multilevel Networks Collaboration, Institutional Affiliation and Journals of the Web of Science)				
Mixed outdegree distribution (without normalization)	0.753	0.696	0.707	0.612

G. Exploratory Analysis using Weighted Networks

Some models are further estimated in this section (Table 24) to explore weighted networks using strong ($> median(X)$) and weak ties ($< median(X)$) incorporated into the model specification as a dyadic covariate. Each model is similar to the results presented in the main section, but the full model and the multiplex did not achieve convergence. In this section, Model 1 is similar to the Multilevel Model adding co-citation (weak ties), in which authors are perceived as cognitive similar by a third party (White, 2003). Co-citation tends to represent authors that are close 'intellectually' or reflect conflict or oppositions between them (White, 2011). Collaboration (weak ties) is the tendency to have from more collaboration to strong collaboration, and the transitivity from the weak collaboration is the tendency to create a strong collaboration with an author if they both share co-authors. Model 2 adds the same dyadic covariates that Model 1 but does not consider the multilevel effects. Model 3 do not estimate the multilevel network but incorporate for the focal actor the relative position of the authors as a covariate such as the number of times the astronomers were first author, second author, and last author in their accumulated papers. This model also incorporates the number of times the astronomers published in papers that were alphabetically ordered. Model 4 also adds the authors' relative position and the alphabetical order, adding if other actors had these attributes and the homophily tendency for each covariate.

Most of the parameters in Model 1, in comparison with the Multilevel Model presented in the main section, present some difference. Model 1 has less level of significance of some parameters, but overall, there are fewer changes in the direction of the coefficients. These differences can be associated with the inclusion of new and more parameters, adding some insights into the differences between weak and strong ties. The main changes are the citation accumulated similarity and age similarity, nationality, and homophily of the nationality for the collaboration network. For multilevel effects, the co-authorship to journal agreement for the journal network and the closure by affiliation in the collaboration (institution) network also change their direction. All the effects before mentioned above change their directions when only strong ties are considered and are less significant. These models seem promising and allowed to use of weighted networks. Further analysed should be considered.

		Model 1		Model 2		Model 3		Model 4	
Citation	Rate (period 1)	2.10	0.46	2.18	0.50	2.19	0.51	2.18	0.48
Network	Rate (period 2)	2.99	0.65	3.30	0.74	3.32	0.75	3.24	0.70
	Rate indegree	0.12***	0.04	0.12 **	0.04	0.11 **	0.04	0.11 **	0.03
	RM: Outdegree (density)	-4.66 ***	1.15	-3.78 ***	1.06	-4.80 ***	1.11	-5.08 ***	1.17
	RM: Reciprocity	3.48 ***	0.59	2.99 ***	0.35	3.04 ***	0.35	3.03 ***	0.37
	RM: Transitive triplets	0.32 †	0.19	0.40 *	0.18	0.37 *	0.17	0.38 *	0.18
	RM: Transitivity reciprocated triplets	-0.53 *	0.27	-0.61 *	0.24	-0.56 *	0.23	-0.57 *	0.24
	RM: Transitive ties	1.57 ***	0.27	1.43 ***	0.26	1.54 ***	0.27	1.56 ***	0.27
	RM: sqrt(indegree) (popularity)	0.06	0.65	0.06	0.59	0.23	0.60	0.36	0.61
	RM: sqrt(outdegree) (popularity)	-0.29	0.25	-0.23	0.22	-0.27	0.23	-0.29	0.25
	RM: sqrt(outdegree) (activity)	0.52	0.61	0.52	0.55	0.85	0.57	0.99 †	0.59
	RM: Outdegree at least one	-0.92	0.57	-1.14 *	0.56	-0.45	0.65	-0.51	0.67
	RM: Assortativity	-0.40	0.31	-0.39	0.28	-0.50 †	0.29	-0.57 †	0.31
	RM: 4-cycles	0.07 ***	0.02	0.06 ***	0.02	0.06 ***	0.02	0.07 ***	0.02
	RM: Co-citation (weak ties)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	C: Nationality alter (1=Chilean)	-0.22	0.22	-0.10	0.21	-0.02	0.21	0.03	0.24
	C: Nationality ego (1=Chilean)	-0.64*	0.25	-0.42 †	0.22	-0.08	0.26	-0.05	0.26
	DSM: Nationality ego x Nationality alter	-0.61 †	0.35	-0.47	0.33	-0.35	0.34	-0.26	0.34
	C: Citations alter	0.55 ***	0.13	0.55 ***	0.12	0.59 ***	0.12	0.56 ***	0.13
	C: Citations ego	0.42 **	0.15	0.44 **	0.15	0.50 **	0.17	0.49 **	0.17
	DSM: Citations accumulated similarity	-0.06	0.10	-0.04	0.09	-0.03	0.10	0.01	0.11
	DSM: Age similarity (year first paper)	0.05	0.07	0.05	0.07	0.06	0.07	0.01	0.09
	DSM: Papers accumulated similarity	-0.05	0.18	-0.11	0.17	-0.14	0.18	-0.24	0.27
	C: Time	-0.95 ***	0.23	-0.91 ***	0.20	-1.00 ***	0.22	-0.99 ***	0.23

Note: RM: Relational mechanisms; RMM: Relational multilevel mechanisms; DSM: Dyadic similarity mechanisms; C: Control; PM: Proximity mechanisms. To control for time heterogeneity, we add a linear time variable (Snijders & Lospinoso, 2019). † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;

Continuation		Model 1		Model 2		Model 3		Model 4	
Citation Network	C: Alphabetic order alter							-0.23	0.14
	C: Alphabetic order ego					-0.48 *	0.21	-0.44 *	0.20
	DSM: Alphabetic order ego x Alphabetic order alter							0.06	0.09
	C: Last position alter							-0.01	0.18
	C: Last position ego					-0.33	0.20	-0.39 †	0.21
	C: Last position ego x Last position alter							0.16	0.11
	C: First author alter							0.37	0.26
	C: First author ego					0.63 *	0.30	0.77 *	0.32
	DSM: First author ego x First author alter							-0.35	0.23
	C: Second author alter							0.05	0.19
	C: Second author ego					0.37 †	0.23	0.34	0.23
	DSM: Second author ego x Second author alter							0.01	0.13

Note: RM: Relational mechanisms; RMM: Relational multilevel mechanisms; DSM: Dyadic similarity mechanisms; C: Control; PM: Proximity mechanisms. To control for time heterogeneity, we add a linear time variable (Snijders & Lospinoso, 2019). † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;

Continuation		Model 1		Model 2		Model 3		Model 4	
	PM: Affiliation closure (Institutions)	1.56***	0.34						
	PM: Affiliation closure (Journals)	0.21	0.15						
	PM: Reciprocity X Affiliation closure (Institutions)	-0.83	0.62						
	PM: Reciprocity X Affiliation closure (Journals)	-0.14	0.26						
Collaboration	Rate (period 1)	9.41	4.18	9.46	4.60	9.55	4.44	9.55	4.38
	Rate (period 2)	7.68	3.52	7.41	2.20	7.42	2.47	7.46	2.70
	RM: Outdegree (density)	-5.11 ***	0.76	-5.10 ***	0.70	-5.10 ***	0.72	-5.12 ***	0.64
	RM: Transitive triads	0.95 *	0.38	0.96 **	0.34	0.95 **	0.36	0.94 *	0.37
	RM: Transitivity ties	0.96	0.68	0.93	0.64	0.93	0.71	0.99	0.71
	RM: Degree	0.06	0.24	0.08	0.21	0.08	0.22	0.08	0.20
	C: Collaboration (weak ties)	-0.67 ***	0.16	-0.68 ***	0.16	-0.67 ***	0.16	-0.67 ***	0.17
	C: Transitivity (weak collaboration)	0.04 ***	0.01	0.04 ***	0.01	0.04 ***	0.01	0.04 ***	0.01
	C: Nationality (1=Chilean)	0.10	0.23	0.09	0.21	0.09	0.22	0.09	0.21
	DSM: Nationality ego x Nationality alter	0.32	0.55	0.30	0.53	0.33	0.54	0.32	0.54
	C: Citations	0.12	0.09	0.12	0.08	0.12	0.08	0.12	0.08
	DSM: Citations accumulated similarity^2	-0.01	0.03	-0.01	0.03	-0.01	0.03	-0.01	0.03
	PM: Affiliation closure (Institutions)	-0.01	0.65						
	PM: Affiliation closure (Journals)	0.06	0.22						

Note: RM: Relational mechanisms; RMM: Relational multilevel mechanisms; DSM: Dyadic similarity mechanisms; C: Control; PM: Proximity mechanisms. To control for time heterogeneity, we add a linear time variable (Snijders & Lospinoso, 2019). † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;

Continuation		Model 1		Model 2		Model 3		Model 4	
Institutional Affiliation	Rate (period 1)	0.24	0.06	0.24	0.06	0.24	0.06	0.24	0.06
	Rate (period 2)	4.67	0.70	4.65	0.64	4.64	0.65	4.65	0.66
	RM: Outdegree (density)	-0.99 *	0.46	-0.99 *	0.47	-0.99 *	0.46	-1.00 *	0.47
	RM: Indegree (popularity)	0.11 **	0.04	0.11 **	0.04	0.11 **	0.04	0.11 **	0.04
	RM: Outdegree (activity)	0.36 *	0.15	0.36 *	0.15	0.36 *	0.16	0.36 *	0.16
	RM: Assortativity	-0.36 *	0.18	-0.36 †	0.19	-0.36 †	0.19	-0.36 †	0.19
	C: Type of Organisation (1=University)	0.04	0.18	0.04	0.18	0.04	0.19	0.04	0.18
	DSM: Citations accumulated similarity	2.11 *	0.83	2.13 *	0.85	2.10 *	0.82	2.10 *	0.87
Journals in the WOS	Rate (period 1)	4.05	0.51	4.07	0.51	4.07	0.51	4.07	0.50
	Rate (period 2)	3.89	0.51	3.87	0.52	3.86	0.50	3.87	0.51
	RM: Outdegree (density)	-3.50 ***	0.22	-3.56 ***	0.22	-3.55 ***	0.22	-3.55 ***	0.22
	RM: Cycle of fourth	0.01 *	0.01	0.01 **	0.00	0.02 **	0.00	0.02 **	0.00
	RM: sqrt(indegree) (popularity)	0.40 ***	0.05	0.43 ***	0.04	0.43 ***	0.05	0.42 ***	0.05
	RM: Outdegree (activity)	0.08 **	0.03	0.08 **	0.03	0.08 **	0.03	0.08 **	0.03
	C: Interdisciplinary Journal (1=Interdisciplinary)	0.02	0.29	0.01	0.28	0.01	0.28	0.01	0.29
	C: National-based journal (1=National)	-0.31	0.36	-0.31	0.36	-0.30	0.35	-0.31	0.36
	C: Astronomical journal (1=Astronomy)	0.27 *	0.13	0.26 *	0.13	0.26 *	0.13	0.26 *	0.13
	C: Impact Factor	0.01	0.10	0.02	0.09	0.02	0.09	0.02	0.09
	DSM: Citations accumulated similarity	-0.19	0.48	-0.04	0.46	-0.03	0.45	-0.04	0.46
	RMM: Citation to journal agreement	0.01	0.04						
	RMM: Collaboration to journal agreement	0.41	0.31						

Note: RM: Relational mechanisms; RMM: Relational multilevel mechanisms; DSM: Dyadic similarity mechanisms; C: Control; PM: Proximity mechanisms. To control for time heterogeneity, we add a linear time variable (Snijders & Lospinoso, 2019). † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;

Table 24 SAOM Models (Weighted) for the Evolution of the Citation Network, Collaboration Network, Scientists Affiliated with Institutions and Scientists Publishing in Journals from the Web of Science

H. Goodness of Fit

Table 25 Goodness of Fit of the Models (Weighted)

Goodness of fit	Full Model	Multiplex Model	Multilevel Model	Baseline Model
Relational Mechanisms Citation Network				
Indegree distribution	0.720	0.638	0.709	0.709
Outdegree distribution	0.404	0.383	0.487	0.49
Triad census	0.002	0.001	0.001	0.002
Geodesic distribution	0.898	0.800	0.807	0.804
Clique distribution	0.427	0.333	0.388	0.433
Eigenvalue distribution	0.097	0.076	0.106	0.107
Relational Mechanisms Collaboration Network				
Degree distribution	0.129	0.129	0.126	0.145
Triad census	0.999	0.999	0.999	0.998
Geodesic distribution	0.006	0.008	0.008	0.010
Clique distribution	0.737	0.728	0.703	0.720
Eigenvalue distribution	0.979	0.976	0.975	0.974
Similarity Based Mechanisms Citation Network				
E-I index	0.292	0.275	0.300	0.326
Yule-Q	0.252	0.169	0.214	0.240
Similarity Distribution	0.107	0.068	0.031	0.076
Average Euclidian	0.428	0.276	0.15	0.305
Constraint	0.019	0.009	0.014	0.017
Effective size	0.582	0.616	0.601	0.616
Similarity Based Mechanisms Collaboration Network				
E-I index	0.412	0.39	0.365	0.395
Yule-Q	0.989	0.987	0.982	0.983
Similarity Distribution	0.797	0.776	0.79	0.796
Average Euclidian	0.613	0.616	0.636	0.633
Constraint	0.021	0.019	0.019	0.022
Effective size	0.356	0.39	0.376	0.391

Continuation

Relational Based Mechanisms Institutional Affiliation Network				
Outdegree distribution	0.094	0.092	0.093	0.095
Indegree distribution	0.408	0.417	0.399	0.416
Triad census	1.000	1.000	1.000	1.000
Clique	1.000	1.000	1.000	1.000
Relational Based Mechanisms Journals of the Web of Science				
Outdegree distribution	0.268	0.24	0.232	0.235
Indegree distribution	0.15	0.158	0.163	0.182
Triad census	0.918	0.919	0.92	0.906
Clique	0.027	0.03	0.03	0.039
Relational Based Mechanisms for Multiplex Networks (Citation and Collaboration)				
Mixed Indegree distribution	0.975	0.973	0.975	0.986
Mixed outdegree distribution	0.652	0.632	0.612	0.703
Mixed triad census	0.716	0.718	0.712	0.689
Mixed geodesic distribution	0.022	0.018	0.017	0.016
Relational Based Mechanisms for Multiplex Networks (Collaboration and Citation)				
Mixed triad census	0.984	0.992	0.988	0.988
Mixed Indegree distribution	0.222	0.194	0.199	0.23
Mixed outdegree distribution	0.973	0.985	0.959	0.958
Mixed geodesic distribution	0.816	0.832	0.914	0.914
Relational Based Mechanisms for Multiplex Networks (Citation and Collaboration)				
Overlapping triadic census	0.072	0.052	0.058	0.063
Overlapping triadic census (extended)	0.046	0.026	0.035	0.039
Proximity Based Mechanisms for Multilevel Networks (Citation and Institutional Affiliation)				
Mixed Indegree distribution	0.732	0.757	0.756	0.752
Mixed outdegree distribution	0.958	0.971	0.970	0.971
Mixed triad census	0.152	0.615	0.545	0.529
Mixed geodesic distribution	0.252	0.25	0.285	0.283

Continuation

Proximity Based Mechanisms for Multilevel Networks (Collaboration and Institutional Affiliation)				
Mixed Indegree distribution	0.780	0.792	0.790	0.794
Mixed outdegree distribution	0.628	0.641	0.601	0.619
Mixed triad census	0.970	0.977	0.977	0.977
Mixed geodesic distribution	0.171	0.17	0.166	0.182
Proximity Based Mechanisms for Multilevel Networks (Citation and Journals of the Web of Science)				
Mixed Indegree distribution	0.733	0.677	0.707	0.687
Mixed outdegree distribution	0.461	0.443	0.413	0.429
Mixed triad census	0.97	0.883	0.864	0.879
Mixed geodesic distribution	0.011	0.006	0.006	0.006
Proximity Based Mechanisms for Multilevel Networks (Collaboration and Journals of the Web of Science)				
Mixed Indegree distribution	0.714	0.648	0.672	0.649
Mixed outdegree distribution	0.359	0.337	0.362	0.347
Mixed triad census	0.487	0.381	0.405	0.384
Mixed geodesic distribution	0.399	0.193	0.198	0.221
Proximity Based Mechanisms for Multilevel Networks (Citations, Institutional Affiliation and Journals of the Web of Science)				
Mixed Indegree distribution	0.682	0.686	0.705	0.663
Mixed outdegree distribution	0.844	0.793	0.826	0.813
Mixed quadratic census	0.687	0.819	0.752	0.835
Mixed geodesic distribution	0.175	0.151	0.149	0.161
Proximity Based Mechanisms for Multilevel Networks (Collaboration, Institutional Affiliation and Journals of the Web of Science)				
Mixed Indegree distribution	0.642	0.635	0.65	0.625
Mixed outdegree distribution	0.293	0.242	0.243	0.235
Mixed quadratic census	0.946	0.946	0.953	0.95
Mixed geodesic distribution	0.075	0.057	0.061	0.075
Relational Based Mechanisms for Multiplex Networks (Citation and Collaboration)				
Mixed indegree distribution (without normalization)	0.951	0.935	0.935	0.931
Mixed outdegree distribution (without normalization)	0.171	0.205	0.201	0.214

Continuation

Proximity Based Mechanisms for Multilevel Networks (Citation and Institutional Affiliation)				
Mixed indegree distribution (without normalization)	0.447	0.420	0.448	0.472
Mixed outdegree distribution (without normalization)	0.712	0.754	0.743	0.724
Proximity Based Mechanisms for Multilevel Networks (Citation and Journals of the Web of Science				
Mixed indegree distribution (without normalization)	0.405	0.323	0.358	0.327
Mixed outdegree distribution (without normalization)	0.336	0.249	0.211	0.227
Proximity Based Mechanisms for Multilevel Networks (Citations, Institutional Affiliation and Journals of the Web of Science)				
Mixed indegree distribution (without normalization)	0.339	0.279	0.315	0.287
Relational Based Mechanisms for Multiplex Networks (Collaboration and Citation)				
Mixed indegree distribution (without normalization)	0.525	0.465	0.500	0.516
Mixed outdegree distribution (without normalization)	0.139	0.167	0.167	0.183
Proximity Based Mechanisms for Multilevel Networks (Collaboration and Institutional Affiliation)				
Mixed indegree distribution (without normalization)	0.503	0.504	0.518	0.499
Mixed outdegree distribution (without normalization)	0.644	0.667	0.666	0.659
Proximity Based Mechanisms for Multilevel Networks (Collaboration and Journals of the Web of Science)				
Mixed indegree distribution (without normalization)	0.215	0.178	0.188	0.199
Mixed outdegree distribution (without normalization)	0.157	0.136	0.139	0.134
Proximity Based Mechanisms for Multilevel Networks Collaboration, Institutional Affiliation and Journals of the Web of Science)				
Mixed outdegree distribution (without normalization)	0.859	0.793	0.799	0.764

0. **Descriptive Analysis**

Table 26 Descriptive of the Citation Network

ID	Organization	Name of Organization	Members	Total	Number of ties	Density	Average Degree
1	AIUC	Center for Astroengineering of the Catholic University	3	9	21	0.29	4.66
2	*CMM*	Center for Mathematical Modeling of the University of Chile	4	20	91	0.24	9.10
3	*CTIO*	The Cerro Tololo Inter-American Observatory	13	25	82	0.14	6.56
4	*ESO*	European Southern Observatory	90	132	206	0.01	3.12
5	*GEMINI*	Gemini Observatory	13	13	0	0.00	0.00
6	*HARLINGTEN*	Caisey Harlingten Observatory	2	2	0	0.00	0.00
7	*INEWTON*	The Isaac Newton Institute	8	14	18	0.10	2.57
8	*JAO*	Joint ALMA Observatory	24	31	13	0.01	0.84
9	*LCO*	Las Campanas Observatory	30	45	128	0.06	5.69
10	*MAD*	Millennium Nucleus Center of Protoplanetary Disks in ALMA Early Science	10	16	23	0.10	2.86
11	*MAS*	The Millennium Institute of Astrophysics	56	115	277	0.02	4.82
12	*NAOJ*	The National Astronomical Observatory of Japan	13	14	3	0.02	0.43
13	*NOAO*	NSF's National Optical Astronomy Observatory	10	10	0	0.00	0.00
14	*NRAO*	NSF's National Radio Astronomy Observatory	1	5	8	0.40	3.20
15	*PUC*	The Pontifical Catholic University of Chile	105	128	151	0.01	2.36

Continuation

16	*SOAR*	The Southern Astrophysical Research	2	2	0	0.00	0.00
17	*UA*	University of Antofagasta	6	6	0	0.00	0.00
18	*UAUTONOMA*	Autonomous University of Chile	2	4	4	0.33	2.00
19	*UCH*	The University of Chile	92	131	235	0.01	3.59
20	*UCN*	The Catholic University of the North	8	9	1	0.01	0.22
21	*UDA*	The University of Atacama	4	4	0	0.00	0.00
22	*UdeC*	The University of Concepcion	30	46	59	0.03	2.57
23	*UDP*	The University of Diego Portales	11	36	112	0.09	6.22
24	*ULS*	The University of La Serena	8	15	40	0.19	5.33
25	*UMI-FCA*	Unité Mixte Internationale Franco-Chilienne d'Astronomie	8	23	53	0.10	4.61
26	*UNAB*	The University of Andres Bello	9	25	56	0.09	4.48
27	*USACH*	The University of Santiago	2	3	1	0.17	0.67
28	*UTFSM*	The Federico Santa Maria Technical University	3	17	93	0.34	10.94
29	*UV*	The University of Valparaiso	38	47	39	0.02	1.66

Table 27 Descriptive of the Institutional Affiliation Network

ID	Organization	Name of Organization	Number of Organizations	Number of ties	Density	Average degree
1	AIUS	Center for Astroengineering of the Catholic University	3	21	0.41	4.00
2	*CMM*	Center for Mathematical Modeling of the University of Chile	2	91	0.53	11.00
3	*CTIO*	The Cerro Tololo Inter-American Observatory	6	82	0.13	3.17
4	*ESO*	European Southern Observatory	13	206	0.07	8.85
5	*GEMINI*	Gemini Observatory	2	0	0.08	1.00
6	*HARLINGTEN*	Caisey Harlingten Observatory	0	0	0.00	0.00
7	*INEWTON*	The Isaac Newton Institute	4	18	0.34	4.75
8	*JAO*	Joint ALMA Observatory	6	13	0.13	4.00
9	*LCO*	Las Campanas Observatory	7	128	0.11	5.00
10	*MAD*	Millennium Nucleus Center of Protoplanetary Disks in ALMA Early Science	6	23	0.21	3.33
11	*MAS*	The Millennium Institute of Astrophysics	11	277	0.11	13.27
12	*NAOJ*	The National Astronomical Observatory of Japan	2	3	0.11	1.50
13	*NOAO*	NSF's National Optical Astronomy Observatory	4	0	0.25	2.75
14	*NRAO*	NSF's National Radio Astronomy Observatory	2	8	0.50	3.00
15	*PUC*	The Pontifical Catholic University of Chile	18	151	0.05	6.44

Continuation

16	*SOAR*	The Southern Astrophysical Research	2	0	0.25	1.00
17	*UA*	University of Antofagasta	2	0	0.17	1.00
18	*UAUTONOMA*	Autonomous University of Chile	2	4	0.38	2.00
19	*UCHASTRO*	The University of Chile, Department Astronomy	12	235	0.07	8.58
20	*UCN*	The Catholic University of the North	6	1	0.11	1.00
21	*UDA*	The University of Atacama	1	0	0.00	0.00
22	*UdeC*	The University of Concepcion	9	59	0.09	4.33
23	*UDP*	The University of Diego Portales	6	112	0.18	6.67
24	*ULS*	The University of La Serena	2	40	0.23	3.5
25	*UMI-FCA*	Unité Mixte Internationale Franco-Chilienne d'Astronomie	3	53	0.29	7.00
26	*UNAB*	The University of Andres Bello	6	56	0.27	7.17
27	*USACH*	The University of Santiago	5	1	0.33	1.00
28	*UTFSM*	The Federico Santa Maria Technical University	2	93	0.29	5.5
29	*UV*	The University of Valparaiso	6	39	0.15	7.17

J. Goodness of Fit Stationary Stochastic Actor-oriented Models

Table 28 Goodness of Fit of Each Group

Group	Indegree	Outdegree	Geodesic	Triad	Clique	Bipartite Indegree	Bipartite Outdegree	Mixed Triad
AIUS	-	-	-	-	-	-	-	-
CMM	0.012	0.612	0.116	0.299	0.494	0.248	0.065	0.793
CTIO	0.178	0.966	0.900	0.995	0.495	0.106	0.523	0.961
ESO	0.124	0.107	0.512	0.516	0.106	0.533	0.002	0.033
GEMINI	-	-	-	-	-	-	-	-
HARLINGTEN	-	-	-	-	-	-	-	-
INEWTON	-	-	-	-	-	-	-	-
JAO	-	-	-	-	-	-	-	-
LCO	0.003	0.915	0.077	0.335	0.452	0.495	0.020	0.447
MAD	-	-	-	-	-	-	-	-
MAS	0.013	0.512	0.055	0.186	0.679	0.173	0.042	0.559
NAOJ	-	-	-	-	-	-	-	-
NOAO	-	-	-	-	-	-	-	-
NRAO	-	-	-	-	-	-	-	-
PUC	0.274	0.209	0.113	0.055	0.520	0.233	0.015	0.055
SOAR	-	-	-	-	-	-	-	-
UA	-	-	-	-	-	-	-	-
UAUTONOMA	-	-	-	-	-	-	-	-
UCH	0.002	0.596	0.049	0.379	0.669	0.324	0.016	0.223
UCN	-	-	-	-	-	-	-	-
UDA	-	-	-	-	-	-	-	-
UdeC	0.610	0.700	0.359	0.288	0.645	0.073	0.064	0.441
UDP	0.946	0.339	0.181	0.444	0.468	0.033	0.898	0.941
ULS	0.931	0.513	0.915	0.593	0.205	0.407	0.310	0.864
UMI-FCA	0.821	0.937	0.289	0.492	0.364	0.383	0.199	0.567
UNAB	-	-	-	-	-	-	-	-
USACH	-	-	-	-	-	-	-	-
UTFSM	0.008	0.698	0.638	0.631	0.598	0.044	0.961	0.149
UV	0.143	0.163	0.452	0.090	0.140	0.798	0.130	0.016

K. Sensitive Analysis

We present some sensitive analysis to identify potential specification and normalization of the *co-citation from weak ties* to compare no normalization, cosine similarity and weighted Jaccard similarity (table 4). From the analysis, the *co-citation from weak ties* without normalization have parameters that are more significantly different across the personal communities ($Q = 47.132$, $\mathrm{Qp} = 0.000$) and the *weighted Jaccard similarity* also present more significant variations ($Q = 43.614$, $\mathrm{Qp} = 0.000$). In the analysis we present the results using the *cosine similarity* ($Q = 20.986$, $\mathrm{Qp} = 0.021$) which has less significant variations.

Table 29 Meta-analysis Considering Co-citation from Weak Ties with No Normalization and Jaccard Weighted Similarity

	No normalization				Jaccard weighted similarity			
	Est	SE	Σ	Q	Est	SE	Σ	Q
Citation Network								
Outdegree (density)	- 4.137**	0.450	1.237	43.057**	-4.192**	0.460	1.232	44.712**
Reciprocity	2.331**	0.190	0.093	8.382	2.312**	0.161	0.000	0.917
Transitive triplets	0.175**	0.029	0.747	28.910*	0.184**	0.020	0.016	11.580
Transitive ties	3.225**	0.294	0.376	13.170	3.351**	0.323	0.826	31.626
Indegree popularity	-0.002	0.018	0.000	7.895	-0.010	0.017	0.018	10.857
√Outdegree popularity	- 0.669**	0.074	0.054	11.710	-0.669**	0.071	0.000	7.785
√Outdegree activity	0.237**	0.053	0.065	12.541	0.172**	0.046	0.000	8.853
Reciprocity degree activity	-0.129**	0.024	0.055	20.152	-0.105**	0.017	0.000	8.945
Closure by affiliation (ego)	-0.055	0.050	0.046	10.078	-0.036	0.048	0.040	7.450
Closure by affiliation	0.007	0.007	0.000	4.325	0.015*	0.007	0.000	7.936
Co-citation from weak ties	0.001**	0.000	0.000	0.000**	0.185*	0.083	0.203	43.614**
Accumulative citations (alter)	0.104**	0.024	0.055	20.152	0.110**	0.021	0.045	18.777
Accumulative citations (ego)	- 0.075**	0.017	0.000	7.131	-0.048**	0.015	0.000	6.170
Absolute difference of the accumulated number of citations	-0.117 **	0.018	0.024	13.577	-0.124**	0.014	0.001	14.454

Σ standard deviation, Q chi-squared test statistic. *Rates functions* fixed $\lambda = 100$ for the citation networks and $\lambda = 30$ for the institutional affiliation network.
$^{*}p < .05$; $^{**}p < .001$;

Continuation

	Est	SE	Σ	Q	Est	SE	Σ	Q
Institutional Affiliation								
Outdegree (density)	-1.335**	0.195	0.000	11.036	-1.116**	0.222	0.000	7.135
√Indegree popularity	0.272**	0.040	0.096	26.570*	0.222**	0.017	0.000	11.853
Outdegree activity	-0.294**	0.048	0.000	5.826	-0.354**	0.057	0.000	1.930
Observatory (ref. University)	-0.094*	0.042	0.000	3.919	-0.073	0.044	0.000	8.410
Research Centre (ref. University)	0.156*	0.045	0.000	4.149	0.147**	0.047	0.000	4.140
Size	0.141**	0.030	0.069	19.989*	0.135**	0.027	0.061	19.245
Closure by association	0.344**	0.052	0.000	2.672	0.312**	0.055	0.000	4.811

Σ standard deviation, Q chi-squared test statistic. *Rates functions* fixed $\lambda = 100$ for the citation networks and $\lambda = 30$ for the institutional affiliation network.
$^{*}p < .05$; $^{**}p < .001$;

Convergence of the *personal communities* in the no normalization specification: CMM, ESO, LCO, MAS, PUC, UCH, UdeC, UDP, ULS, UNAB, and UV. For the *personal communities* in the Jaccard weighted similarity specification: CMM, CTIO, ESO, LCO, MAS, PUC, UCH, UdeC, UDP, ULS, UNAB, and UV

We also perform some extra analysis to control for the retrospective citation of the actors from 2017 treated as an additional rate function or as a closure by affiliation considering the retrospective network (table 5). From the two variations the model with retrospective citations as a rate function seems promising and more stable than the model that controls for the closure by affiliation in the retrospective citations. Nonetheless, in general the direction, significance and size effects of the parameters behave similarly in the different specifications.

Table 30 Meta-Analysis Including Retrospective Citations

	Model with retrospective citations as a rate function				Model with closure by affiliation considering the retrospective citations			
	Est	SE	Σ	Q	Est	SE	Σ	Q
Citation Network								
Outdegree (density)	-5.093**	0.254	0.000	6.934	- 3.838**	0.518	1.297	37.118**
Reciprocity	2.225**	0.181	0.000	6.589	2.228**	0.167	0.000	10.217
Transitive triplets	0.177**	0.022	0.000	7.949	0.128**	0.036	0.063	15.111
Transitive ties	3.235**	0.251	0.439	14.029	3.230**	0.341	0.850	35.302**
Indegree popularity	-0.016	0.020	0.000	6.001	0.016	0.014	0.000	5.531
√Outdegree popularity	-0.640**	0.082	0.000	4.515	-0.751**	0.078	0.000	9.811
√Outdegree activity	0.201**	0.057	0.000	8.623	0.090	0.075	0.114	12.877
Reciprocity degree activity	-0.103**	0.020	0.000	5.339	-0.100**	0.018	0.022	9.259
Closure by affiliation (ego)	-0.003	0.051	0.000	10.222	-0.057	0.061	0.089	12.400
Closure by affiliation	0.016	0.009	0.000	4.029	0.012	0.007	0.000	3.434
Co-citation from weak ties	0.010**	0.004	0.005	23.408*	0.001**	0.000	0.001	54.356**
Accumulative citations (alter)	0.087**	0.015	0.013	13.243	0.104**	0.024	0.048	15.565
Accumulative citations (ego)	-0.043**	0.016	0.000	4.246	-0.076**	0.017	0.000	6.813
Absolute difference of the accumulated number of citations	-0.117**	0.016	0.002	14.035	-0.139**	0.020	0.031	14.532
Closure by affiliation in the retrospective citations					0.004**	0.000	0.000	5.134

Σ standard deviation, Q chi-squared test statistic. *Rates functions* fixed $\lambda = 100$ for the citation and the retrospective citation network, and $\lambda = 30$ for the institutional affiliation network.
$^*p < .05$; $^{**}p < .001$;

Continuation

	Est	SE	Σ	Q	Est	SE	Σ	Q
Institutional Affiliation								
Outdegree (density)	-0.975**	0.254	0.000	7.823	-1.171**	0.264	0.485	14.668
√Indegree popularity	0.269**	0.044	0.098	19.469*	0.229**	0.014	0.000	8.730
Outdegree activity	-0.372**	0.064	0.000	2.342	-0.289**	0.045	0.000	6.843
Observatory (ref. University)	-0.056	0.050	0.000	7.104	-0.097*	0.043	0.000	7.751
Research Centre (ref. University)	0.136**	0.052	0.000	3.665	0.167**	0.047	0.000	3.537
Size	0.115**	0.025	0.042	13.956	0.146**	0.030	0.067	18.896*
Closure by association	0.309**	0.060	0.011	6.091	0.310**	0.044	0.000	5.227

Σ standard deviation, Q chi-squared test statistic. *Rates functions* fixed $\lambda = 100$ for the citation and the retrospective citation network, and $\lambda = 30$ for the institutional affiliation network.
$^{*}p < .05$; $^{**}p < .001$;

Convergence of the *personal communities* adding the retrospective citation as a rate function: CMM, CTIO, LCO, MAS, PUC, UCH, UdeC, UDP, ULS, UTFSM, and UV. For the *personal communities* adding the retrospective citation as a dyadic covariable specification: CMM, CTIO, ESO, MAS, PUC, UCH, UdeC, UDP, ULS, UNAB, and UTFSM

In the following and for further consideration, we follow the operationalization of Snijders and Steglich (2015) an perform different models that have "short-term dynamic equilibrium". However, we will prefer for the interpretation of the main results a very high value to approximate a stationary distribution (Block et al., 2019). When the actors have different opportunities for changing its strong aggregated citations, we identify that most of the parameters behave similarly, nonetheless, there are some exceptions. In the three cases we can see differences in terms of the indegree tendency. For the shorter iteration process the effects seem more significant and positive ($\beta = 0.048,\ \mathrm{s.e.} = 0.017$), then the model with rate functions fixed to $\lambda = 50$ for the citation network and $\lambda = 20$ for the institutional affiliation the effects decrease its significance ($\beta = 0.023,\ \mathrm{s.e.} = 0.016$), and the longest model become negative and less significant as well ($\beta = -0.007,\ \mathrm{s.e.} = 0.063$). A similar issue appears in the *closure by affiliation* (ego), where the coefficient reverse direction and significance in the network with rate function fixed to $\lambda = 50$ for the citation network, and $\lambda = 20$ for the institutional affiliation ($\beta = 0.002,\ \mathrm{s.e.} = 0.049$), in comparison with the models that allowed shorter and large opportunities to changes ($\beta = -0.005,\ \mathrm{s.e.} = 0.054$ and $\beta = -0.009,\ \mathrm{s.e.} = 0.046$, respectively).

Table 31 Robustness of Different Specification in Fixing the Rate Functions with Cosine Similarity

	Rate function fixed to $\lambda = 30$ for the citation network and $\lambda = 10$ for the institutional affiliation				Rate function fixed to $\lambda = 50$ for the citation network and $\lambda = 20$ for the institutional affiliation			
	Est	SE	Σ	Q	Est	SE	Σ	Q
Citation Network								
Outdegree (density)	-5.430**	0.542	1.290	23.479*	-5.041**	0.492	1.114	20.102*
Reciprocity	2.208**	0.187	0.000	7.613	2.178**	0.181	0.000	6.363
Transitive triplets	0.163**	0.038	0.071	14.675	0.171**	0.021	0.000	12.522
Transitive ties	2.772**	0.220	0.376	13.170	2.917**	0.234	0.447	15.873
Indegree popularity	0.048*	0.017	0.000	4.806	0.023	0.016	0.000	4.549
√Outdegree popularity	-0.574**	0.078	0.001	10.987	-0.646**	0.077	0.000	8.305
√Outdegree activity	0.301**	0.071	0.094	14.114	0.218**	0.052	0.000	10.538
Reciprocity degree activity	-0.113**	0.019	0.000	5.386	-0.105**	0.018	0.000	5.593
Closure by affiliation (ego)	-0.005	0.054	0.000	10.480	0.002*	0.049	0.001	10.114
Closure by affiliation	0.028**	0.007	0.000	3.882	0.024*	0.007	0.000	4.723
Co-citation from weak ties	0.016**	0.005	0.007	16.786	0.015*	0.005	0.007	20.951
Accumulative citations (alter)	0.092**	0.016	0.001	12.781	0.091**	0.014	0.002	12.462
Accumulative citations (ego)	-0.041*	0.016	0.000	5.355	-0.042*	0.015	0.000	5.335
Absolute difference of the accumulated number of citations	-0.128**	0.018	0.007	14.595	-0.121**	0.017	0.011	13.648

Σ standard deviation, Q chi-squared test statistic.

$^{*}p < .05$; $^{**}p < .001$;

Continuation

	Est	SE	Σ	Q	Est	SE	Σ	Q
Institutional Affiliation Outdegree (density)	-1.339**	0.285	0.593	18.988	-1.306**	0.253	0.401	13.401
√Indegree popularity	0.308**	0.043	0.101	25.728	0.302**	0.044	0.106	31.074*
Outdegree activity	-0.249**	0.052	0.080	11.540	-0.276**	0.049	0.000	7.169
Observatory (ref. University)	-0.076	0.056	0.000	7.470	-0.072	0.048	0.000	8.936
Research Centre (ref. University)	0.146*	0.061	0.000	2.601	0.150*	0.050	0.000	3.607
Size	0.132**	0.027	0.037	12.033	0.130**	0.028	0.060	17.439
Closure by association	0.291**	0.041	0.000	7.447	0.279**	0.045	0.000	5.881

Σ standard deviation, Q chi-squared test statistic.

$^{*}p < .05$; $^{**}p < .001$;

Convergence of the *personal communities* fixing the rate function $\lambda = 30$ for the citation network and $\lambda = 10$ for the institutional affiliation: CMM, CTIO, LCO, MAS, PUC, UCH, UdeC, UDP, ULS, UNAB, UTFSM, and UV. For the *personal communities* fixing the rate function $\lambda = 50$ for the citation network and $\lambda = 20$ for the institutional affiliation: CMM, CTIO, LCO, MAS, PUC, UCH, UdeC, UDP, ULS, UNAB, UTFSM, and UV.

L. Separate Stochastic Actor-oriented Models

Table 32 Separate SAOM models

	CMM		UDP		UdeC		PUC	
Citation Network								
Outdegree (density)	-7.46*	2.50	-4.03**	1.20	-6.27**	1.18	-5.91**	0.77
Reciprocity	2.76*	0.96	2.25**	0.49	2.27*	0.89	3.41**	0.59
Transitive triplets	0.14	0.10	0.41**	0.10	0.21	0.11	-0.06	0.22
Transitive ties	3.49	1.99	2.19**	0.57	3.44**	0.69	2.81**	0.40
Indegree popularity	-0.07	0.10	-0.09	0.06	0.01	0.08	0.03	0.05
√Outdegree popularity	-0.58*	0.29	-0.99**	0.28	-0.84*	0.32	-0.18	0.21
√Outdegree activity	0.85*	0.29	0.04	0.14	0.08	0.29	0.46*	0.22
Reciprocity degree activity	-0.20*	0.09	-0.20*	0.07	-0.12	0.13	-0.20*	0.09
Closure by affiliation (ego)	0.87*	0.38	-0.21	0.21	-0.08	0.19	-0.00	0.14
Closure by affiliation	-0.01	0.03	0.05*	0.02	0.04	0.05	0.02	0.03
Co-citation from weak ties	0.27*	0.10	0.08*	0.04	0.09*	0.03	0.01	0.01
Accumulative citations (alter)	0.01	0.06	0.22**	0.07	0.09	0.08	0.16**	0.05
Accumulative citations (ego)	-0.01	0.05	-0.07	0.06	-0.13	0.09	-0.05	0.05
Absolute difference of the accumulated number of citations	-0.05	0.07	-0.16*	0.07	-0.19*	0.08	-0.13*	0.04
Institutional Affiliation								
Outdegree (density)	0.08	1.49	0.32	1.04	-1.09	0.78	-1.88*	0.92
√Indegree popularity	0.35	0.23	0.14	0.14	0.40**	0.10	0.22**	0.05
Outdegree activity	-0.45	0.31	-0.53*	0.25	-0.38*	0.19	-0.23	0.25
Observatory (ref. University)	-0.00	0.17	-0.13	0.16	-0.13	0.14	-0.23	0.15
Research Centre (ref. University)	0.23	0.20	0.03	0.18	0.09	0.17	0.12	0.18
Size	0.09	0.07	0.07	0.05	0.13	0.07	0.28**	0.08
Closure by association	0.17	0.15	0.56*	0.25	0.36	0.41	0.65	0.59

$^{*}p < .05$; $^{**}p < .001$;

Continuation	ESO		LCO		MAS		UCH	
Citation Network								
Outdegree (density)	-6.21**	0.66	-5.58**	0.76	-4.77**	0.56	-5.02**	0.45
Reciprocity	2.88**	0.50	1.82*	0.61	2.11**	0.29	2.29**	0.41
Transitive triplets	0.23**	0.04	0.13*	0.06	0.17**	0.04	0.17**	0.04
Transitive ties	4.65**	0.41	4.36**	0.66	2.99**	0.33	3.78**	0.39
Indegree popularity	-0.08	0.06	-0.02	0.04	0.02	0.02	-0.00	0.04
√Outdegree popularity	-0.64*	0.25	-0.43*	0.19	-0.69**	0.11	-0.63**	0.15
√Outdegree activity	0.21	0.13	0.28	0.18	0.13	0.08	0.22*	0.09
Reciprocity degree activity	-0.16	0.05	-0.08	0.05	-0.07*	0.03	-0.09*	0.04
Closure by affiliation (ego)	-0.07	0.12	0.04	0.16	-0.09	0.09	0.10	0.10
Closure by affiliation	0.03	0.02	0.00	0.02	0.02	0.01	0.01	0.02
Co-citation from weak ties	0.01	0.01	0.02	0.01	0.01*	0.01	0.00	0.00
Accumulative citations (alter)	0.23**	0.06	0.05	0.03	0.10**	0.02	0.10**	0.03
Accumulative citations (ego)	-0.11*	0.05	-0.02	0.03	-0.04	0.03	-0.04	0.04
Absolute difference of the accumulated number of citations	-0.14*	0.05	-0.06	0.04	-0.13**	0.03	-0.13**	0.03
Institutional Affiliation								
Outdegree (density)	-1.51**	0.44	-1.16	0.71	-0.60	0.77	-1.31*	0.46
√Indegree popularity	0.21**	0.03	0.38**	0.08	0.11	0.07	0.22**	0.03
Outdegree activity	-0.30*	0.11	-0.36	0.19	-0.43*	0.20	-0.37*	0.13
Observatory (ref. University)	-0.09	0.10	0.08	0.15	-0.21	0.15	-0.05	0.12
Research Centre (ref. University)	0.20	0.12	0.10	0.16	0.31*	0.14	0.19	0.11
Size	0.28**	0.08	0.08	0.05	0.30**	0.09	0.21**	0.05
Closure by association	0.35*	0.14	0.26	0.16	0.52	0.27	0.43*	0.18

$^*p < .05$; $^{**}p < .001$;

Continuation	ULS		CTIO		UV	
Citation Network						
Outdegree (density)	-6.05*	2.92	-6.12*	2.18	-4.18**	1.13
Reciprocity	0.32	1.79	2.95*	1.50	3.80	6.22
Transitive triplets	0.10	0.35	0.27*	0.10	0.41	2.07
Transitive ties	1.65	1.20	5.44*	1.97	3.51**	0.75
Indegree popularity	0.28	0.26	-0.13	0.09	0.07	0.34
√Outdegree popularity	-0.64	0.70	-0.67*	0.27	-1.81	4.51
√Outdegree activity	0.74	0.86	0.35	0.49	0.09	0.48
Reciprocity degree activity	-0.03	0.15	-0.25	0.17	-0.26	0.77
Closure by affiliation (ego)	0.97	0.83	-0.03	0.37	0.03	0.32
Closure by affiliation	0.03	0.10	0.05	0.07	0.04	0.06
Co-citation from weak ties	0.36	0.21	0.03	0.06	0.02	0.04
Accumulative citations (alter)	0.20	0.14	0.05	0.06	0.36	0.38
Accumulative citations (ego)	-0.04	0.18	-0.03	0.06	-0.24	0.25
Absolute difference of the accumulated number of citations	-0.51*	0.25	-0.05	0.06	-0.44	0.24
Institutional Affiliation						
Outdegree (density)	0.30	1.39	-0.09	1.03	-2.20*	0.73
√Indegree popularity	0.13	0.40	0.27	0.19	0.48**	0.08
Outdegree activity	-0.48	0.33	-0.38	0.21	-0.12	0.18
Observatory (ref. University)	-0.11	0.23	0.34	0.21	-0.03	0.15
Research Centre (ref. University)	-0.18	0.30	0.16	0.23	0.04	0.15
Size	0.02	0.09	0.07	0.06	0.03	0.07
Closure by association	0.42	0.22	0.18	0.12	0.74	0.94

$^{*}p < .05$; $^{**}p < .001$;

References

Ahlgren, P., & Colliander, C. (2009). Document–document similarity approaches and science mapping: Experimental comparison of five approaches. *Journal of Informetrics*, *3*(1), 49–63. https://doi.org/10.1016/j.joi.2008.11.003

Ahlgren, P., Jarneving, B., & Rousseau, R. (2003). Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology*, *54*(6), 550-560. https://doi.org/10.1002/asi.10242

Akbaritabar, A., Traag, V. A., Caimo, A., & Squazzoni, F. (2020). Italian sociologists: A community of disconnected groups. *Scientometrics*, *124*, 2361–2382. https://doi.org/10.1007/s11192-020-03555-w

Arancibia, D.; Bayo, A.; Cabrera-Vives, G.; Föster, F.; González, R.; Hamuy, M.; Maureira, J.C.; Quinn, P.; Rada, J.; Rodriguez, G.; Santander-Vela, J.; Tarenghi, M.; Ruiz, M. T.; San-Martin, M. & Williams, R. (2018). *Proceedings of the International Astronomical Union*, 14 (A30), 591 – 593. https://doi.org/10.1017/S174392131900557X

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Barandiaran, J. (2015). Reaching for the stars? Astronomy and growth in Chile. *Minerva*, *53*(2), 141-164.

Barnes, J. A. (1954). Class and committees in a Norwegian island parish. *Human Relations*, *7*(1), 39–58. https://doi.org/10.1177/001872675400700102

Barnes, J. A. (1969). Networks and political process. In J. C. Mitchell (Ed.), *Social networks in urban situations: Analyses of personal relationships in Central African towns* (pp. 51–76). Manchester University Press.

Basov, N., Breiger, R., & Hellsten, I. (2020). Socio-semantic and other dualities. *Poetics*, *78*, 1-12.

Batagelj, V. (2020). On Fractional Approach to analysis of linked networks. *Scientometrics*, *123*, 621–633. https://doi.org/10.1007/s11192-020-03383-y

Batagelj, V., & Cerinšek, M. (2013). On bibliographic networks. *Scientometrics*, *96*(3), 845–864.

Batagelj, V., Doreian, P., Ferligoj, A., & Kejzar, N. (2014). *Understanding large temporal networks and spatial networks: Exploration, pattern searching, visualization and network evolution* (Vol. 2). John Wiley & Sons.

Bear, J. B., & Woolley, A. W. (2011). The role of gender in team collaboration and performance. *Interdisciplinary Science Reviews*, *36*(2), 146–153. https://doi.org/10.1179/030801811X13013181961473

Bellotti, E. (2012). Getting funded. Multi-level network of physicists in Italy. *Social Networks*, *34*(2), 215–229. https://doi.org/10.1016/j.socnet.2011.12.002

Bellotti, E. (2015). *Qualitative networks: Mixed methods in sociological research*. Routledge.

Bellotti, E., Guadalupi, L., & Conaldi, G. (2016a). Comparing Fields of Sciences: Multilevel Networks of Research Collaborations in Italian Academia. In E. Lazega & T. A. B. Snijders (Eds.) *Multilevel Network Analysis for the Social Sciences* (pp. 213–244). Springer International Publishing. https://doi.org/10.1007/978-3-319-24520-1_9

Bellotti, E., Kronegger, L., & Guadalupi, L. (2016b). The evolution of research collaboration within and across disciplines in Italian Academia. *Scientometrics*, *109*(2), 783–811. https://doi.org/10.1007/s11192-016-2068-1

Bianconi, G., & Barabási, A. L. (2001). Bose-Einstein condensation in complex networks. *Physical review letters*, *86*(24), 5632. https://doi.org/10.1103/PhysRevLett.86.5632

Block, P. (2015). Reciprocity, transitivity, and the mysterious three-cycle. *Social Networks*, *40*, 163–173. https://doi.org/10.1016/j.socnet.2014.10.005

Block, P., Koskinen, J., Hollway, J., Steglich, C., & Stadtfeld, C. (2018). Change we can believe in: comparing longitudinal network models on consistency, interpretability and predictive power. *Social Networks*, *52*, 180–191. https://doi.org/10.1016/j.socnet.2017.08.001

Block, P., Stadtfeld, C., & Snijders, T. A. B. (2019). Forms of dependence: Comparing SAOMs and ERGMs from basic principles. *Sociological Methods & Research*, *48*(1), 202–239. https://doi.org/10.1177/0049124116672680

Boissevain, J. (1968). The place of non-groups in the social sciences. *Man*, *3*(4), 542-556. https://doi.org/10.2307/2798578

Boissevain, J. (1974). Friends of friends: networks, manipulators and coalitions. Oxford: Blackwell.

Bonacich, P. (2004). The invasion of the physicists. *Social Networks*, *26*(3), 285–288.

Borgatti, S. P. (1997). Structural holes: Unpacking Burt's redundancy measures. *Connections*, *20*(1), 35-38.

Borgatti, S. P. (2005). Centrality and network flow. *Social networks*, *27*(1), 55-71. https://doi.org/10.1016/j.socnet.2004.11.008

Borgatti, S. P., & Everett, M. G. (1997). Network analysis of 2-mode data. *Social Networks*, *19*(3), 243–269. https://doi.org/10.1016/S0378-8733(96)00301-2

Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social Networks*, *21*(4), 375–395. https://doi.org/10.1016/S0378-8733(99)00019-2

Borgatti, S. P., & Halgin, D. S. (2011a). Analyzing affiliation networks. In: J. Scott and P. J. Carrington (Eds.) *The Sage handbook of social network analysis* (pp. 417-433), Sage.

Borgatti, S. P., & Halgin, D. S. (2011b). On network theory. *Organization Science*, *22*(5), 1168–1181. https://doi.org/10.1287/orsc.1100.0641

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). Ucinet for Windows: Software for social network analysis. *Harvard, MA: Analytic Technologies*, *2006*.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). *Analyzing social networks*. Sage.

Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional Studies*, *39*(1), 61–74. https://doi.org/10.1080/0034340052000320887

Bott, E. (1968). *Family and Social Network: Roles, Norms, and External Relationships in Ordinary Urban Families*. Routledge.

Bottero, W., & Crossley, N. (2011). Worlds, fields and networks: Becker, Bourdieu and the structures of social relations. *Cultural Sociology*, *5*(1), 99–119. https://doi.org/10.1177/1749975510389726

Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information*, *14*(6), 19–47. https://doi.org/10.1177/053901847501400602

Bourdieu, P. (1988). *Homo academicus*. Stanford University Press.

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, *61*(12), 2389–2404.

Brailly, J., Favre, G., Chatellet, J., & Lazega, E. (2016). Embeddedness as a multilevel problem: A case study in economic sociology. Social Networks, 44, 319-333. https://doi.org/10.1016/j.socnet.2015.03.005

Breiger, R. L. (1974). The duality of persons and groups. *Social Forces*, *53*(2), 181-190. https://doi.org/10.1093/sf/53.2.181

Breiger, R. L. (1976). Career attributes and network structure: A blockmodel study of a biomedical research specialty. *American Sociological Review*, 117-135. https://doi.org/10.2307/2094376

Broccatelli, C., Everett, M., & Koskinen, J. (2016). Temporal dynamics in covert networks. *Methodological Innovations*, *9*, 1–14. https://doi.org/10.1177/2059799115622766

Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications*, *10*(1), 1-10.

Bronfman, L. (2002). A panorama of Chilean astronomy. *The Messenger*, *107*, 14-18.

Bu, Y., Liu, T., & Huang, W. (2016). MACA: A modified author co-citation analysis method combined with general descriptive metadata of citations. *Scientometrics*, *108*(1), 143–166. https://doi.org/10.1007/s11192-016-1959-5

Bu, Y., Wang, B., Huang, W. B., Che, S., & Huang, Y. (2018). Using the appearance of citations in full text on author co-citation analysis. *Scientometrics*, *116*(1), 275-289. https://doi.org/10.1007/s11192-018-2757-z

Burt, R. (1992). *Structural holes: The social structure of competition.* Harvard University Press.

Burt, R. S., & Doreian, P. (1982). Testing a structural model of perception: conformity and deviance with respect to journal norms in elite sociological methodology. *Quality and Quantity*, *16*(2), 109-150. https://doi.org/10.1007/BF00166880

Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology*, *38*(1), 155–200.

Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, *22*(2), 191-235.

Camacho-Miñano, M. D. M., & Núñez-Nickel, M. (2009). The 221o-author221red nature of reference selection. *Journal of the American Society for Information Science and Technology*, *60*(4), 754-777. https://doi.org/10.1002/asi.21018

Cárdenas, J. P., Cabrera, F., Moguillansky, G., & Olivares, G. (2015). *Cartografía del conocimiento en Chile* (pp. 1–81). CNID.

Carley, K. M. (2002). Smart agents and organizations of the future. In: Leah Lievrouw & Sonia Livingstone (Eds.), *The Handbook of New Media* (pp. 206-220). Thousand Oaks, CA, Sage.

Chang, Y.-T., Pantazis, D., & Leahy, R. M. (2010). Statistically optimal modular partitioning of directed graphs. *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, 1075–1079.

Chang, Y.-T., Pantazis, D., & Leahy, R. M. (2011). Partitioning directed graphs based on modularity and information flow. *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 1105–1108.

Chubin, D. E. (1976). State of the field the conceptualization of scientific specialties. *The Sociological Quarterly*, *17*(4), 448-476. https://doi.org/10.1111/j.1533-8525.1976.tb01715.x

Chubin, D. E. (1985). Beyond invisible colleges: Inspirations and aspirations of post-1972 social studies of science. *Scientometrics*, *7*(3-6), 221-254.

Chubin, D. E., & Moitra, S. D. (1975). Content analysis of references: Adjunct or alternative to citation counting? *Social studies of science*, *5*(4), 423-441.

Chubin, D. E., & Studer, K. (1979). Knowledge and structures of scientific growth measurement of a cancer problem domain. *Scientometrics*, *1*(2), 171-193. https://doi.org/10.1007/bf02016969

Cimenler, O., Reeves, K. A., & Skvoretz, J. (2015). An evaluation of collaborative research in a college of engineering. *Journal of Informetrics*, *9*(3), 577–590. https://doi.org/10.1016/j.joi.2015.05.003

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661-703.

Cole, J. R., & Cole, S. (1973). *Social stratification in science*. University of Chicago Press.

Collins, H. M. (1974). The TEA set: Tacit knowledge and scientific networks. *Science Studies*, 4(2), 165-185. https://www.jstor.org/stable/284473

Collins, H. M. (1981). The place of the 'core-set' in modern science: social contingency with methodological propriety in science. *History of Science*, *19*(1), 6-19. https://doi.org/10.1177/007327538101900102

Collins, H. M. (1988). Public experiments and displays of virtuosity: The core-set revisited. *Social Studies of Science*, *18*(4), 725-748. https://www.jstor.org/stable/284968

Collins, H. M. (1998). The Meaning of Data: Open and Closed Evidential Cultures in the Search for Gravitational Waves. *American Journal of Sociology*, 104(2), 293–337.

CONICYT. (2012). *Astronomy, technology, industry: Roadmap for the fostering of technology development and innovation in the field of astronomy in chile*. CONICYT.

Crane, D. (1969). Social Structure in a Group of Scientists: A Test of the 'Invisible College' Hypothesis. *American Sociological Review*, 34(3), 335-352. https://doi.org/10.2307/2092499

Crane, D. (1972). *Invisible colleges: Invisible colleges; diffusion of knowledge in scientific communities*. University of Chicago Press.

Crawford, S. (1971). Informal communication among scientists in sleep research. *Journal of the American Society for Information Science*, *22*(5), 301-310.

Crossley, N. (2011). *Towards relational sociology*. Routledge.

Crossley, N., Bellotti, E., Edwards, G., Everett, M. G., Koskinen, J., & Tranmer, M. (2015). *Social network analysis for ego-nets: Social network analysis for actor-centred networks*. Sage.

Davis, J. A., & Leinhardt, S. (1967). The Structure of Positive Interpersonal Relations in Small Groups. In J. Berger (Ed.), *Sociological Theories in Progress* (Vol. 2, pp. 218–251). Houghton Mifflin.

Davis, J.A. & Leinhardt, S. (1972). The structure of positive relations in small groups. In: J. Berger, M. Zelditch & B. Anderson (Eds.), *Sociological Theories in Progress* (pp. 218–51). Houghton Mifflin.

De Nooy, W. (2003). Fields and Networks: Correspondence Analysis and Social Network Analysis in the Framework of Field Theory. *Poetics*, *31*(5), 305–327. https://doi.org/10.1016/S0304-422X(03)00035-4

Dekker, D., Krackhardt, D., & Snijders, T. A. (2007). Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika*, *72*(4), 563-581. https://doi.org/10.1007/S11336-007-9016-1

Desmarais, B. A., & Cranmer, S. J. (2012). Micro-level interpretation of exponential random graph models with application to estuary networks. *Policy Studies Journal*, *40*(3), 402–434. https://doi.org/10.1111/j.1541-0072.2012.00459.x

Dhand, A., Luke, D. A., Carothers, B. J., & Evanoff, B. A. (2016). Academic cross-pollination: The role of disciplinary affiliation in research collaboration. *PLOS ONE*, *11*(1), e0145916. https://doi.org/10.1371/journal.pone.0145916

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160. https://doi.org/10.2307/2095101

Egghe, L., & Leydesdorff, L. (2009). The relation between Pearson's correlation coefficient r and Salton's cosine measure. *Journal of the American Society for Information Science and Technology*, *60*(5), 1027–1036. https://doi.org/10.1002/asi.21009

Elmer, T., & Stadtfeld, C. (2020). Depressive symptoms are associated with social isolation in face-to-face interaction networks. *Scientific reportsReports*, *10*(1), 1-12. https://doi.org/10.1038/s41598-020-58297-9

Elmer, T., Boda, Z., & Stadtfeld, C. (2017). The co-evolution of emotional well-being with weak and strong friendship ties. *Network Science*, *5*(3), 278-307. https://doi.org/10.1017/nws.2017.20

Espinosa-Rada, A. (2015). *La Productividad y la colaboración en el reconocimiento científico de los astrónomos de Chile* (MA). Pontificia Universidad Católica de Chile.

Espinosa-Rada, A., Ortiz-Ruiz, F., & Cereceda-Lorca, T. (2019). Astroinformática y prospección de la astronomía chilena: Sub-disciplina en el escenario global y desarrollo científico local. *Cultura-hombre-sociedad*, *29*(1), 276-305. http://doi.org/10.7770/0719-2789.2019.cuhso.01.a02

Espinosa, A. (2015). Asignación de científicos en el observatorio astronómico Atacama Large Millimeter/Submillimeter Array (ALMA): lógica virtual, sociomaterialidad y régimen de control en la ciencia. *Persona y Sociedad*, *29*(2), 45,66. https://doi.org/10.11565/pys.v29i2.88

Etzkowitz, H., Kemelgor, C., & Uzzi, B. (2000). *Athena unbound: The advancement of women in science and technology*. Cambridge University Press.

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. *Research policy*, *29*(2), 109-123. https://doi.org/10.1016/S0048-7333(99)00055-4

Everett, M. G., & Borgatti, S. P. (2012). Categorical attribute based centrality: E–I and G–F centrality. *Social Networks*, *34*(4), 562–569. https://doi.org/10.1016/j.socnet.2012.06.002

Everett, M. G., & Borgatti, S. P. (2013). The dual-projection approach for two-mode networks. *Social Networks*, *35*(2), 204–210.

Everett, M. G., & Borgatti, S. P. (2019). Partitioning multimode networks. In *Advances in Network Clustering and Blockmodeling* (pp. 251–265). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119483298.ch9

Everett, M. G., & Borgatti, S. P. (2020). Unpacking Burt's constraint measure. *Social Networks*, 62, 50-57. https://doi.org/10.1016/j.socnet.2020.02.001

Everett, M. G., Broccatelli, C., Borgatti, S. P., & Koskinen, J. (2018). Measuring knowledge and experience in two mode temporal networks. *Social Networks*, *55*, 63–73. https://doi.org/10.1016/j.socnet.2018.05.003

Fagan, J., Eddens, K. S., Dolly, J., Vanderford, N. L., Weiss, H., & Levens, J. S. (2018). Assessing research collaboration through co-authorship network analysis. *The Journal of Research Administration*, *49*(1), 76–99.

Fararo, T. J., & Doreian, P. (1984). Tripartite structural analysis: Generalizing the Breiger-Wilson formalism. *Social Networks*, *6*(2), 141–175. https://doi.org/10.1016/0378-8733(84)90015-7

Feld, S. L. (1981). The focused organization of social ties. *American Journal of Sociology*, *86*(5), 1015–1035. https://doi.org/10.1086/227352

Feld, S. L. (1991). Why your friends have more friends than you do. *American Journal of Sociology*, *96*(6), 1464–1477. https://doi.org/10.1086/229693

Felt, U., Fouché, R., Miller, C. A., & Smith-Doerr, L. (2017). *The handbook of science and technology studies*. MIT Press.

Ferligoj, A., Kronegger, L., Mali, F., Snijders, T. A. B., & Doreian, P. (2015). Scientific collaboration dynamics in a national scientific system. *Scientometrics*, *104*(3), 985–1012. https://doi.org/10.1007/s11192-015-1585-7

Fligstein, N., & McAdam, D. (2012). *A theory of fields*. Oxford University Press.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3–5), 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, *659*, 1–44.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185. https://doi.org/10.1126/science.aao0185

Frank, O., & Snijders, T. (1994). Estimating the size of hidden populations using snowball sampling. *Journal of Official Statistics*, *10*(1), 53-67.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

Friedkin, N. E. (1978). University social structure and social networks among scientists. *American Journal of Sociology*, *83*(6), 1444-1465.

Fujimoto, K., Snijders, T. A. B., & Valente, T. W. (2017). Popularity breeds contempt: The evolution of reputational dislike relations and friendships in high school. *Social Networks*, *48*, 100–109. https://doi.org/10.1016/j.socnet.2016.07.006

Galaskiewicz, J., & Wasserman, S. (1989). Mimetic processes within an interorganizational field: An empirical test. *Administrative science quarterly*, 34(3), 454-479. https://doi.org/10.2307/2393153

Gansner, E. R., Koren, Y., & North, S. (2004). Graph drawing by stress majorization. *In International Symposium on Graph Drawing* (pp. 239-250). Springer. https://doi.org/10.1007/978-3-540-31843-9_25

Garfield, E. (1955). Citation indexes for science. *Science*, *122*(3159), 108-111.

Garfield, E., Sher, I. H., & Torpie, R. J. (1964). *The use of citation data in writing the history of science*. Institute for Scientific Information, INC Philadelphia PA.

Gibert, J. (2011). La construcción social del científico. Notas sobre identidad intelectual y social de comunidades científicas en Chile. *Estudios Sociales*, *119*, 169–206.

Gilbert, G. N. (1977). Referencing as persuasion. *Social studies of science*, *7*(1), 113-122. https://www.jstor.org/stable/284636

Gile, K. J., & Handcock, M. S. (2010). Respondent-driven sampling: An assessment of current methodology. *Sociological Methodology*, *40*(1), 285–327. https://doi.org/10.1111/j.1467-9531.2010.01223.x

Glänzel, W., & Czerwon, H. J. (1996). A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level. *Scientometrics*, *37*(2), 195–221. https://doi.org/10.1007/BF02093621

Glänzel, W., & Thijs, B. (2017). Using hybrid methods and 'core documents' for the representation of clusters and topics: the astronomy dataset. *Scientometrics*, *111*(2), https://doi.org/10.1007/s11192-017-2301-6

Gläser, J., (2001). Scientific Specialties as the (Currently Missing) link between scientometrics and the sociology of science. In M. Davis & C. S. Wilson (Eds.), *Conference on Scientometrics and Infometrics Proceedings* (pp. 191-210). Bibliometric & Informetric Research Group: Sydney.

Gómez, M., & Mérida, F. (2007). Evaluating ADS, ISI Web of Knowledge and Scopus in the context of two astronomy libraries in Spain. In S. Ricketts, C. Birdie, & E. Isaksson (Eds.), *Library and Information Services in Astronomy V ASP Conference Serie* (Vol. 377, pp. 175–183).

Gondal, N. (2011). The local and global structure of knowledge production in an emergent research field: An exponential random graph analysis. *Social Networks*, *33*(1), 20–30. https://doi.org/10.1016/j.socnet.2010.09.001

Gondal, N. (2018). Duality of departmental specializations and PhD exchange: A Weberian analysis of status in interaction using multilevel exponential random graph models (mERGM). *Social Networks*, *55*, 202–212. https://doi.org/10.1016/j.socnet.2018.07.005

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360-1380.

Guridi, J. A., Pertuze, J. A., & Pfotenhauer, S. M. (2020). Natural laboratories as policy instruments for technological learning and institutional capacity building: The case of Chile's astronomy cluster. *Research Policy*, *49*(2), 103899. https://doi.org/10.1016/j.respol.2019.103899

Hagen, N. T. (2013). Harmonic 227o-author credit: A parsimonious quantification of the byline hierarchy. *Journal of Informetrics*, *7*(4), 784–791. https://doi.org/10.1016/j.joi.2013.06.005

Hagstrom, W. O. (1965). *The scientific community*. Basic Books.

Hagstrom, W.O. (1976). The production of culture in science. *American Behavioral Scientist*, *19*(6), 753-768. https://doi.org/10.1177/000276427601900606

Hanneke, S., Fu, W., & Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, *4*, 585–605. https://doi.org/10.1214/09-EJS548

Harris, J. K., Wong, R., Thompson, K., Haire-Joshu, D., & Hipp, J. A. (2015). Networks of collaboration among scientists in a center for diabetes translation research. *PLOS ONE*, *10*(8), e0136457. https://doi.org/10.1371/journal.pone.0136457

Hedström, P., & Bearman, P. S. (2009). What is analytical sociology all about? An introductory essay. In P. Hedström & P. S. Bearman (Eds.), *The Oxford handbook of analytical sociology* (pp. 3–24). Oxford University Press.

Heidler, R. (2011). Cognitive and social structure of the elite collaboration network of astrophysics: a case study on shifting network structures. *Minerva*, *49*(4), 461-488. https://doi.org/10.1007/s11024-011-9184-0

Heidler, R. (2017). Epistemic cultures in conflict: The case of astronomy and high energy physics. *Minerva*, *55*(3), 249-277. https://doi.org/10.1007/s11024-017-9315-3

Hodge, S. E., & Greenberg, D. A. (1981). Publication credit. *Science*, *213*(4511), 950–952.

Hoeppe, G. (2014). Working data together: The accountability and reflexivity of digital astronomical practice. *Social Studies of Science*, *44*(2), 243-270.

Holland, P. W., & Leinhardt, S. (1976). Local structure in social networks. *Sociological methodology*, *7*, 1-45. https://doi.org/10.2307/270703

Holland, P.W. & Leinhardt, S. (1970) A method for detecting structure in sociometric data. American Journal of Sociology, 70, 492-513

Holland, P.W. & Leinhardt, S. (1974). Measurement error in sociometry. Blalock, H. M. (Eds.), *Measurement in the Social Science: Theories and Strategy* (pp. 187-211). University of Washington.

Hollway, J., Lomi, A., Pallotti, F., & Stadtfeld, C. (2017). Multilevel social spaces: The network dynamics of organizational fields. *Network Science*, *5*(2), 187–212. https://doi.org/10.1017/nws.2017.8

Homans, G. C. (1950). *The Human Group*. Harcourt, Brace & World, Inc.

Huang, M.-H., & Chang, C.-P. (2016). A comparative study on three citation windows for detecting research fronts. *Scientometrics*, *109*(3), 1835–1853. https://doi.org/10.1007/s11192-016-2133-9

Hummon, N. P., & Dereian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, *11*(1), 39-63. https://doi.org/10.1016/0378-8733(89)90017-8

Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, *103*(481), 248–258. https://doi.org/10.1198/016214507000000446

Indlekofer, N., & Brandes, U. (2013). Relative importance of effects in stochastic actor-oriented models. *Network Science*, *1*(3), 278–304.

Jacobs, J. A., & Frickel, S. (2009). Interdisciplinarity: A critical assessment. *Annual Review of Sociology*, *35*, 43-65

Jacomy, M. (2020). Epistemic clashes in network science: Mapping the tensions between idiographic and nomothetic subcultures. *Big Data & Society*, *7*(2), 1-21.

Jansen, D., von Goertz, R., & Heidler, R. (2010). Knowledge production and the structure of collaboration networks in two scientific fields. *Scientometrics*, *83*(1), 219–241. https://doi.org/10.1007/s11192-009-0022-1

Jha, Y., & Welch, E. W. (2010). Relational mechanisms governing multifaceted collaborative behavior of academic scientists in six fields of science and engineering. *Research Policy*, *39*(9), 1174–1184. https://doi.org/10.1016/j.respol.2010.06.003

Kadushin, C. (1966). The Friends and Supporters of Psychotherapy: On Social Circles in Urban Life. *American Sociological Review*, 31(6), 786–802. https://doi.org/10.2307/2091658

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18. https://doi.org/10.1016/S0048-7333(96)00917-1

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, *14*(1), 10–25.

Kim, Y., Son, S.-W., & Jeong, H. (2010). Finding communities in directed networks. *Physical Review E*, *81*(1), 016103. https://doi.org/10.1103/PhysRevE.81.016103

Klavans, R., & Boyack, K. W. (2017). Which Type of Citation Analysis Generates the Most Accurate Taxonomy of Scientific and Technical Knowledge? *Journal of the Association for Information Science and Technology*, *68*(4), 984–998. https://doi.org/10.1002/asi.23734

Knorr-Cetina, K. D. (1982). Scientific communities or transepistemic arenas of research? A critique of quasi-economic models of science. *Social Studies of Science*, *12*(1), 101-130.

Koskinen, J., & Edling, C. (2012). Modelling the evolution of a bipartite network—Peer referral in interlocking directorates. *Social Networks*, *34*(3), 309–322. https://doi.org/10.1016/j.socnet.2010.03.001

Koskinen, J., & Lomi, A. (2013). The local structure of globalization. *Journal of Statistical Physics*, *151*(3), 523–548. https://doi.org/10.1007/s10955-013-0732-x

Koskinen, J., Caimo, A., & Lomi, A. (2015). Simultaneous modeling of initial conditions and time heterogeneity in dynamic networks: An application to foreign direct investments. *Network Science*, *3*(1), 58–77. https://doi.org/10.1017/nws.2015.3

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, *10*(4), 359-381. https://doi.org/10.1016/0378-8733(88)90004-4

Krackhardt, D. (1994). Graph theoretical dimensions of informal organizations. In K. M. Carley & M. J. Prietula (Eds.), *Computational organization theory* (pp. 107–130). Psychology Press.

Krackhardt, D., & Carley, K. M. (1998). *PCANS model of structure in organizations* (pp. 113-119). Pittsburgh, Pa, USA: Carnegie Mellon University, Institute for Complex Engineered Systems.

Krackhardt, D., & Stern, R. N. (1988). Informal networks and organizational crises: An experimental simulation. *Social Psychology Quarterly*, *51*(2), 123–140. https://doi.org/10.2307/2786835

Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks. *Electronic journal of statistics*, *6*, 1100-1128. https://doi.org/10.1214/12-EJS696

Krivitsky, P. N., & Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 29–46. https://doi.org/10.1111/rssb.12014

Krivitsky, P. N., Koehly, L. M., & Marcum, C. S. (2020). Exponential-family random graph models for multi-layer networks. *Psychometrika*, *85*(3), 630-659. https://doi.org/10.1007/s11336-020-09720-7

Kronegger, L., Mali, F., Ferligoj, A., & Doreian, P. (2012). Collaboration structures in Slovenian scientific communities. *Scientometrics*, *90*(2), 631–647. https://doi.org/10.1007/s11192-011-0493-8

Kuhn, T. S. (2012). The structure of scientific revolutions. University of Chicago Press.

Latapy, M., Magnien, C., & Del Vecchio, N. (2008). Basic notions for the analysis of large two-mode networks. *Social Networks*, *30*(1), 31–48. https://doi.org/10.1016/j.socnet.2007.04.006

Latour, B. & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton University Press.

Latour, B., (1987). *Science in action: How to follow scientists and engineers through society.* Harvard University Press.

Latour, B., Jensen, P., Venturini, T., Grauwin, S., & Boullier, D. (2012). 'The whole is always smaller than its parts'–a digital test of Gabriel Tardes' monads. *The British Journal of Sociology*, *63*(4), 590-615.

Laumann, E. O., Galaskiewicz, J., & Marsden, P. V. (1978). Community structure as interorganizational linkages. *Annual Review of Sociology*, *4*, 455–484.

Laumann, E. O., Marsden, P. V., & Prensky, D. (1983). The boundary specification problem in network analysis. In R. S. Burt & M. J. Minor (Eds.), *Applied Network Analysis: A Methodological Introduction* (pp. 18–34). Sage.

Lazarsfeld, P. F., & Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. In: M. Berger, T. Abel & C.H. Page (Eds.) *Freedom and Control in Modern Society* (pp. 18–66). D. Van Nostrand Company.

Lazega, E. (2001). *The collegial phenomenon: The social mechanisms of cooperation among peers in a corporate law partnership*. Oxford University Press.

Lazega, E. (2018). Networks and institutionalization: A neo-structural approach. *Connections*, *37*, 7–22. https://doi.org/10.21307/connections-2017-001

Lazega, E., & Jourdá, M. T. (2016). The structural wings of Matthew effects: The contribution of three-level network data to the analysis of cumulative advantage. *Methodological Innovation*, *9*, 1–13. https://doi.org/10.1177/2059799115622764

Lazega, E., & Snijders, T. A. B. (Eds.). (2016). *Multilevel network analysis for the social sciences: Theory, methods and applications*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24520-1

Lazega, E., Bar-Hen, A., Barbillon, P., & Donnet, S. (2016). Effects of competition on collective learning in advice networks. *Social Networks*, *47*, 1-14. https://doi.org/10.1016/j.socnet.2016.04.001

Lazega, E., Jourda, M., & Mounier, L. (2013). Network lift from dual alters: Extended opportunity structures from a multilevel and structural perspective. *European Sociological Review*, *29*(6), 1226–1238. https://doi.org/10.1093/esr/jct002

Lazega, E., Jourda, M., Mounier, L., & Stofer, R. (2008). Catching up with big fish in the big pond? Multi-level network analysis through linked design. *Social Networks*, *30*(2), 159–176. https://doi.org/10.1016/j.socnet.2008.02.001

Lazega, E., Mounier, L., Jourda, M. T. S., & Stofer, R. (2006). Organizational vs. personal social capital in scientists' performance: A multi-level network study of elite French cancer researchers (1996-1998). *Scientometrics*, *67*(1), 27-44. https://doi.org/10.1007/s11192-006-0049-5

Leicht, E. A., & Newman, M. E. J. (2008). Community structure in directed networks. *Physical Review Letters*, *100*(11), 118703. https://doi.org/10.1103/PhysRevLett.100.118703

Leifeld, P., & Cranmer, S. J. (2019). A theoretical and empirical comparison of the temporal exponential random graph model and the stochastic actor-oriented model. *Network Science*, *7*(1), 20–51. https://doi.org/10.1017/nws.2018.26

Leydesdorff, L. (1997). Why words and co-words cannot map the development of the sciences. *Journal of the American Society for Information Science*, *48*(5), 418–427. https://doi.org/10.1002/(SICI)1097-4571(199705)48:5<418::AID-ASI4>3.0.CO;2-Y

Leydesdorff, L. (2008). On the normalization and visualization of author co-citation data: Salton's cosine versus the Jaccard index. *Journal of the American Society for Information Science and Technology*, *59*(1), 77-85. https://doi.org/10.1002/asi.20732

Leydesdorff, L., & Park, H. W. (2016). Full and fractional counting in bibliometric networks. *ArXiv Preprint ArXiv:1611.06943*.

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the Association for Information Science and Technology*, *60*(2), 348–362. https://doi.org/10.1002/asi.20967

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, *60*(2), 348-362. https://doi.org/10.1002/asi.20967

Leydesdorff, L., & Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment. *Journal of the American Society for Information Science and Technology*, *57*(12), 1616-1628. https://doi.org/10.1002/asi.20335

Lievrouw, L. A. (1989). The invisible college reconsidered: Bibliometrics and the development of scientific communication theory. *Communication Research*, *16*(5), 615-628. https://doi.org/10.1177/009365089016005004

Lievrouw, L. A. (1990). Reconciling structure and process in the study of scholarly communication. In C.L. Borgman (Ed.), *Scholarly communication and bibliometrics* (pp. 59–69). Sage.

Lievrouw, L. A., Rogers, E. M., Lowe, C. U., & Nadel, E. (1987). Triangulation as a research strategy for identifying invisible colleges among biomedical scientists. *Social Networks*, 9(3), 217-248. https://doi.org/10.1016/0378-8733(87)90021-9

Liu, R.-L. (2017). A new bibliographic coupling measure with descriptive capability. *Scientometrics*, *110*(2), 915–935. https://doi.org/10.1007/s11192-016-2196-7

Lomi, A., & Stadtfeld, C. (2014). Social networks and social settings: Developing a coevolutionary view. *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, *66*(1), 395–415. https://doi.org/10.1007/s11577-014-0271-8

López, S., Hamuy, M., Barrientos, L. F., Maza, J., & Ruiz, M. T. (2005). Astronomía. In J. E. Allende, J. Babul, S. Martinez, & T. Ureta (Eds.), *Análisis y proyecciones de la ciencia chilena 2005* (pp. 239–260). Academia Chilena de Ciencias, Consejo de Sociedades Científicas de Chile, Programa Bicentenario de Ciencia y Tecnología (Conicyt), Santiago.

Lospinoso, J. A. (2012). *Statistical methods for social network dynamics* (PhD Thesis). Dphil Thesis, University of Oxford.

Lospinoso, J., & Snijders, T. A. (2019). Goodness of fit for stochastic actor-oriented models. *Methodological Innovations*, *12*(3), 1–18. https://doi.org/10.1177/2059799119884282

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, *16*(12), 317-323. https://www.jstor.org/stable/24529203

Luke, D. A., Baumann, A. A., Carothers, B. J., Landsverk, J., & Proctor, E. K. (2016). Forging a link between mentoring and collaboration: A new training model for implementation science. *Implementation Science*, *11*(1), 137. https://doi.org/10.1186/s13012-016-0499-y

Lusher, D., Koskinen, J., & Robins, G. (2013). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge University Press.

Malliaros, F. D., & Vazirgiannis, M. (2013). Clustering and community detection in directed networks: A survey. *Physics Reports*, *533*(4), 95–142.

Marra, M. (2014). *Astrophysics as an excellent object for bibliometric analysis, myth or reality? A sample study*. Presented at the 5th Conference on Library and Information Services in Astronomy: Common Challenges, Uncommon Solutions, San Francisco.

Marshakova, I. V. (1973). Co-Citation in Scientific Literature: A New Measure of the Relationship Between Publications. *Scientific and Technical Information Serial of VINITI*, *6*, 3–8.

Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & López-Cózar, E. D. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics*, 126, 871-906. https://doi.org/10.1007/s11192-020-03690-4

Mayer, A. C. (1966). The significance of quasi-groups in the study of complex societies. In Michael Banton (Ed.) *The Social Anthropology of Complex Societies* (pp. 97-122). London: Tavistock.

McCain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science*, *41*(6), 433–443. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<433::AID-ASI11>3.0.CO;2-Q

McCray, W. P. (2000). Large telescopes and the moral economy of recent astronomy. *Social Studies of Science*, *30*(5), 685–711. https://doi.org/10.1177/030631200030005002

McCray, W. P. (2014). How astronomers digitized the sky. *Technology and Culture*, 908-944.

McCray, W. P. (2017). The biggest data of all: Making and sharing a digital universe. *Osiris*, *32*(1), 243-263. https://doi.org/10.1086/693912

McLevey, J., Graham, A. V., McIlroy-Young, R., Browne, P., & Plaisance, K. S. (2018). Interdisciplinarity and insularity in the diffusion of knowledge: An analysis of disciplinary boundaries between philosophy of science and the sciences. *Scientometrics*, *117*(1), 331–349. https://doi.org/10.1007/s11192-018-2866-8

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Merton, R. K. (1968a). The Matthew effect in science. *Science*, *159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56

Merton, R. K. (1968b). *Social theory and social structure*. Simon and Schuster.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press.

Merton, R. K. (1988). The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *Isis*, *79*(4), 606–623. https://doi.org/10.1086/354848

Merton, R. K. (2000). On the Garfield Input to the Sociology of Science: A Retrospective Collage. In B. Cronin & H. B. Atkins (Eds.), *The web of knowledge: A festschrift in honor of Eugene Garfield* (pp. 435–448). Information Today.

Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, *83*(2), 340–363. https://www.jstor.org/stable/2778293

Milard, B. (2014). The social circles behind scientific references: Relationships between citing and cited authors in chemistry publications. *Journal of the Association for Information Science and Technology*, *65*(12), 2459-2468. https://doi.org/10.1002/asi.23149

Mitchell, J. C. (1969). The concept and use of social networks. In J. C. Mitchell (Ed.), *Social networks in urban situations: Analyses of personal relationships in Central African towns* (pp. 1–50). Manchester University Press.

Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, *69*(2), 213–238. https://doi.org/10.1177/000312240406900204

Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social studies of science*, *5*(1), 86-92.

Moreno, J. L., & Jennings, H. H. (1938). Statistics of social configurations. *Sociometry*, *1*(3/4), 342–374. https://doi.org/10.2307/2785588

Morris, S. A., & Van der Veer Martens, B. (2009). Mapping research specialties. *Annual Review of Information Science and Technology*, *42*(1), 213-295. https://doi.org/10.1002/aris.2008.1440420113

Morris, S.A. (2005). *Unified Mathematical Treatment of Complex Cascaded Bipartite Networks: The Case of Collections of Journal Papers*. PhD Thesis, Oklahoma State University. Retrieved from http://digital.library.okstate.edu/etd/umi-okstate-1334.pdf

Mulkay, M. (1976). The mediating role of the scientific elite. *Social Studies of Science*, *6*(3-4), 445-470.

Mulkay, M. J., (1974). Methodology in the sociology of science: Some reflections on the study of radio astronomy. *Social Science Information*, 13, 107–119. https://doi.org/10.1177/053901847401300206

Mulkay, M. J., & Edge, D. O. (1973). Cognitive, technical and social factors in the growth of radio astronomy. *Social Science Information*, *12*(6), 25-61.

Mulkay, M. J., Gilbert, G. N., & Woolgar, S. (1975). Problem areas and research networks in science. *Sociology*, 9(2), 187-203. https://doi.org/10.1177/003803857500900201

Mullins, N. C. (1972). The development of a scientific specialty: The phage group and the origins of molecular biology. *Minerva*, 10(1), 51-82. https://doi.org/10.1007/BF01881390

Mullins, N. C. (1973). The development of specialties in social science: The case of ethnomethodology. *Science Studies*, 3(3), 245-273. https://doi.org/10.1177/030631277300300302

Mullins, N. C. (1983). Theories and theory groups revisited. *Sociological theory*, *1*, 319-337.

Mullins, N. C., & Mullins, C. J. (1973). *Theories and theory groups in contemporary American sociology*. Harper & Row.

Mullins, N. C., Hargens, L. L., Hecht, P. K., & Kick, E. L. (1977). The group structure of cocitation clusters: A comparative study. *American Sociological Review*, 552-562. https://doi.org/10.2307/2094554

Neal, Z. (2014). The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. *Social Networks*, *39*, 84-97. https://doi.org/10.1016/j.socnet.2014.06.001

Newman, M. (2018). *Networks*. Oxford University Press.

Newman, M. E. (2001a). Clustering and preferential attachment in growing networks. *Physical Review E*, *64*(2), 1–4. https://doi.org/10.1103/PhysRevE.64.025102

Newman, M. E. (2001b). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*(2), 404-409.

Newman, M. E. (2001c). Scientific collaboration networks. II. Shortest paths, weighted networks, and cen-trality. *Physical Review E*, *64*(1), 016132.

Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*(suppl 1), 5200-5205. https://doi.org/10.1073/pnas.0307545100

Newman, M. E. (2009). The first-mover advantage in scientific publication. *EPL (Europhysics Letters)*, *86*(68001), 1-6. https://doi.org/10.1209/0295-5075/86/68001

Nicolaisen, J. (2008). Citation analysis. *Annual Review of Information Science and Technology*, *41*(1), 609-641. https://doi.org/10.1002/aris.2007.1440410120

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251. https://doi.org/10.1016/j.socnet.2010.03.006

Peng, T.-Q. (2015). Assortative mixing, preferential attachment, and triadic closure: A longitudinal study of tie-generative mechanisms in journal citation networks. *Journal of Informetrics*, *9*(2), 250–262. https://doi.org/10.1016/j.joi.2015.02.002

Perianes-Rodriguez, A., Waltman, L., & Van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics*, *10*(4), 1178–1195.

Perry, B. L., Pescosolido, B. A., & Borgatti, S. P. (2018). *Egocentric network analysis: Foundations, methods, and models*. Cambridge University Press.

Persson, O. (2010). Identifying research themes with weighted direct citation links. *Journal of Informetrics*, *4*(3), 415–422.

Powell, W. W., & Colyvas, J. A. (2008). Microfoundations of institutional theory. In: Royston Greenwood, Christine Oliver, Roy Suddaby and Kerstin Sahlin-Andersson (Eds.). *The Sage Handbook of Organizational Institutionalism* (pp. 276-298).

Powell, W. W., & Rerup, C. (2017). Opening the black box: The microfoundations of institutions. In: Royston Greenwood, Christine Oliver, Thomas B. Lawrence and Renate E. Meyer (Eds.). *The SAGE Handbook of Organizational Institutionalism* (pp. 311-337). Sage.

Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational Collaboration collaboration and the lLocus of iInnovation: Networks of lLearning in Biotechnologybiotechnology. *Administrative Science Quarterly*, *41*(1), 116–145. https://doi.org/10.2307/2393988

Price, D. J. (1963). *Little science, big science*. Columbia University Press.

Price, D.J. (1965). Networks of scientific papers. *Science*, *149*(3683), 510–515. https://doi.org/10.1126/science.149.3683.510

Price, D. J. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American society for Information Science*, *27*(5), 292-306. https://doi.org/10.1002/asi.4630270505

Purwitasari, D., Fatichah, C., Sumpeno, S., Steglich, C., & Purnomo, M. H. (2020). Identifying collaboration dynamics of bipartite author-topic networks with the influences of interest changes. *Scientometrics*, *122*, 1407–1443. https://doi.org/10.1007/s11192-019-03342-2

Raimbault, B. & Joly, P.B. (2021). The emergence of technoscientific fields and the new political sociology of science. In: K. Kastenhofer & S. Molyneux-Hodgson (Eds.). *Community and Identity in Contemporary Technosciences* (pp. 85-106). Sociology of the Sciences Yearbook, Vol 31. Springer.
https://doi.org/10.1007/978-3-030-61728-8_4

Ramos Zincke, C. (2014). Local and global communications in Chilean social science: Inequality and relative autonomy. *Current Sociology*, *62*(5), 704–722. https://doi.org/10.1177/0011392114521374

Rawlings, C. M., & McFarland, D. A. (2011). Influence flows in the academy: Using affiliation networks to assess peer effects among researchers. *Social Science Research*, *40*(3), 1001-1017.

Rawlings, C.M., McFarland, D. A., Dahlander, L., & Wang, D. (2015). Streams of thought: Knowledge flows and intellectual cohesion in a multidisciplinary era. *Social Forces*, *93*(4), 1687-1722.

Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The structure of founding teams: Homophily, strong ties, and isolation among US entrepreneurs. *American Sociological Review*, *68*(2), 195–222. https://doi.org/10.2307/1519766

Rigney, D. (2010). *The Matthew effect: How advantage begets further advantage*. Columbia University Press.

Ripley, R. M., Snijders, T. A. B., Boda, Z., Vörös, A., & Preciado, P. (2021). *Manual for RSiena*. University of Oxford: Department of Statistics; Nuffield College; University of Groningen: Department of Sociology.

Rivera, M. T., Soderstrom, S. B., & Uzzi, B. (2010). Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *Annual Review of Sociology*, *36*, 91–115. https://doi.org/10.1146/annurev.soc.34.040507.134743

Robins, G., & Pattison, P. (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, *25*(1), 5–41. https://doi.org/10.1080/0022250X.2001.9990243

Robins, G., Pattison, P., & Wang, P. (2009). Closure, connectivity and degree distributions: Exponential random graph (p*) models for directed social networks. *Social Networks*, *31*(2), 105–117. https://doi.org/10.1016/j.socnet.2008.10.006

Robins, G., Pattison, P., & Woolcock, J. (2005). Small and Other Worlds: Global Network Structures from Local Processes. *American Journal of Sociology*, *110*(4), 894–936. https://doi.org/10.1086/427322

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*, 351–357. https://doi.org/10.2307/2087176

Rosengren, K. E. (1968). *Sociological aspects of the literary system*. Natur och Kultur.

Roth, C., & Cointet, J. P. (2010). Social and semantic coevolution in knowledge networks. *Social Networks*, *32*(1), 16-29. https://doi.org/10.1016/j.socnet.2009.04.005

Salton, G., & McGill, M.J. (1983). *Introduction to modern information retrieval*. McGraw-Hill.

Schneider, J. W., Larsen, B., & Ingwersen, P. (2009). A comparative study of first and all-author co-citation counting, and two different matrix generation approaches applied for author co-citation analyses. *Scientometrics*, *80*(1), 103–130.

Schoch, D. (2021). Projecting signed two-mode networks. *The Journal of Mathematical Sociology*, *45*(1), 37-50. https://doi.org/10.1080/0022250X.2019.1711376

Schrum, W. & Mullins, N. (1988). Network analysis in the study of science and technology. In A.F.J van Raan (Ed.), *Handbook of Quantitative Studies of Science and Technology* (pp. 107–133). Elsevier Science Publishers B.V. (North-Holland). https://doi.org/10.1016/B978-0-444-70537-2.50009-X

Schubert, A. (2013). Measuring the similarity between the reference and citation distributions of journals. *Scientometrics*, *96*(1), 305-313. https://doi.org/10.1007/s11192-012-0889-0

Schubert, A., & Telcs, A. (2014). A note on the Jaccardized Czekanowski similarity index. *Scientometrics*, *98*(2), 1397-1399. https://doi.org/10.1007/s11192-013-1044-2

Schweinberger, M., & Luna, P. (2018). HERGM: Hierarchical exponential-family random graph models. *Journal of Statistical Software*, *85*(1).

Schweinberger, M., Krivitsky, P. N., Butts, C. T., & Stewart, J. R. (2020). Exponential-family models of random graphs: Inference in finite, super and infinite population scenarios. *Statistical Science*, *35*(4), 627–662. https://doi.org/10.1214/19-STS743

Sciabolazza, V. L., Vacca, R., Okraku, T. K., & McCarty, C. (2017). Detecting and analyzing research communities in longitudinal scientific networks. *PLOS ONE*, *12*(8), 1-23. https://doi.org/10.1371/journal.pone.0182516

Scott, J. (2011). Social physics and social networks. In J. Scott & P. J. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 55–66). Sage.

Scott, W.R. (1995). *Institutions and Organizations*. Thousand Oaks, CA: Sage.

Scott, W.R. & Meyer, J.W. (1991). The organization of societal sectors: Propositions and early evidence. In: W.W. Powell & P.J. DiMaggio (Eds.). *The New Institutionalism in Organizational Analysis* (pp. 108-140). University of Chicago Press.

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2009). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science and Technology*, *60*(3), 571–580.

Simmel, G. (1950). *The sociology of Georg Simmel*. Simon and Schuster.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, *42*(3/4), 425-440. https://doi.org/10.2307/2333389

Small H. (1997). Update on science mapping: Creating large document spaces. *Scientometrics*, *38*(2), 275-293. https://doi.org/10.1007/bf02457414

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265–269.

Small, H. (1977). A co-citation model of a scientific specialty: A longitudinal study of collagen research. *Social studies of science*, *7*(2), 139-166.

Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying and graphing specialties. *Science studies*, *4*(1), 17-40.

Smith, L. C. (1981). Citation analysis. *Library Trends*, 30(1), 83-106.

Snijders, T. A. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, *31*(1), 361–395. https://doi.org/10.1111/0081-1750.00099

Snijders, T. A. (2005). Models for Longitudinal longitudinal nNetwork dData. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and Methods in Social Network Analysis* (pp. 215–247). Cambridge University Press.

Snijders, T. A. (2016). The multiple flavours of multilevel issues for networks. In E. Lazega & T. A. B. Snijders (Eds.), *Multilevel Network Analysis for the Social Sciences: Theory, Methods and Applications* (Vol. 12, pp. 15–46). Springer. https://doi.org/10.1007/978-3-319-24520-1_2

Snijders, T. A. B. (2017). Stochastic actor-oriented models for network dynamics. *Annual Review of Statistics and Its Application*, *4*(343–363). https://doi.org/10.1146/annurev-statistics-060116-054035

Snijders, T. A. B., & Baerveldt, C. (2003). A multilevel network study of the effects of delinquent behavior on friendship evolution. *The Journal of Mathematical Sociology*, *27*(2–3), 123–151. https://doi.org/10.1080/00222500305892

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed). Sage.

Snijders, T. A. B., & Koskinen, J. (2013). Longitudinal models. In D. Lusher, J. Koskinen, & G. Robins (Eds.), *Exponential random graph models for social networks: Theory, methods and applications* (pp. 130–140). Cambridge University Press.

Snijders, Tom T. A. B., Faye, M., & Brailly, J. (2020). Network dynamics with a nested node set: Sociability in seven villages in Senegal. *Statistica Neerlandica*. https://doi.org/10.1111/stan.12208

Snijders, Tom T. A. B., Lomi, A., & Torló, V. J. (2013). A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social Networks*, *35*(2), 265–276. https://doi.org/10.1016/j.socnet.2012.05.005

Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, *36*(1), 99-153. https://doi.org/10.1111/j.1467-9531.2006.00176.x

Snijders, Tom T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New Specifications for Exponential Random Graph Models. *Sociological Methodology*, *36*(1), 99–153. https://doi.org/10.1111/j.1467-9531.2006.00176.x

Snijders, T. A. B., Van de Bunt, G. G., & Steglich, C. E. (2010). Introduction to sStochastic aActor-bBased Models models for Network network

Dynamicsdynamics. *Social Networks*, *32*(1), 44–60. https://doi.org/10.1016/j.socnet.2009.02.004

Snijders, T. A. B., & Steglich, C. E. (2015). Representing micro–macro linkages by actor-based dynamic network models. *Sociological Methods & Research*, *44*(2), 222–271. https://doi.org/10.1177/0049124113494573

Snijders, T. A.B. (2011). Network Dynamics. In J. Scott & P. J. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 501–513). SAGE Sage Publications.

Snijders, T.A.B. & Pickup, M. (2017). Stochastic actor-oriented models for network dynamics. In: J. N. Victor, A. H. Montgomery, M. Lubell (Eds.), *Oxford handbook Handbook of political Political Nnetworks* (pp.221–247). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190228217.013.10

Stadtfeld, C. (2018). The micro–macro link in social networks. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–15). https://doi.org/10.1002/9781118900772.etrds0463

Stadtfeld, C., & Block, P. (2017). Interactions, actors, and time: Dynamic network actor models for relational events. *Sociological Science*, *4*, 318–352.

Stadtfeld, C., Mascia, D., Pallotti, F., & Lomi, A. (2016). Assimilation and differentiation: A multilevel perspective on organizational and network change. *Social Networks*, *44*, 363–374. https://doi.org/10.1016/j.socnet.2015.04.010

Stadtfeld, C., Takács, K., & Vörös, A. (2020). The emergence and stability of groups in social networks. *Social Networks*, *60*, 129–145. https://doi.org/10.1016/j.socnet.2019.10.008

Stark, T. H., Rambaran, J. A., & McFarland, D. A. (2020). The meeting of minds: Forging social and intellectual networks within universities. *Sociological Science*, *7*, 433-464. https://doi.org/10.15195/v7.a18

Steglich, C., Snijders, T. A., & Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, *40*(1), 329-393. https://doi.org/10.1111/j.1467-9531.2010.01225.x

Stivala, A. D., Koskinen, J. H., Rolls, D. A., Wang, P., & Robins, G. L. (2016). Snowball sampling for estimating exponential random graph models for large networks. *Social Networks*, *47*, 167–188. https://doi.org/10.1016/j.socnet.2015.11.003

Tuire, P., & Erno, L. (2001). Exploring invisible scientific communities: Studying networking relations within an educational research community. A Finnish case. *Higher Education*, *42*(4), 493-513.

van den Besselaar, P., & Heimeriks, G. (2006). Mapping research topics using word-reference co-occurrences: A method and an exploratory case study. *Scientometrics*, *68*(3), 377–393. https://doi.org/10.1007/s11192-006-0118-9

van Eck, N. J., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, *60*(8), 1635-1651. https://doi.org/10.1002/asi.21075

Velden, T., Yan, S., & Lagoze, C. (2017). Mapping the cognitive structure of astrophysics by infomap clustering of the citation network and topic affinity analysis. *Scientometrics*, *111*(2), 1033–1051. https://doi.org/10.1007/s11192-017-2299-9

Visser, M., van Eck, N. J., & Waltman, L. (2020). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *ArXiv:2005.10732 [Cs]*. http://arxiv.org/abs/2005.10732

Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, *34*(10), 1608-1618.

Wallace, M. L., Larivière, V., & Gingras, Y. (2012). A small world of citations? The influence of collaboration networks on citation practices. *PLOS ONE*, *7*(3), 1-10. https://doi.org/10.1371/journal.pone.0033339

Wang, B., Bu, Y., & Xu, Y. (2018). A quantitative exploration on reasons for citing articles from the perspective of cited authors. *Scientometrics*, *116*(2), 675–687. https://doi.org/10.1007/s11192-018-2787-6

Wang, C., Butts, C. T., Hipp, J., & Lakon, C. M. (2020). Model adequacy checking/goodness-of-fit testing for behavior in joint dynamic network/behavior models, with an extension to two-mode networks. *Sociological Methods & Research*, 1-34. https://doi.org/10.1177/0049124120914933

Wang, D. & Barabási, L. (2021). *The science of science*. Cambridge University Press.

Wang, F., Jia, C., Wang, X., Liu, J., Xu, S., Liu, Y., & Yang, C. (2019). Exploring all-author tripartite citation networks: A case study of gene editing. *Journal of Informetrics*, *13*(3), 856–873. https://doi.org/10.1016/j.joi.2019.08.002

Wang, P., Robins, G., Pattison, P., & Lazega, E. (2013). Exponential random graph models for multilevel networks. *Social Networks*, *35*(1), 96–115. https://doi.org/10.1016/j.socnet.2013.01.004

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. Cambridge University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Wenger, E. (1998) *Communities of practice: learning, meaning, and identity*. Cambridge University Press.

White, H. C., & Breiger, R. L. (1975). Pattern across networks. *Society*, *12*(5), 68-74. https://doi.org/10.1007/BF02699922

White, H. C., Boorman, S. A., & Breiger, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology*, *81*(4), 730–780. https://doi.org/10.1086/226141

White, H. D. (2001). Authors as citers over time. *Journal of the American Society for Information Science and Technology*, *52*(2), 87–108. https://doi.org/10.1002/1097-4571(2000)9999:9999<::AID-ASI1542>3.0.CO;2-T

White, H. D. (2003). Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology*, *54*(5), 423–434. https://doi.org/10.1002/asi.10228

White, H. D. (2011). Scientific and Scholarly Networks. In J. Scott & P. J. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 271–285). Sage.

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, *32*(3), 163–171. https://doi.org/10.1002/asi.4630320302

White, H. D., Wellman, B., & Nazer, N. (2004). Does citation reflect social structure? Longitudinal evidence from the 'Globenet' interdisciplinary research group. *Journal of the American Society for information Science and Technology*, *55*(2), 111-126. https://doi.org/10.1002/asi.10369

White, H.D. (2004). Reward, persuasion, and the Sokal Hoax: A study in citation identities. *Scientometrics*, *60*(1), 93-120.

Whitley, R. D. (1970). Black boxism and the sociology of science: A discussion of the major developments in the field. *The Sociological Review*, *18*(1), 61-92.

Whitley, R. D. (1983). From the sociology of scientific communities to the study of scientists' negotiations and beyond. *Social Science Information*, *22*(4-5), 681-720.

Woolgar, S. W. (1976). Writing an intellectual history of scientific development: The use of discovery accounts. *Social Studies of Science*, *6*(3-4), 395-422.

Wooten, M., & Hoffman, A. J. (2017). Organizational fields past, present and future. In R. Greenwood, C. Oliver, T. B. Lawrence, & R. E. Meyer (Eds.), *The SAGE Handbook of Organizational Institutionalism* (pp. 130–148). Sage.

Zhang, C., Bu, Y., Ding, Y., & Xu, J. (2018). Understanding scientific collaboration: Homophily, transitivity, and preferential attachment. *Journal of the Association for Information Science and Technology*, *69*(1), 72–86. https://doi.org/10.1002/asi.23916

Zhao, D. (2006). Towards all-author co-citation analysis. *Information Processing & Management*, *42*(6), 1578–1591. https://doi.org/10.1016/j.ipm.2006.03.022

Zhao, D., & Strotmann, A. (2008). Author bibliographic coupling: Another approach to citation-based author knowledge network analysis. *Proceedings of the American Society for Information Science and Technology*, *45*(1), 1–10. https://doi.org/10.1002/meet.2008.1450450292

Zhao, D., & Strotmann, A. (2014). The knowledge base and research front of information science 2006–2010: An author cocitation and bibliographic coupling analysis. *Journal of the Association for Information Science and Technology*, *65*(5), 995–1006. https://doi.org/10.1002/asi.23027

Zhou, Q., & Leydesdorff, L. (2016). The normalization of occurrence and C o-occurrence matrices in bibliometrics using Cosine similarities and Ochiai coefficients. *Journal of the Association for Information Science and Technology*, *67*(11), 2805-2814. https://doi.org/10.1002/asi.23603

Žiberna, A. (2014). Blockmodeling of multilevel networks. *Social networks*, *39*, 46-61. https://doi.org/10.1016/j.socnet.2014.04.002

Zinilli, A. (2016). Competitive project funding and dynamic complex networks: Evidence from Projects of National Interest (PRIN). *Scientometrics*, *108*(2), 633–652. https://doi.org/10.1007/s11192-016-1976-4

Zuccala, A. (2006). Modeling the invisible college. *Journal of the American Society for information Science and Technology*, 57(2), 152-168. https://doi.org/10.1002/asi.20256

Zuckerman, H. (1967). Nobel laureates in science: Patterns of productivity, collaboration, and authorship. *American Sociological Review*, 32(3), 391-403. https://doi.org/10.2307/2091086