



Socio-cognitive networks between researchers: Investigating scientific dualities with the group-oriented relational hyperevent model^{☆,☆☆}

Alejandro Espinosa-Rada^{a,b},^{*} Jürgen Lerner^c, Cornelius Fritz^d

^a Instituto de Sociología, Pontificia Universidad Católica de Chile, Chile

^b Social Networks Lab, ETH Zurich, Switzerland

^c University of Konstanz, Germany

^d School of Computer Science and Statistics, Trinity College Dublin, Ireland

ARTICLE INFO

Keywords:

Duality
Scientific networks
Science of science
Relational event model
Hyperevents

ABSTRACT

Understanding why researchers cite certain works remains a key question in the study of scientific networks. Prior research has identified factors such as relevance, group cohesion, and source crediting. However, the interplay between cognitive and social dimensions in citation behavior – often conceptualized as a socio-cognitive network – is frequently overlooked, particularly regarding the intermediary steps that lead to a citation. Since a citation first requires a work to be published by a set of authors, we examine how the structure of coauthorship networks influences citation patterns. To investigate this relationship, we analyze the citation and collaboration behavior of Chilean astronomers from 2013 to 2015 using the Group-Oriented Relational Hyperevent Model, which allows us to study coauthorship and citation networks in a joint framework. Our findings suggest that when selecting which works to cite, authors favor recent research and maintain cognitive continuity across cited works. At the same time, we observe that coherent groups – closely connected coauthors – tend to be co-cited more frequently in subsequent publications, reinforcing the interdependence of collaboration and citation networks.

1. Introduction

Why do researchers cite each other? Citation is one of the most relevant indicators to measure the history of knowledge, and the impact and recognition of researchers, which, as a consequence, reinforces some level of hierarchy in science (de Solla Price, 1965; Crane, 1972; Cole and Cole, 1973; Merton, 1988; Bellotti and Espinosa-Rada, 2025). Still, understanding why researchers cite each other is a longstanding conjecture in studying scientific networks. Existing theories suggest that authors cite relevant contributions, cite their group of reference, or cite other work because of its honest contribution to their publications (Nicolaisen, 2007). White et al. (2004) also emphasize interpersonal networks, such as coauthorship networks. However, prior research often overlooked the dual nature of cognitive and social dimensions (Mützel and Breiger, 2020).

Citations help researchers trace existing knowledge, understand its diffusion, and examine how peer recognition shapes intellectual contributions. Knowledge evolves through recursive framing, influenced

by social structures such as communication, interactions, and relations among scholars, mentorship, and groups. This interplay between knowledge and social structures forms a socio-cognitive network, where thinking itself is shaped by social interactions and the knowledge available. In citation practices, references reflect both intrinsic intellectual contributions and the influence of scholarly networks. The concept of dualities (Breiger, 1974) allows researchers to analyze these intertwined processes, highlighting the mutual impact of groups of scholars and knowledge development.

While different interpretations exist in the literature, there is no agreed-upon theory on how social or cognitive ties are interrelated and which are the main mechanisms driving accumulation processes. Are citations accumulated because of intellectual merits or because individuals know each other? To unravel this puzzle, we investigate the tendency of researchers' citations by examining the concurrent interplay of socio-cognitive ties through coauthorship and citation networks.

[☆] The authors acknowledge support from Fondecyt Regular (ANID) under grant number 1220560 (AER) and the German Research Foundation under grants 321869138 (JL) and 520770522 (CF).

^{☆☆} We thank members of the Social Networks Lab at ETH Zürich for helpful comments and suggestions that considerably improved this work.

^{*} Corresponding author at: Instituto de Sociología, Pontificia Universidad Católica de Chile, Chile.

E-mail address: anespinosa@uc.cl (A. Espinosa-Rada).

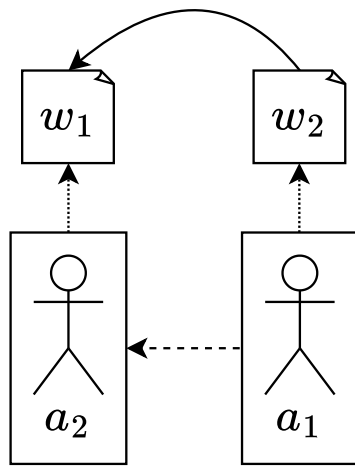


Fig. 1. Illustrative example of indirect and direct interactions between actors and publications.

Understanding citation patterns is challenging, although they have been extensively criticized as a simplistic measure (Edge, 1979) that is often under-theorized (Leydesdorff, 1998; Nicolaisen, 2007). However, to disentangle their structural properties, recent research has analyzed how they are embedded in a social context to identify the main mechanisms underlying why authors cite each other and how different networks are interrelated to explain citation tendencies (White et al., 2004; Espinosa-Rada et al., 2024; Lerner et al., 2024). This literature followed the tradition of the network researchers working on the sociology of science and knowledge that investigates social circles (Bellotti and Espinosa-Rada, 2025), as groups that are comprised of scientists who work on similar research problems that are usually aware of each other and maintain a high level of informal communication allowing them to navigate the complexity of science by creating social organizations of this kind beyond their institutional affiliations (Crane, 1969). These social circles or invisible colleges can be investigated using bibliographic data through the lens of the duality of socio-cognitive ties (Kadushin, 1966; Crane, 1969; Breiger, 1974).

The literature on author citation networks often relies on aggregate data and has not yet explored the intermediary role of works without aggregating the information (e.g., Small, 1973 and Zhao and Strotmann, 2008). This gap arises due to difficulties considering the complex and interdependent latent mechanisms in citation data. As illustrated in Fig. 1, for actor a_1 to cite actor a_2 , a_1 must first publish work¹ w_2 referencing another work w_1 (co-)authored by actor a_2 . The numerous ways of representing the citation process underscore the socio-cognitive dimension of scientific networks and give rise to multiple interdependent “dualities”. Authors become indirectly related by coauthoring the same work, citing the same works or authors, and being cited by the same works or authors. The authors become indirectly related via common authors—but also by citing, or being cited by, the same works or authors. Considering multiple paths reveals various “dualities” that depend on one another. Fig. 2 further illustrates these dualities. For instance, the citation relationship between two works may depend on the coauthorship relations between the authors. Disregarding this dependence results in information loss and biased results. In what follows, nodes correspond to scholars and works, and ties to (co)authorship and citation.

In this paper, we explore the dual relationship between academic citations and coauthorship to shed light on this complex phenomenon by focusing on how authors cite other authors through works. We start

by laying out the available theory on the concept of “duality” and bibliographic patterns in Section 2. In Section 3, we theorize about academic authors’ interdependencies between citation and collaboration behavior to identify whether the collaboration begets citations and more collaborations. As a case study, we analyze the collaboration and citation data among Chilean astronomers from 2013 to 2015, introduced in Section 4. We then propose a statistical model for the available data in Section 5 unifying previous research findings on citation and coauthor behavior (e.g., Matthew effect and group effect) and our novel theoretical insights about the interplay between coauthorship and citation. We hypothesize that effects drawing on “dualities” in socio-cognitive networks among scientific works and researchers provide additional explanations for “who coauthors with whom” and “who cites whom”. This model allows us to simultaneously consider different “dualities” in scientific networks by studying citations and collaboration ties in a joint framework. Thus, we advance the duality approach by using all available information in bibliometric data. Against this background, we assess our claims by employing a group-oriented variant of the relational hyperevent model (Lerner and Lomi, 2023; Lerner et al., 2024) in Section 6. Finally, Section 7 provides a discussion of the main findings.

2. Theory

Citations help researchers identify the knowledge available to others and examine how this knowledge spreads and evolves through peer recognition and scholarly contributions. Network researchers argue that the emergence and evolution of ideas – understood as sets of beliefs and bodies of knowledge – are produced, in part, by groups of scientists and that it is the social order governing such groups that shapes the way ideas are formed and accepted (Bellotti and Espinosa-Rada, 2025). Knowledge is framed by existing ideas about what is already known, serving as an initial reference point. These frames are constructed from available and necessary information, shaped by how it is gathered and processed, and refined through a recursive process that enables researchers to develop new conceptual frames (Carley, 1986). Groups of scientists further modulate these frames, encompassing regular communication, interactions, and relationships among researchers, participation in “coherent groups”, intergenerational mentor–mentee networks, and other intellectual settings. These structures serve as pathways to an individual’s cognitive processes. As an individual’s frame evolves and becomes accessible to others, it can influence their position within the social network and the knowledge they engage with, and vice versa (Mullins and Mullins, 1973; Carley, 1986; Collins, 2002).

The intrinsic relationship between knowledge production, dissemination, and underlying social structures is known as a socio-cognitive network, as these elements are fundamentally interconnected (Carley, 1986; White et al., 2004). As a cultural practice, knowledge is learned, discussed, and shared among scholars with similar interests. Thus, thinking can be perceived as internalizing cognitive structures shaped by available knowledge and social networks. As an individual’s frame evolves, their cognitive understanding and position relative to others within the network may shift.

In the context of citations, the socio-cognitive network is partly reflected in how references to others’ works and their authors diffuse—whether or not the cited authors are personally known to the person citing. Behind each act of referencing are authors and research teams that collaborate and collectively decide which works to cite as a foundation for further research. These references may be selected based on the intrinsic ideas they convey – regardless of authorship as publicly available knowledge – or due to the relevance of the scholars behind them, who function as “coalitions in the mind” (Collins, 2002) and serve as intellectual reference groups. This interconnectedness underscores the socio-cognitive nature of citation networks. Moreover, developments using dualities (Breiger, 1974) have enabled researchers

¹ Note that we use the terms publication and work interchangeably.

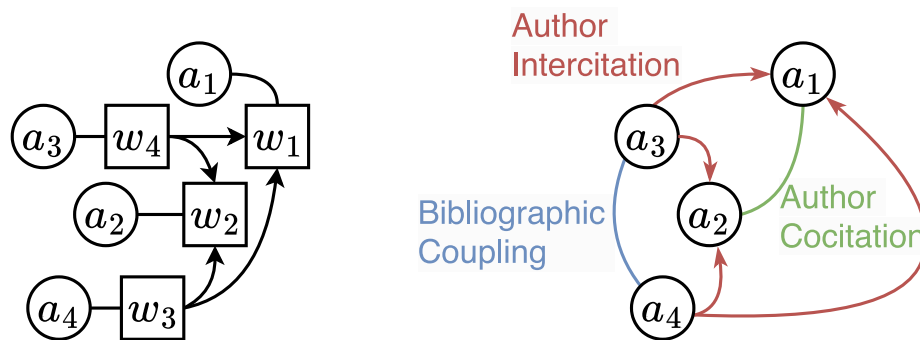


Fig. 2. Illustrative examples of three types of representation. The observed citation and collaboration data are represented in the left graph. On the right, the green, red, and blue lines represent the ties resulting from author cocitation, bibliographic coupling, and author intercitation, respectively.

to represent these relationships and assess the relative importance of different processes shaping the network.

“Duality”, as a concept, was initially associated with the intersection of social circles following the tradition of Georg Simmel. This concept was related to relationships among actors of different levels (e.g., individuals and organizations) through membership relations (Breiger, 1974). The main reasons for the intersections are individuals’ shared interests, personal affinities, or ascribed status of members who regularly participate in collective activities. Breiger (1974) also demonstrated that the representation of a two-mode network can act as a proxy to create two different networks, where a set of actors can be connected due to a shared affiliation just like groups to which the actors belong are connected via overlapping memberships. By projecting the matrix, a rectangular matrix resulted in two square matrices. New extensions of the concept of “duality” aimed to go beyond structural representations and consider cultural forms such as shared objects, symbols, or expressions of taste (Mützel and Breiger, 2020).

Network researchers working on the sociology of science and knowledge often associate social circles – the sociological phenomenon behind the analytical concept of “duality” – with invisible colleges when analyzing researchers. In this literature, researchers are grouped together because they interact, have a common interest in shared topics, and do not need to know each other to be influenced by other members (Kadushin, 1966; Crane, 1969). This type of social circle requires both a social and cognitive dimension. As Zuccala (2006) clarified, an invisible college is a set of interacting scholars that share similar research interests concerning a subject specialty, as the propositional knowledge available from an intellectual group of references (“a coalition in the mind”). A subject specialty informs the invisible college of its rules and research problems and supports the intellectual motivation for social activity. These invisible colleges should not separate the subject specialty they come from the social aspect of science (Mullins, 1972) such as personal communication, interactions, social relationships, groups, mentor–mentees or contexts in which knowledge takes place, making the socio-cognitive dimension of science explicit. In the postscript of Kuhn (2012), he mentioned that paradigms are better understood when the community structure of science is taken into account, as was investigated by researchers of the time using the social network perspective (e.g., Hagstrom, 1965, de Solla Price and Beaver, 1966, and Crane, 1969). And Merton (2000, p. 437) equated the concept of the “invisible college” (Crane, 1969) with socio-cognitive networks to explain the genesis and transmission of knowledge.

Duality can leverage science’s socio-cognitive dimension by explicitly explaining how individuals create social ties by considering these two dimensions together. Regarding scientific networks, there is a longstanding tradition emphasizing that researchers should consider social and cognitive ties to analyze scientific networks (Crane, 1972; Merton, 2000; White et al., 2004; White, 2011; Bellotti and Espinosa-Rada, 2025). The main reason therein is that each type of tie measures

something different, and separating them can lead to distorted representations of the underlying network (Holland and Leinhardt, 1974; Chubin, 1976). Still, researchers often analyze social or cognitive ties separate from one another. For example, for Moody (2004), citation networks are not social networks because the social ties do not capture the informational interaction structure of the latter. Schrum and Mullins (1988) distinguished between “interactions” and “interest”. The former mechanism implies communication, information flow, or general contact (such as coauthorship and “in-house” citation). In contrast, the latter is represented by citing the same papers (i.e., co-occurrence of citations in bibliographies). Leydesdorff and Vaughan (2006) argues that co-occurrences in bibliometric research represent variables attributed to texts, which is different from social networks that often refer to concrete relations (such as “affiliations”).

Collaboration and citation. Citations are a manifestation of formal but asymmetric communication between two scientists: if researcher a_1 refers to a work authored by another researcher a_2 , it is presumed that the cited work was helpful in a_1 ’s research (Chubin, 1976). For Small (1978), citations are symbols of concepts and ideas expressed in language, as cited works embody ideas that authors discuss in their work. On the other hand, to characterize the structure of a scientific field, researchers often use coauthorship networks (Newman, 2001; Moody and Light, 2006) as a proxy of interpersonal relationships to identify scientists’ communication as a social dimension of science. These researchers distinguished between cognitive and social ties.

Separating citation and collaboration as cognitive and social dimensions added another layer of distortion by assuming the citation is cognitive without any social component interweaving with cognitive ties. The overlapping nature of these two types of relations is referred to as a socio-cognitive network (Merton, 2000; White et al., 2004), and has been recently studied as a co-evolving process that allowed disentangling whether the two networks influence each other (Espinosa-Rada et al., 2024; Lerner et al., 2024) specifically by considering a delimited context. Researchers relied on intercitation, defined as “the record of who has cited whom within a fixed set of authors” (White, 2011, p. 275) to explore in bounded contexts, and assuming awareness, the dual nature of socio-cognitive networks such as “in-house” (i.e., same institution) relationships (Chubin and Studer, 1979; Schrum and Mullins, 1988). By delimiting the context, researchers can investigate in more detail how the social dimension of scientific networks unfolds by identifying whether citations also have a social component. The resulting socio-cognitive process is inherently part of bibliometric data, leading to many possible linked networks (Batagelj and Cerinšek, 2013) manifesting the duality (Breiger, 1974; Mützel and Breiger, 2020) of scientific networks. From the same scientific work, different networks can be derived, such as coauthorships, in which two authors are linked if they produce a joint work (e.g., paper, book, presentation), or a citation network, in which authors can cite many other different works, which in both cases represents a two-mode network. Bibliographic data, as a

product of science, allow for the tracing of many formal communication channels in science.

Types of representations. In the study of author-to-author citation networks, researchers often use aggregated network representations to investigate social and cognitive ties without explicitly considering the intermediate role of works. For instance, author cocitation, where “*authors whose works are generally seen to be related, and are repeatedly cited as such in later documents, tend to cluster together on the map, while authors who are rarely or never cited together are relatively far apart*” (White and Griffith, 1981, p. 164). While a single work is implicitly considered a building block of the representation, an author cocitation approximates how the same works cite two authors through a two-mode network.

Another strategy that uses citation-based representations is author bibliographic coupling (Zhao and Strotmann, 2008). Contrasting cocitation, two individuals are assumed to be closer if they cite the same references. In this case, the focus lies on the research front, i.e., who they cite, rather than the knowledge agreement, which relates to how other publications cite them.

Finally, author intercitation (also referred to as author direct citation or cross-citation) is a third representation based on direct relationships between authors through citations without including a third-party work (White, 2011). By considering authors and works in a chain, it is apparent that for a_1 to cite actor a_2 , a_1 must first publish work w_2 referencing another work w_1 (co-)authored by actor a_2 (see Fig. 1). As in cocitation and bibliographic coupling, the intercitation is often treated as an aggregated matrix, assuming that the frequency of citation among a_1 and a_2 measures direct relation strength between them (Wang et al., 2012).

In all these representations (represented in Fig. 2), the available timestamped interaction events are boiled down to the frequency of shared works to indicate the strength of a dyadic tie. However, treating the data as weighted ties raises new problems because works with more authors or references can overestimate their prevalence, requiring new techniques to normalize the credit given to authors (e.g., Batagelj, 2020). At the same time, it is unlikely to explicitly address the contribution of the chain of the entire set of authors and the whole sets of works in these network representations. A representation that can explore the duality of persons and groups as pushed forward by Breiger (1974) can capture higher-order dependencies typically present in networks with these characteristics. For example, if work w_1 cites another work w_2 , it may depend on the coauthorship relations between the authors of both papers.

The aggregated representation of the ties assumes that the frequency of links between entities reflects the durability of the underlying structure of scientific networks. Nonetheless, analyses of the precise order and repeated interactions via authorship or citations in science have received little attention (some exceptions are Lerner and Hânclean, 2023 and Lerner et al., 2024). For instance, works are events or instances in science that are scientific productions generated by an author or a team of researchers that refer to previous works by citing the references that justify the stands of the work. As Garfield (1964, p. iii) mentioned, the history of science depends on the sequence of events on which each discovery depends. Compared with aggregated measures, events in scientific networks allow the study of links between researchers and other entities (Hummon and Doreian, 1989).

3. Socio-cognitive mechanisms

Researchers investigating mechanisms underlying scientific networks consider different network mechanisms (Rivera et al., 2010; Espinosa-Rada et al., 2024). These relational mechanisms are based on dyadic similarity (e.g., homophily), relationships (e.g., Matthew effect or group structures), and proximity-based mechanisms (e.g., focuses of activity). Some mechanisms represent general patterns, while others can be dissected into concrete network representations that explicitly

show entities' internal structure and relations (Stadtfeld and Amati, 2021).

Socio-cognitive networks are complex structures that can be understood through the lens of mechanisms involving a mixture of different entities (e.g., authors and works) and ties (e.g., citations and collaborations). One can explore these structures in fine-grained data as micro-temporal patterns (Butts et al., 2023), considering the temporal order of relational events over specific time scales. We present four general mechanisms involving socio-cognitive structures and then suggest more concrete micro-temporal mechanisms to explore these patterns.

Matthew effect of authors. We investigate the Matthew effect (Zuckerman, 1967; Merton, 1968) by focusing on the authors of the papers as one of the primary explanations for why researchers receive more recognition over time. Originally, Harriet Zuckerman and Robert K. Merton were interested in how the allocation of credit in cases of collaboration affects the flow of ideas through the communication network of science. The Matthew effect highlighted the bias of allocating more recognition to renowned researchers while reducing the visibility of contributions of less well-known authors (Zuckerman, 1967; Merton, 1968; Cole and Cole, 1973; Merton, 1988). One instance of this mechanism is the accumulation of work citations (de Solla Price, 1965) and collaborators (Newman, 2001; Barabási et al., 2002). By analyzing collaboration and citations separately, the pattern becomes a self-fulfilled prophecy in which collaborations lead to more collaborators. Citations of a particular work lead to more citations of that work, thus the authors of that work accumulate more recognition. Nonetheless, it is unclear how authors, collaboration, and citations imply cumulative processes when they are analyzed together. Authors can gain more recognition by receiving more citations; this accumulation might occur because a single work becomes highly visible, the author's portfolio of documents accumulates more citations, or both. These accumulation processes can also result from a group of researchers reinforcing the recognition or a consensus among the broader scientific community. In other words, is it the work or the author that gets repeatedly cited? Further, do they get cited repeatedly by the same or different authors? Due to the dual accumulation process by individual papers or the author's entire portfolio, an assortativity degree process underlies the mechanism. This is reasonable in scientific networks, as it represents the reinforcement of active actors according to the Matthew effect and their more visible positions within groups (Briege, 1976; Mullins et al., 1977).

Hypothesis 1 (H1): *Actors tend to send more citations to other actors that have received more citations before.*

Intercitation. Intercitation or author-to-author citation occurs when members of a contextually bounded group cite each other. White et al. (2004) investigate whether intercitation varies according to acquaintanceship and communication between members of these groups, intellectual affinities that are paramount regardless of the social dimension, or a combination of both. Intercitation allows focusing on asymmetric relationships since one author citing another author does not imply that the latter cites the former. For intercitations, we can distinguish if an author cites another author because they are collaborators or if, because they cite each other, they will collaborate. We further dissect this mechanism by considering the hyperevent taking into consideration author a_1 publishing a work w_1 that cites another work w_2 to a different author a_2 .

Hypothesis 2a (H2a): *Authors tend to cite other authors they had cited before.*

Hypothesis 2b (H2b): *Authors tend to collaborate with other authors they had cited before.*

Hypothesis 2c (H2c): *Authors tend to cite coauthors' works.*

Table 1

Descriptive statistics on the number of authors and works per year in the Chileans Citation Network.

Year	2013	2014	2015
Number of authors	87	87	87
Number of citing papers	322	345	367
Average number of coauthors	2.39	2.18	2.25
Average number of references	59.27	56.74	56.33

Author cocitation. White (2011) challenged that intercitation was one of the main effects that explain citation because he believes that the “true glue” binding scientists and scholars together is what members can competently write about rather than whom they know. To him, social and affective ties are secondary to intellectual relevance. To achieve this conclusion, he explored the author cocitation mechanism as a measure that controls for the propensity of any work by an author that also appears in any work of another author to appear in the references of a later work (White and Griffith, 1981). White et al. (2004) write that “[b]ecause scholars are cited together for many reasons, cocitation data can be noisy, but in the aggregate, they are a robust measure of how citers view the intellectual linkages in a research domain” (p. 115). This approach considers the global community, and as long as two researchers are cocited from anyone else, they would appear together in a network representation (as a symmetric weighted tie). White et al. (2004) considered that cocitation and intercitation could be conflated in bounded settings because if author a_1 cites himself and another author a_2 in the same network, it will increment the author a_1 to author a_2 intercitation and the author a_1 to author a_2 cocitation. Nonetheless, this conflation appears when self-citations are considered.

Hypothesis 3 (H3): *Authors that are frequently cocited will tend to be cocited together more often in future publications.*

Group effect. The last mechanism we are interested in is related to the relevance of groups. Mullins (1972, 1973) and Griffith and Mullins (1972) proposed that “coherent groups” – which are considered small and intensely interacting research groups – are the primary drivers of scientific change and seed larger invisible colleges that develop around them. Dyadic and triadic structures of multiple ties represent these groups, allowing the emergence of more complex morphological structures (Mullins and Mullins, 1973). Recently, researchers have used similar structures by considering one layer in collaboration networks (Kronegger et al., 2011; Ferligoj et al., 2015; Stark et al., 2020; Witte et al., 2023).

However, Espinosa-Rada et al. (2024) proposed a cross-network closure that involves citation and collaboration network mechanisms. They operationalized it as “the tendency to cite an author if two researchers share a joint coauthor and the tendency of two actors collaborating to be cited by the same authors” (pp. 98–99). A further possible distinction of cross-network closure is between influence and selection processes. The “coherent group” can influence other members by encouraging the authors to adopt the past citations of their past coauthors or to cite the papers of their coauthors. They might also expand the group by being cocited by third authors. In both cases, the group expands by indirectly agreeing with the accumulated knowledge shared by a paradigmatic group or by directly being cocited with those with whom they agree.

Hypothesis 4a (H4a): *Coauthors tend to cite similar references in their publications.*

Hypothesis 4b (H4b): *Coauthors tend to be cocited in subsequent publications.*

4. Data

For the exploration of the socio-cognitive mechanisms, we make use of data involving Chilean astronomers (for details, see Espinosa-Rada, 2021 and Espinosa-Rada et al., 2024). The studied time frame corresponds to the local group formation period of astronomers and astrophysics a few years after the Atacama Large Millimeter/submillimeter Array became fully operative in 2013. The bibliographic data were initially gathered from the Web of Science, and additional information, such as the researcher’s nationality, was collected from the academics’ CVs. We use the Web of Science ID as a proxy of the time order, representing the date the paper was indexed into the database.

Network boundary. We restrict the data to researchers affiliated with organizations settled in Chile and, thereby, have access to all the telescopes in the country. For the analysis, we consider the authors and publications in which at least one researcher is affiliated with an organization settled in Chile participating in 2013 – 2015. All authors that are not settled in Chile are excluded from the analysis. Note, however, that foreigners can be settled in Chile. The cited papers are the works of this cohort published between 1947 and 2015.

Descriptive statistics. For each year, additional information on the number of authors and citing papers is given in Table 1. The aggregated network data can be considered dual because it involves different types of nodes (works and authors) and its multiplex nature (connections based on coauthorship and citation). The two networks that distinguish between coauthor and citation relations are visualized in Figs. 3(a) and 3(b), respectively. While this visualization can be helpful for the exploration of the network, recent trends in network modeling for scientific networks aim to go beyond projections to analyze the two-mode structure of the network (for examples and further discussion, see Espinosa-Rada and Ortiz, 2022, Fritz et al., 2023, and Gallagher et al., 2023).

5. Methods

5.1. Group-oriented hyperevent model

Denote the set of all N astronomers by $\mathcal{A} = \{1, \dots, N\}$ and the set of documents published up until but not including time point $t \in \mathcal{T}$ by $\mathcal{W}(t)$, with \mathcal{T} being the set including all possible timestamps between 2013 and 2015 for which the bibliographic data is available. To shorten the notation, we assume that the static information on whether a particular astronomer in the data is Chilean is also contained in $\mathcal{W}(t)$. In this context, the observed coauthorship and citation data are group-to-set relational events, i.e., where the sender is a group of actors and the receiver a set of works. A publication $w = (A(w), C(w), t(w))$ encompasses an author set $A(w) \subseteq \mathcal{A}$, citation set $C(w) \subseteq \mathcal{W}(t(w))$, and timestamp $t(w) \in \mathcal{T}$. As described in Section 4, only the order, not the exact timing of each work, is available in the publication records. Therefore, we assume that each work w corresponds to an arbitrary timestamp $t(w)$ such that the order of the times corresponds to the observed order. Further, $A(w)$ represents a group of authors, while $C(w)$ represents a set of works. Therefore, we term this type of data group-to-set relational events. The differentiation between groups of authors and sets of works also highlights that one can comprehend these sets as a two-mode structure of authors and works (Breiger, 1974) or a three-mode structure (Fararo and Doreian, 1984) if we include time, as they also involve authors, citing, and cited work. The set of citable works at time point t is then defined by $\mathcal{W}(t) = \{w; t(w) < t\}$, and the number of authors and references in $A(w)$ and $C(w)$ is denoted by $|A(w)|$ and $|C(w)|$, respectively. Finally, the set of sets of all possible citation lists of length $k \in \{1, 2, \dots\}$ with information available up until but not including time point $t \in \mathcal{T}$ is given by $\mathcal{I}(t, k)$.

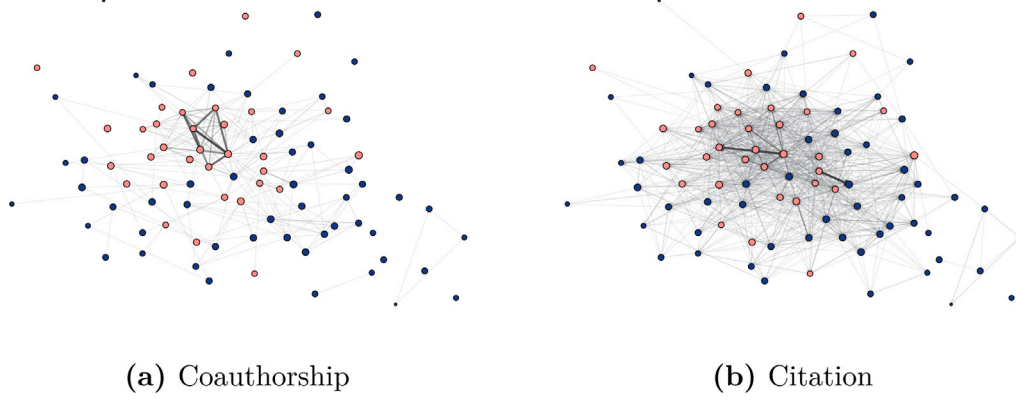


Fig. 3. Intercitation network of Chilean astronomers (2013–2015). The size of the nodes is the log-transformed number of accumulated citations, the edges are the weighted ties (number of works), and the colors represent if the nodes are Foreigners (pink) or Chileans (blue).

REM for higher-order events. Lerner et al. (2021) extended the relational event model (REM) to events involving multiple actors. The available data from Section 4 can, however, best be comprehended as events between groups of authors and sets of cited works, a data type for which we use the name *group-to-set events*. Lerner et al. (2024) proposed an extension of undirected hyperevents (Lerner et al., 2021) and directed one-to-many hyperevents (Lerner and Lomi, 2023) in a one-mode network to group-to-set events in a two-mode network, which we next amend to our setting. Instead of specifying the intensity of a dyadic interaction, such as author a_1 to cite some work by author a_2 at time t , we state a joint intensity to write a work with coauthors A and citations to works C at time t . This intensity characterizes a multivariate counting process that counts how often each possible work (encompassing any number of coauthors and cited works) was written until arbitrary time point t . A similar model – analyzing the interrelation between collaboration networks and references to previous work in cultural production – has been applied by Burgdorf et al. (2024). We discuss alternative models in the Appendix A.

Group-oriented formulation. For the empirical setting of Chilean astronomers, we argue that coauthors affect which works are cited, but, reversely, the cited work does not affect the coauthors (Espinosa-Rada et al., 2024). Therefore, we assume that the set of authors (“sender”) is first determined for publication, then the references (“receiver”) are decided upon conditional on the set of authors. Our model, thereby, parallels other network actor-oriented models such as the stochastic actor-oriented model (Snijders, 2001) or the dynamic network actor model (Stadtfeld et al., 2017; Stadtfeld and Block, 2017).

Mathematically, this implies a conditional independence assumption between the set of authors and its reference list, yielding the *Group-Oriented Relational Hyperevent Model*. This model comprises an author model determining the coauthors and a citation model governing the citations conditional on the set of authors and the size of the citation list. Both models take the general form of a relational event model for hyperevents involving multiple actors as proposed by Lerner et al. (2021). Since the citation model governs the decision on a particular set of citations conditional on the set of authors and the size of the citation list, we state it as a multinomial choice model proposed by McFadden (1973). For both models, the available information at timepoint $t \in \mathcal{T}$ includes the entire coauthorship and citation past, denoted by $\mathcal{W}(t)$. As a result, the intensity to observe a publication of the set of authors $A \subseteq \mathcal{A}$ with references to $C \subseteq \mathcal{W}(t)$ at time point t is given by:

$$\lambda_{A,C}(t | \theta, \gamma, |C|) = \underbrace{\lambda_A(t | \theta)}_{\text{Author Model}} \underbrace{p_{C|A}(t | \gamma, A, |C|)}_{\text{Citation Model}} \quad (1)$$

with

$$\lambda_A(t | \theta) = \lambda_{0,A}(|A|, t) \exp(\theta^\top s(\mathcal{W}(t), A)) \quad (2)$$

and

$$p_{C|A}(t | \gamma, A, |C|) = \frac{\exp(\gamma^\top h(\mathcal{W}(t), C, A))}{\sum_{W \in \mathcal{H}(t, |C|)} \exp(\gamma^\top h(\mathcal{W}(t), W, A))}, \quad (3)$$

defining the author and citation model, respectively, where

- $\theta \in \mathbb{R}^P$ and $\gamma \in \mathbb{R}^Q$ are parameter vectors estimated from data that govern the author and citation model, respectively (see Appendix D, for additional information on how they are estimated);
- $\lambda_0^A(|A|, t)$ is a nonparametric baseline intensity depending on the size of the author set and time t ;
- $s(\mathcal{W}(t), A) \in \mathbb{R}^P$ and $h(\mathcal{W}(t), C, A) \in \mathbb{R}^Q$ are vectors of statistics characterizing for the author and citation model, separately. These statistics detail how the intensity of observing author set A and citation set C given author set A at time $t \in \mathcal{T}$ are affected by the bibliographic data of the past, denoted by $\mathcal{W}(t)$. This means that observing a specific group of authors or cited works depends on the authors of previous works and who they cited.

5.2. Specification

To adapt this general framework to the theory at hand, we need to specify the vectors of statistics $s(\mathcal{W}(t), A)$ and $h(\mathcal{W}(t), C, A)$ to act as proxies for the hypotheses developed in Section 2 and control for other effects representing alternative explanations (or “control effects”) for citations or coauthorship relations (Lerner et al., 2024). The mathematical formulation of and further details on all effects employed in our model specification are provided in Appendix C. Most of the hypotheses are based on the citation side; however, we also consider the author model to investigate intercitation.

For H1, the accumulation of citations of works or authors may be the result of several simultaneous processes or mechanisms, which we capture by three effects illustrated in Fig. 4. We consider the first two effects as control variables for basic patterns in citation events. First, the “*Citation Popularity of Work*” (Fig. 4(a)) effect accounts for the effect of a paper’s popularity as a process of preferential attachment. This effect evaluates if more frequently cited publications are more likely to receive additional citations from the academic community. Second, “*Citation Repetition*” (Fig. 4(b)) approximates the Matthew effect as a ritual process, where the same researchers repeatedly cite the same papers, typically within specific scientific specialties, promoting the cognitive group. Finally, we use the “*Cite much Cited Authors*” (Fig. 4(c)) effect to capture the socio-cognitive structures we are interested in. Authors and works are interrelated, since researchers cite works of authors that have published highly cited other publications before.

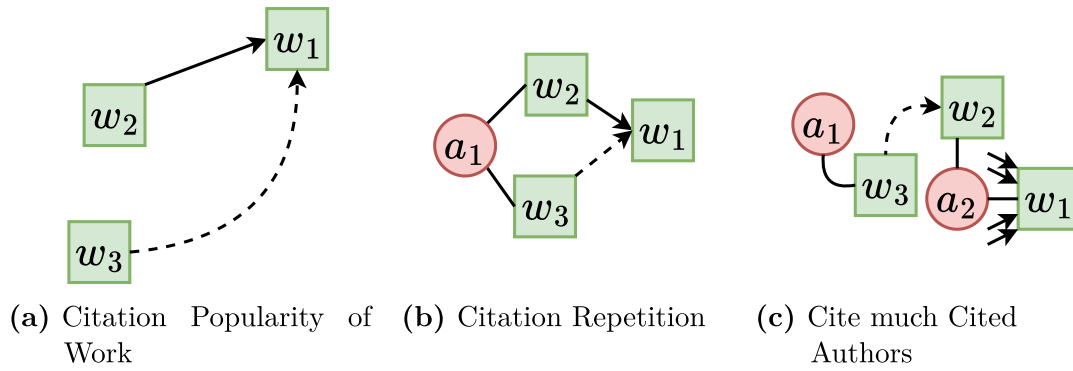


Fig. 4. Statistics included for H1.

Table 2

Operationalization of the main Hypotheses.

Hypothesis	Sub-hypothesis	Effects	Model
H1: Matthew effect of Authors	H1 (Fig. 4(c))	Cite much Cited Authors	Citation
H2: Intercitation	H2a (Fig. 5(a))	Author cites Author Repetition	Citation
	H2b (Fig. 5(c))	Collaborate with Citing Author	Author
	H2c (Fig. 5(d))	Cite Work of Coauthor	Citation
H3: Author Cocitation	H3 (Fig. 6(a))	Author cocitation	Citation
H4: Group effect	H4a (Fig. 7(a))	Adopt Citation of Coauthor	Citation
	H4b (Fig. 7(b))	Cocite Coauthor Pairs	Citation

The following effects approximate the intercitation mechanism. The first effect, “*Author cites Author Repetition*” (Fig. 5(a)), tests H2a and controls for researchers’ inclination to follow the work of prominent figures, often leaders in their specialties, by promoting their research agenda (Mullins and Mullins, 1973). The second effect, “*Author cites Author Reciprocation*” (Fig. 5(b)), also complements the first effect by incorporating reciprocity as a controlling effect, to evaluate if there is mutual admiration or recognition that may occur among researchers interested in similar topics. A third effect for H2b considers the author’s model by investigating whether scientists are more likely to coauthor papers with those who cited their previous work (“*Collaborate with Citing Author*” in Fig. 5(c)). Finally, for H2c, we employ the “*Cite Coauthor Works*” effect (Fig. 5(d)), which can be comprehended as an effect reversing the temporal order of “*Collaborate with Citing Author*” in that scientists are first coauthors and then cite each other.

As in H1, we use three effects to investigate H3. The first effect (Fig. 6(a)) represents the hypothesis related to author cocitation (H3), and the other two (Figs. 6(b) and 6(c)) control for simpler explanations for cocitation, which are lower-order terms. The effect of “*Author Cocitation*” (Fig. 6(a)) aims to identify the recurrence of authors perceived as working on similar topics by later publications. However, in the case of “*Cocitation Popularity (Pair)*” (Fig. 6(b)) and “*Cocitation Popularity (Triple)*” (Fig. 6(c)), instead of using author cocitation (White and Griffith, 1981), we control for standard cocitation at the level of works (Small, 1973). The main difference is that in the former case, we emphasize the duality of authors and works, while in the latter, the focus is on the cognitive dimension without considering which authors are behind the publication.

We use two variants specified in H4a and H4b to investigate the described group mechanisms. We test hypothesis H4a using “*Adopt Coauthor Citation*” (Fig. 7(a)), which is a “*social influence*” effect in which scientists cite some of the works cited by their former coauthors. To test H4b, we consider the statistic “*Cocite Coauthor Pair*”(Fig. 7(b)), which represents the pattern of cociting papers of a pair of coauthors.

Table 2 summarizes the main hypotheses and effects and to which model each effect corresponds. In the author model, we include additional control effects such as the ratio of Chileans in author teams compared to foreigners (“*Ratio Chileans*”) and the heterogeneity of the team of coauthors concerning Chilean nationality (“*Heterogeneity*

Chilean”). We also consider some degree-based effects such as “*Citation Popularity of Authors*”, “*Publication Activity*”, and “*Coauthor Repetition*” (for pairs, triples, and quartets of authors). Additionally, we explore other transitivity-based effects for social (“*Closure by Coauthor*”) and cognitive (“*Closure by Citing Same Work*”) structures. In the citation model, we account for further effects including “*Outdegree Popularity*”, which controls for the tendency to cite papers with long reference lists, “*Cite Work and its Citations*”, as an appropriation of knowledge from the baseline publication, and “*Self-Citation*”, as the effect of authors citing their own past work.

6. Results

Before presenting the model results, note that the coefficients θ and γ can be interpreted similarly to those in proportional hazard models. For $p \in 1, \dots, P$, consider two distinct author sets $A \subseteq \mathcal{A}$ and $A^* \subseteq \mathcal{A}$. Let θ_p and $s_p(W(t), A)$ denote the p th coefficient and statistic in (2). If $s_k(W(t), A) = s_k(W(t), A^*)$ for all $k \neq p$ and $s_p(W(t), A) = s_p(W(t), A^*) + 1$, then $\theta_p > 0$ indicates that A^* is more likely than A . The coefficient γ can be interpreted in a similar manner. Further details are provided in Appendix B.

6.1. Author model

The author model explains the set of authors $A(w)$ of the published work w , irrespective of the citations $C(w)$. Findings are given in the upper part of Table 3. We also assess for each included covariate the differences in the Akaike Information Criterion (AIC – negative values indicate improvement) over the null model and in the full model. To measure contributions over the null model, we compare the AIC of the model containing just the focal effect with the AIC of the model containing only a baseline intensity. To measure contributions in the full model, we compare the AIC of the model containing all effects (that is, all effects of the author model or all effects of the citation model described in Section C) with the AIC of the model containing all effects but the focal one. The upper part of Table 4 presents the differences in AIC values in the author model, ordered by their contributions over the null model.

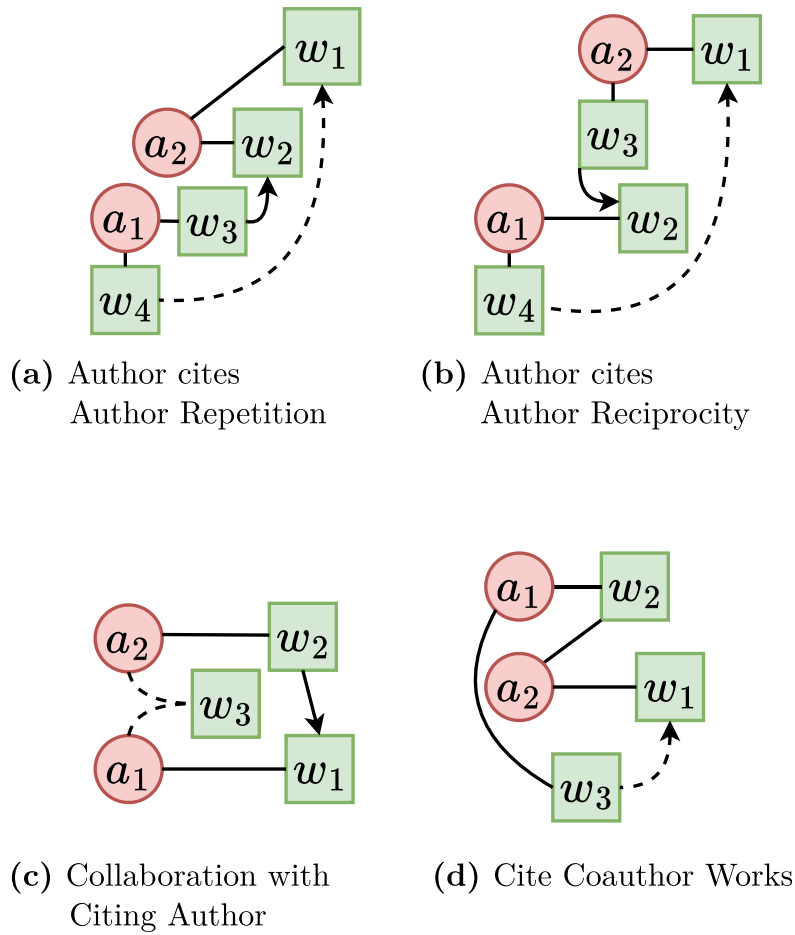


Fig. 5. Statistics included for H2.

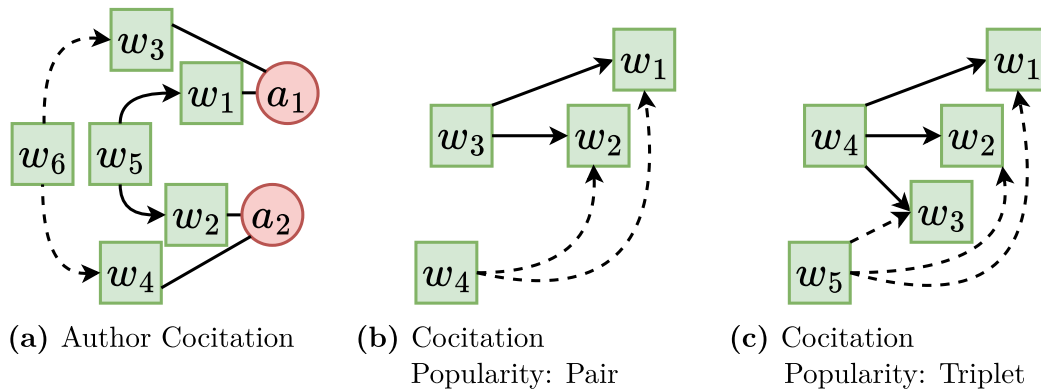


Fig. 6. Statistics included for H3.

Hypotheses effects. Regarding H2b, we find a positive effect to “Collaborate with Citing Author”. This suggests that scientists tend to coauthor works with those who cited their own work in the past, as predicted by H2b. That is, there is a social selection of coauthors having cited their own work. From another point of view, there is a cross-network effect in the sense that if author a_1 has cited the work of author a_2 , then it is more likely that a_1 and a_2 become coauthors in the future. The effect to collaborate with an author who has cited the own work makes a contribution of intermediate strength over the null model, compared to the contributions of the control effects. However, its contribution in the full model is much weaker, albeit it leads to a model improvement.

Other effects. Foreigners are more likely to be included in sets of authors than Chileans (negative effect of “Ratio Chilean”, significant at the 10% level), and groups of coauthors are more heterogeneous than groups of randomly sampled scientists. That is, teams of coauthors are often mixed with Chileans and foreigners (positive effect of “Heterogeneity Chilean”).

There is a positive effect of the number of citations that an author’s works have received in the past (“Citation Popularity”). That is, those scientists whose works have been cited more in the past publish at a higher rate in the future, i.e., they are more likely to be included as coauthors. There is a negative effect of “Publication Activity” on the publication rate. Which, scientists who have published more in the

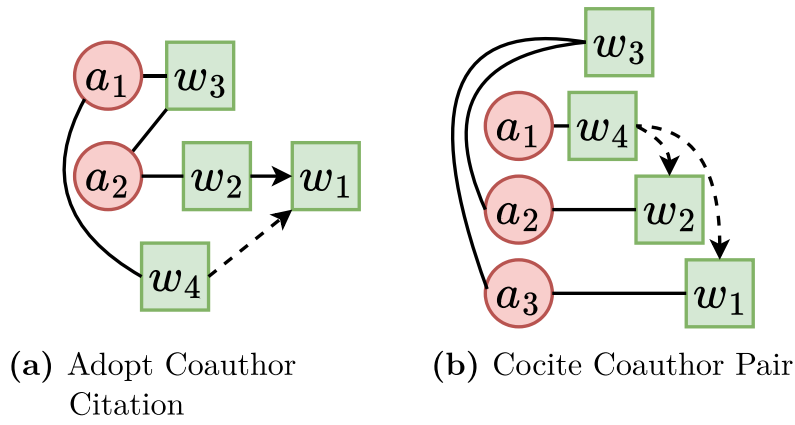


Fig. 7. Statistics included for H4.

Table 3

Results: The first column shows the estimated coefficients, the second column the standard errors, and the third column the p-values. Note that H1 stands for Hypothesis 1, H2 for Hypothesis 2 and so on.

	Estimate	Std. error	p-Value
Author model			
Ratio Chileans	−0.166	0.074	0.026
Heterogeneity Chilean	0.791	0.168	<0.001
Citation Popularity of Author	−0.320	0.074	<0.001
Publication Activity	0.506	0.049	<0.001
Coauthor-pair Repetition	1.157	0.043	<0.001
Coauthor-triple Repetition	0.224	0.011	<0.001
Coauthor-quartet Repetition	0.070	0.009	<0.001
H2b: Collaborate with Citing Author	0.114	0.046	0.013
Closure by Coauthor	−0.861	0.082	<0.001
Closure by Citing same Work	−0.115	0.029	<0.001
Citation model			
Citation Popularity of Work	−0.153	0.140	0.274
Cocitation Popularity: Pair	0.265	0.029	<0.001
Cocitation Popularity: Triple	0.073	0.005	<0.001
Citation Repetition	0.151	0.047	0.001
Outdegree Popularity	−0.620	0.103	<0.001
Cite Work and its Citations	0.297	0.021	<0.001
Self Citation	1.531	0.115	<0.001
H1: Cite much Cited Authors	−1.388	0.210	<0.001
H2a: Author cites Author Repetition	0.730	0.209	<0.001
Author cites Author Reciprocation	−0.112	0.206	0.585
H2c: Cite Work of Coauthor	−0.384	0.138	0.006
H3: Author Cocitation	−0.165	0.109	0.130
H4a: Adopt Citation of Coauthor	−0.063	0.079	0.422
H4b: Cocite Coauthor Pairs	0.480	0.110	<0.001

past will publish at a lower rate in the future. This latter effect works towards equalizing publication counts in the population of scientists.

There is evidence for “*Coauthorship Repetition*” among groups of scientists of sizes two, three, and four. That is, those groups of the given sizes who have coauthored more in the past are more likely to coauthor work in the future.

There is a significant negative triadic closure (“*Closure by Coauthor*”) in the coauthoring network. According to Lerner and Lomi (2022), Lerner and Hâncean (2023), negative closure points to actors occupying stable broker positions (actors surrounded by structural holes, that is, actors bridging between communities). A positive closure effect would imply a tendency to close structural holes. Still, the negative closure effect found in our model suggests that authors are likely to keep structural holes open so that the “third author”, i.e., the one connected to the two others, is likely to keep her broker position. Likewise, if two scientists cited the same works, they are less likely to jointly publish a work (“*Closure by Citing Same Work*”).

Table 4

AIC Contributions: Over null models and in full models. More negative values point to stronger contributions. Effects are ordered by the contributions over the null models. The percentages in brackets give the share of the AIC differences between the respective full models and null models. In the Author Model, the AIC of the full model minus the AIC of the null model is −2629.438, corresponding to 100%. In the Citation Model, the AIC of the full model minus the AIC of the null model is −13914.82, corresponding to 100%.

Model term	Over null model	In full model
Author model		
Coauthor-pair Repetition	−1507.248 (57.3%)	−593.090 (22.6%)
Closure by Citing same Work	−997.218 (37.9%)	−12.278 (0.5%)
Coauthor-triple Repetition	−825.247 (31.4%)	−263.097 (10.0%)
Publication Activity	−763.299 (29.0%)	−101.118 (3.8%)
H2b: Collaborate with Citing Author	−760.261 (28.9%)	−3.890 (0.1%)
Closure by Coauthor	−593.533 (22.6%)	−200.996 (7.6%)
Citation Popularity of Author	−472.499 (18.0%)	−20.720 (0.8%)
Coauthor-quartet Repetition	−371.534 (14.1%)	−81.990 (3.1%)
Ratio Chileans	−110.515 (4.2%)	−3.217 (0.1%)
Heterogeneity Chilean	0.198 (−0.0%)	−20.708 (0.8%)
Citation model		
Self citation	−11 657.594 (83.8%)	−1340.308 (9.6%)
Citation Repetition	−6920.160 (49.7%)	−21.127 (0.2%)
Cocitation Popularity: Triple	−6315.904 (45.4%)	−249.832 (1.8%)
Cocitation Popularity: Pair	−4813.124 (34.6%)	−187.990 (1.4%)
Cite work and its Citations	−4785.378 (34.4%)	−360.689 (2.6%)
H2a: Author cites Author Repetition	−2967.863 (21.3%)	−30.325 (0.2%)
Author cites Author Reciprocation	−2637.315 (19.0%)	1.258 (−0.0%)
H4a: Adopt Citation of Coauthor	−1235.190 (8.9%)	0.847 (−0.0%)
H2c: Cite work of Coauthor	−568.488 (4.1%)	−17.708 (0.1%)
Citation Popularity of Work	−525.869 (3.8%)	−1.731 (0.0%)
H4b: Cocite Coauthor Pairs	−172.407 (1.2%)	−35.353 (0.3%)
H3: Author Cocitation	−16.773 (0.1%)	−1.999 (0.0%)
H1: Cite much Cited Authors	0.205 (−0.0%)	−102.173 (0.7%)
Outdegree Popularity	0.942 (−0.0%)	−78.691 (0.6%)

6.2. Citation model

The citation model explains the list of citations $C(w)$ of a published work w , conditioning on its group of authors $A(w)$. Findings are given in the lower part of Table 3. The lower part of Table 4 presents the differences in AIC values implied by the various effects, ordered by their contributions over the null model.

Hypotheses effects. Hypothesis H1 is represented by the effect “*Cite Much Cited Authors*”, displayed in Fig. 4(c). In the example given in that figure, a_2 authored the work w_1 , which received many citations in the past, and a_2 also authored work w_2 . H1 predicts that there is an increase probability of w_2 being cited in the future. Put differently, we expect a “*spill-over*” effect of the popularity of w_1 onto the work w_2 written by the same author a_2 . Contrary to these expectations, we find a negative effect of “*Cite Much Cited Authors*”. In the setting

above, this suggests that w_2 gets cited at a lower rate in the future. Still, this finding must be interpreted alongside several other effects that control for repeatedly citing the same works or the same authors: We find a positive tendency for “Citation Repetition”, displayed in Fig. 4(b), suggesting that if author a_1 has already cited work w_1 , then the same author a_1 is more likely to cite w_1 again when publishing another paper. The second control effect, “Citation Popularity of Work”, displayed in Fig. 4(a), has a negative parameter (significant at the 10% level). This suggests that works that received many citations in the past get cited at a lower rate in the future, controlling for all other effects. The contribution of the effect “Cite Much Cited Authors” to the AIC is of moderate strength in the full model – compared to the contributions of other effects – but its contribution over the null model is non-existent. In general, we observe that – especially in the references model but also in the authors model – several of the control variables make stronger contributions than any of the effects related with our hypotheses. This does not come as a surprise and, when interpreting results, we have to distinguish between “explaining the data” and “testing relevant hypotheses”. Indeed, for example, the tendency of authors to cite their own work (“Self citation”) or to repeatedly cite the same papers (“Citation repetition”) are fairly obvious patterns that empirically explain a large share of the variance in the data; on the other hand there is hardly any novel or unexpected insight coming from these patterns.

Hypothesis H2a is represented by “Author Cites Author Repetition” (Fig. 5(a)) and we also control for “Author Cites Author Reciprocation” (Fig. 5(b)). Consistent with H2a, we find a positive effect of repetition, implying that if author a_1 has cited work of a_2 , then a_1 is more likely to cite a (possibly different) work of a_2 in the future. There is no significant finding for reciprocation. The latter effect predicts that if author a_1 has cited work of a_2 , then a_2 is more likely to cite work of a_1 in the future; thus, it reverses the roles of the citing and cited authors. Looking at the contributions to the AIC of these effects we find that both make an intermediate contribution over the null model and “Author Cites Author Repetition” makes a small but positive contribution in the full model, while “Author Cites Author Reciprocation” entails no improvement in the full model (consistent with its statistical insignificance).

Hypothesis H2c is represented in Fig. 5(d) “Cite Work of Coauthor”. In the example given in this figure, a_1 and a_2 have coauthored work w_2 and a_2 has published w_1 . Given this precondition, H2c predicts that a_1 is more likely to cite the work w_1 of her coauthor in the future. Contrary to the predictions of H2c, we find a negative effect to cite the work of former coauthors. We recall that H2b, “Collaborate with Citing Author”, has been (positively) tested with the author model, discussed above. The contribution to the AIC of “Cite Work of Coauthor” entails a small to intermediate improvement over the null model and a small but positive improvement in the full model.

Hypothesis H3 is represented by “Author Cocitation”, displayed in Fig. 6(a). This effect predicts that if (possibly different) works of authors a_2 and a_3 jointly appear in the reference list of a past work w_5 , then a future work w_6 is more likely to cocite (possibly yet other and possibly different) works of a_2 and a_3 . With the effects displayed in Figs. 6(b) and 6(c) – cocitation popularity of pairs and triples of works – we control for the baseline effect to repeatedly cocite the same (pairs or triples of) works, rather than to cocite different works of the repeated pair of authors. Contrary to the predictions of H3, we find a negative effect (significant at the 10% level) to repeatedly cocite authors. In contrast, repeated cocitation to pairs and triples of papers is significantly positive. The contributions to the AIC over the null model or in the full model of “Author Cocitation” are almost inexistent, while the contributions of the cocitation popularity of papers are very strong.

Hypothesis H4a is represented in Fig. 7(a) “Adopt Citations of Coauthors” and H4b is represented in Fig. 7(b) “Cocite Coauthor Pairs”. In the example given in these figures, “Adopt Citations of Coauthors” has as precondition that a_1 and a_2 are coauthors (having coauthored work w_3) and a_2 has cited w_1 when publishing w_2 . Given this precondition,

a_1 is predicted to be more likely to also cite work w_1 when publishing another future work w_4 . However, we find no significant effect in adopting citations of coauthors, contrary to the predictions of H4a. The Effect “Cocite Coauthor Pairs” has as a precondition that a_2 and a_3 have coauthored work w_3 and have individually published w_2 and w_1 . Given this precondition, the effect predicts that w_1 and w_2 have an increased probability to be cocited by the future work w_4 . Consistent with the predictions of H4b we find a positive effect to cocite works individually published by former pairs of coauthors. Consistent with the significance or insignificance of their parameters “Adopt Citations of Coauthors” entails no improvement to the AIC in the full model while “Cocite Coauthor Pairs” entails a small improvement. Both effects make small to moderate contributions over the null model.

Other effects. We find a positive tendency to “Cite Work and its Citations”. There are different possible explanations of this effect, as discussed in Lerner et al. (2024). Among others, it may be that authors copy parts of the reference lists of works they cite; that authors search for citations to a work they cite (and subsequently cite some of the citing works); or that work w_2 is topically similar to the work w_1 it cites, increasing the probability that w_2 and w_1 get cocited in the future. Works with longer reference lists – Effect “Outdegree Popularity” – are less likely to be cited. Moreover, authors frequently cite their own works (“Self Citation”).

With respect to the contribution of the other effects over the null model and in the full model (Table 4), we have already noted above that some controlling effects are more prominent in both cases. This is the case of coauthor-pair repetition, coauthor-triple repetition, and publication activity for the author model, and self citation (the most noticeable effect in the analysis), cocitation popularity triple, and cocitation popularity pair for the citation model. Regarding our main hypotheses, in both cases, some inter citations effects (for H2a: author cites author repetition and H2c: cite work of coauthor) are consistently contributing over the null model and in the full model. For the case of the Matthew effect of authors (H1), when we consider its contribution to the full model, we observe that the effect is more prominent than the other hypotheses (but relatively low in the null model) followed by the group effect to cocite coauthor pairs (H4b). The importance of these effects in the full model reveals the relevance of considering collaboration and citation events embedded in higher-order relations, that is, embedded in a hypergraph. Finally, the contribution of the author’s cocitation (H3) effect in both cases is fairly modest in comparison with the other effects.

7. Discussion

Using bibliometric data to analyze a scientific community of astronomers in Chile, we investigated social and cognitive ties. By defining the context, we explored the social dimension – often linked to invisible colleges (Crane, 1972; Zuccala, 2006) but also to underlying knowledge dissemination (Carley, 1986; Collins, 2002) – as a dual process involving both authors and their works (Bellotti and Espinosa-Rada, 2025). This approach follows the tradition of “duality” (Breiger, 1974; Mützel and Breiger, 2020), explicitly considering science’s structural and cultural forms as socio-cognitive networks. Examining these processes and the micro-temporal mechanisms at play allowed us to identify patterns that influence authors’ citation behaviors. While citation is often used to assess researchers’ impact, it also plays a broader role in shaping the stratification of science.

While the literature consistently states that social stratification exists in science (de Solla Price, 1965; Cole and Cole, 1973; Newman, 2001; Barabási et al., 2002), our results indicate that authors do not preferentially cite those who have received more citations in the past. These findings highlight the benefits of our approach in not projecting the data, as they show that although citation repetition occurs between, it is more predominant at the level of works rather than authors. Although stratification may be more apparent in larger networks, authors

in this bounded community do not accumulate recognition and are cited at a lower rate over time. Similarly, works that have received many citations in the past tend to be cited less frequently in the future. Nonetheless, the accumulation of recognition through citation is linked to previously cited works. Thus, we interpret these findings as evidence of a preference for contemporaneous work among Chilean astronomers in scientific knowledge.

Regarding the preference for recent research among Chilean astronomers, many important control effects do not take the cited authors into account, as shown in Table 4. This phenomenon suggests that subject specialization, sets of beliefs and bodies of knowledge, is a key dimension underlying this community. One instance of this is citation repetition, where certain papers function as accumulative processes in which future work builds on previous research. Moreover, we obtain positive coefficients for the cocitation popularity of pairs and triples, reflecting cognitive closure processes at the level of scientific knowledge (i.e., works). Some authors, from an intercitation perspective, are still cited repetitively and collaborate with citing authors, but the coefficient for citing highly cited authors is negative. We interpret the relevance of intrinsic ideas, regardless of authorship, as an indication of a tendency towards specialization.

For socio-cognitive ties, White (2011) argues that the “true glue” binding scientists and scholars together is what they can competently write about rather than whom they know. Recent research (Espinosa-Rada et al., 2024) has challenged this claim by reaffirming the relevance of the social dimension in the study of socio-cognitive networks. However, our findings differ from those of Espinosa-Rada et al. (2024). While their study found that prior collaboration leads to citation but prior citation does not lead to collaboration, we observe the opposite effect. Similarly, Lerner et al. (2024) found that scientists tend to coauthor with those who have cited their work and are inclined to cite their coauthors’ papers.

Our results indicate that authors are more likely to collaborate with individuals who have previously cited their work. Conversely, they tend to cite the work of former coauthors less frequently in subsequent research compared to non-collaborators. Some of these patterns may be linked to implicit local hierarchy practices. Controlling for other effects, we find that scientists tend to cite their former coauthors or highly cited authors less frequently than others. However, these results must be interpreted in light of a strong positive effect of citation repetition, which reflects a lower-order representation of the Matthew effect.

We conjecture that unobserved status-related processes – such as academic seniority or forms of reputation not captured by citation counts – may be shaping citation dynamics. The divergence of our findings from previous expectations highlights the need for further investigation.

The prominence of ideas is also observed in the case of cocitation, which is considered a key driver of why researchers are cited. Classical research (Small and Griffith, 1974) established that cocitation identifies relationships between works deemed important by authors within a specialty. White and Griffith (1981) extended this concept to author cocitation, showing how authors whose works are generally regarded as related tend to cluster in knowledge maps. Our results indicate that authors frequently cited together do not tend to be cited more often in future publications. A similar effect is observed when considering repeated cocitation at the level of works (pairs and triples) rather than authors. Both effects make a strong contribution to the author and citation model shown in Table 4. These findings suggest that in the local astronomical community, cocited authors tend to be less prominent than cocited works in shaping the development of scientific specialties.

We also highlight the importance of exploring the role of ‘coherent groups’—groups that tend to expand over time. These groups have historically been considered crucial for explaining how scientific specialties evolve (Mullins and Mullins, 1973). More recently, research on scientific networks suggests that science is becoming increasingly team-oriented (Wuchty et al., 2007; Jones et al., 2008; Leahey, 2016).

We build on this perspective by explicitly examining the effect of these groups on citation patterns, focusing on dualities in the interplay of works and authors. This is particularly relevant, as some researchers have argued that, in such communities, a researcher’s group of colleagues is the most significant source of social influence on their work (Hagstrom, 1965). Our findings indicate that collaborating authors do not tend to cite similar references in their publications. However, authors who collaborate are more likely to be cocited in subsequent publications. This effect suggests that the astronomical community is more inclined to cite dyadic teams, as our analysis did not account for triadic or higher-order structures.

From a methodological perspective, we contribute to the study of scientific networks by modeling citation dynamics without aggregating data at the author or paper level—a common practice in empirical studies and classical two-mode projections, such as those proposed by Breiger (1974). Instead, we model raw group-to-set citation events directly. This approach enhances the capacity of group-oriented relational event models.

We demonstrate the utility of this method through an empirical case study and further extend the model by incorporating novel structural effects that capture dependencies among overlapping sets of nodes and ties. Our framework allows for a more granular analysis of citation mechanisms, enabling the investigation of dualities, such as the interplay between authors and works through (co)authorships or citation networks. The Group-Oriented Relational Hyperevent Model provides a way to model hyperevents in situations where groups are first formed and then decide on subsequent actions. This perspective allows for testing effects that consider group decisions on fine-grained temporal micro-mechanisms. Additionally, the model distinguishes between selection effects – where authors preferentially cite based on prior relational structures – and influence effects, in which network exposure shapes future citation behavior. In our case, we identify whether coherent author groups influence one another through shared citation practices (e.g., adopting the citation patterns of coauthors) or whether they extend their influence indirectly, such as by being cocited.

While we have advanced the knowledge about socio-cognitive networks among researchers, several limitations need to be addressed. Our study’s time frame was restrictive and only covered a short period (2013–2015). Further applications should incorporate extended periods. However, increasing the time window requires careful consideration of how reasonable the assumption is that authors are aware of all previous publications necessary to identify inter citations. Additionally, our study is only a case study. We believe that comparisons between disciplines and including interdisciplinary research areas can enhance the exploration of complex networks. Topics are a relevant area of research that should further explore the relevance of specializations. There are many possibilities for exploring this dimension. For instance, one can independently run a Large Language Model on the textual data to derive embeddings for the papers and then use this in our framework as covariates for the respective papers. Alternatively, one may derive keywords from past papers to encode the areas of expertise of each author who contributed to the paper. One can use this information as exogenous covariates, such as the nationality of the authors (variable whether the authors are Chilean or not). Finally, moving beyond bibliometric data can also help explore the social dimension underpinning the scientific network. Further research should go beyond formal communication channels in science.

CRediT authorship contribution statement

Alejandro Espinosa-Rada: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jürgen Lerner:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation,

Formal analysis, Data curation, Conceptualization. **Cornelius Fritz:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Financial support

The authors acknowledge support from Fondecyt Regular (ANID) under grant number 1220560 (AER) and the German Research Foundation, Germany under grants 321869138 (JL) and 520770522 (CF).

Acknowledgments

We thank members of the Social Networks Lab at ETH Zürich for helpful comments and suggestions that considerably improved this work.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.socnet.2025.04.005>.

Data availability

For legal reasons, data from Clarivate Web of Science cannot be made openly available.

References

- Barabási, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., Vicsek, T., 2002. Evolution of the social network of scientific collaborations. *Phys. A* 311 (3), 590–614.
- Batagelj, V., 2020. On fractional approach to analysis of linked networks. *Scientometrics* 123 (2), 621–633.
- Batagelj, V., Cerinšek, M., 2013. On bibliographic networks. *Scientometrics* 96 (3), 845–864.
- Bellotti, E., Espinosa-Rada, A., 2025. Scientific networks. In: Crossley, N., Widdop, P. (Eds.), *Handbook of Social Networks and Culture*. Edward Elgar, pp. 154–167.
- Breiger, R.L., 1974. The duality of persons and groups. *Soc. Forces* 53 (2), 181–190.
- Breiger, R.L., 1976. Career attributes and network structure: A blockmodel study of a biomedical research specialty. *Am. Sociol. Rev.* 117–135.
- Burgdorf, K., Wittek, M., Lerner, J., 2024. Communities of style: Artistic transformation and social cohesion in hollywood, 1930 to 1999. *Socius* 10, 23780231241257334.
- Butts, C.T., Lomi, A., Snijders, T.A.B., Stadtfeld, C., 2023. Relational event models in network science. *Netw. Sci.* 11 (2), 175–183.
- Carley, K., 1986. An approach for relating social structure to cognitive structure. *J. Math. Sociol.* 12 (2), 137–189.
- Chubin, D.E., 1976. The conceptualization of scientific specialties. *Sociol. Q.* 17 (4), 448–476.
- Chubin, D.E., Studer, K.E., 1979. Knowledge and structures of scientific growth measurement of a cancer problem domain. *Scientometrics* 1 (2), 171–193.
- Cole, J.R., Cole, S., 1973. *Social Stratification in Science*. University of Chicago Press.
- Collins, R., 2002. *The Sociology of Philosophies: A Global Theory of Intellectual Change*. Belknap Press of Harvard University Press.
- Crane, D., 1969. Social structure in a group of scientists: A test of the 201CInvisible College201DHypothesis. *Am. Sociol. Rev.* 34, 335–352.
- Crane, D., 1972. *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. University of Chicago Press, Chicago.
- de Solla Price, D.J., 1965. Networks of scientific papers. *Science* 149 (3683), 510–515.
- de Solla Price, D.J., Beaver, D., 1966. Collaboration in an invisible college. *Am. Psychol.* 21, 1011–1018.
- Edge, D., 1979. Quantitative measures of communication in science: A critical review. *Hist. Sci.* 17 (2), 102–134.
- Espinosa-Rada, A., 2021. *A Network Approach for the Sociological Study of Science and Knowledge: Modelling a Dynamic Multilevel Network* (Ph.D. thesis). The University of Manchester.
- Espinosa-Rada, A., Bellotti, E., Everett, M., Stadtfeld, C., 2024. Co-evolution of a socio-cognitive scientific network: A case study of citation dynamics among astronomers. *Soc. Netw.* 78, 92–108.
- Espinosa-Rada, A., Ortiz, F., 2022. Gender and researchers with institutional affiliations in the global south/north in social network science. *Appl. Netw. Sci.* 7 (1), 40.
- Fararo, T.J., Doreian, P., 1984. Tripartite structural analysis: Generalizing the Breiger-Wilson formalism. *Soc. Netw.* 6 (2), 141–175.
- 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.
- Fritz, C., De Nicola, G., Kevork, S., Harhoff, D., Kauermann, G., 2023. Modelling the large and dynamically growing bipartite network of German patents and inventors. *J. R. Stat. Soc. Ser. A: Stat. Soc.* 186 (3), 557–576.
- Gallagher, C., Lusher, D., Koskinen, J., Roden, B., Wang, P., Gosling, A., Polyzos, A., Stenzel, M., Hegarty, S., Spurling, T., Simpson, G., 2023. Network patterns of university-industry collaboration: A case study of the chemical sciences in Australia. *Scientometrics* 128 (8), 4559–4588.
- Garfield, E., 1964. *Citation Data in Writing the History of Science*. Philadelphia, Pennsylvania, USA.
- Griffith, B.C., Mullins, N.C., 1972. Coherent social groups in scientific change. *Science* 177 (4053), 959–964.
- Hagstrom, W.O., 1965. *The Scientific Community*. basic books, New York.
- Holland, P.W., Leinhardt, S., 1974. Measurement error in sociometry. In: Blalock, H. (Ed.), *Measurement in the Social Science: Theories and Strategy*, University of Washington Washington, D.C, pp. 187–211.
- Hummon, N.P., Doreian, P., 1989. Connectivity in a citation network: The development of DNA theory. *Soc. Netw.* 11 (1), 39–63.
- Jones, B.F., Wuchty, S., Uzzi, B., 2008. Multi-university research teams: Shifting impact, geography, and stratification in science. *Science* 322 (5905), 1259–1262.
- Kadushin, C., 1966. The friends and supporters of psychotherapy: On social circles in urban life. *Am. Sociol. Rev.* 31 (6), 786–802.
- Kronegger, L., Ferligoj, A., Doreian, P., 2011. On the dynamics of national scientific systems. *Qual. Quant.* 45 (5), 989–1015.
- Kuhn, T.S., 2012. *The Structure of Scientific Revolutions*. University of Chicago Press.
- Leahey, E., 2016. From sole investigator to team scientist: Trends in the practice and study of research collaboration. *Annu. Rev. Sociol.* 42, 81–100.
- Lerner, J., Hâncian, M.-G., 2023. Micro-level network dynamics of scientific collaboration and impact: relational hyperevent models for the analysis of coauthor networks. *Netw. Sci.* 11 (1), 5–35.
- Lerner, J., Hâncian, M.-G., Lomi, A., 2024. Relational hyperevent models for the coevolution of coauthoring and citation networks. *J. R. Stat. Soc. Ser. A: Stat. Soc.* qnae068.
- Lerner, J., Lomi, A., 2022. A dynamic model for the mutual constitution of individuals and events. *J. Complex Netw.* 10 (2), cnac004.
- Lerner, J., Lomi, A., 2023. Relational hyperevent models for polyadic interaction networks. *J. R. Stat. Soc. Ser. A: Stat. Soc.* 186 (3), 577–600.
- Lerner, J., Lomi, A., Mowbray, J., Rollings, N., Tranmer, M., 2021. Dynamic network analysis of contact diaries. *Soc. Netw.* 66, 224–236.
- Leydesdorff, L., 1998. Theories of citation? *Scientometrics* 43, 5–25.
- Leydesdorff, L., Vaughan, L., 2006. Co-occurrence matrices and their applications in information science: Extending ACA to the web environment. *J. Am. Soc. Inf. Sci. Technol.* 57 (12), 1616–1628.
- McFadden, D., 1973. Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P. (Ed.), *Frontiers in Econometrics*. Academic Press New York, New York, NY, USA, pp. 105–142.
- Merton, R.K., 1968. The matthew effect in science: The reward and communication systems of science are considered. *Science* 159 (3810), 56–63.
- Merton, R.K., 1988. The Matthew Effect in science, II: Cumulative advantage and the symbolism of intellectual property. *Isis* 79 (4), 606–623.
- Merton, R.K., 2000. On the Garfield input to the sociology of science: A retrospective collage. In: Cronin, B., Atkins, H.B. (Eds.), *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*. Information Today, Inc., Medford, New Jersey, pp. 435–448.
- Moody, J., 2004. The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *Am. Sociol. Rev.* 69 (2), 213–238.
- Moody, J., Light, R., 2006. A view from above: The evolving sociological landscape. *Am. Sociol.* 37 (2), 67–86.
- Mullins, N.C., 1972. The development of a scientific specialty: The phage group and the origins of molecular biology. *Minerva* 10, 51–82.
- Mullins, N.C., 1973. The development of specialties in social science: The case of ethnomethodology. *Sci. Stud.* 3 (3), 245–273.
- Mullins, N.C., Hargens, L.L., Hecht, P.K., Kick, E.L., 1977. The group structure of cocitation clusters: A comparative study. *Am. Sociol. Rev.* 552–562.
- Mullins, N.C., Mullins, C.J., 1973. *Theories and Theory Groups in Contemporary American Sociology*. Harper & Row, New York.
- Mützel, S., Breiger, R.L., 2020. Duality beyond persons and groups. In: Light, R., Moody, J. (Eds.), *The Oxford Handbook of Social Networks*. Oxford University Press, pp. 392–413.
- Newman, M.E.J., 2001. Clustering and preferential attachment in growing networks. *Phys. Rev. E* 64 (2), 025102.
- Nicolaisen, J., 2007. Citation analysis. *Annu. Rev. Inf. Sci. Technol.* 41 (1), 609–641.
- Rivera, M.T., Soderstrom, S.B., Uzzi, B., 2010. Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *Annu. Rev. Sociol.* 36, 91–115.
- Schrum, W., Mullins, N.C., 1988. Network analysis in the study of science and technology. In: van Raan, A.F.J. (Ed.), *Handbook of Quantitative Studies of Science and Technology*. Elsevier Science Publishers B.V. (North-Holland), pp. 107–133.

- Small, H., 1973. Co-citation in the scientific literature: A new measure of the relationship between two documents. *J. Am. Soc. Inf. Sci.* 24 (4), 265–269.
- Small, H.G., 1978. Cited documents as concept symbols. *Soc. Stud. Sci.* 8 (3), 327–340.
- Small, H., Griffith, B.C., 1974. The structure of scientific literatures I: Identifying and graphing specialties. *Sci. Stud.* 4 (1), 17–40.
- Snijders, T.A.B., 2001. The statistical evaluation of social network dynamics. *Sociol. Methodol.* 31 (1), 361–395.
- Stadtfeld, C., Amati, V., 2021. Network mechanisms and network models. In: Gianluca, M. (Ed.), *Research Handbook on Analytical Sociology*. Edward Elgar Publishing, pp. 432–452.
- Stadtfeld, C., Block, P., 2017. Interactions, actors, and time: Dynamic network actor models for relational events. *Sociol. Sci.* 4, 318–352.
- Stadtfeld, C., Hollway, J., Block, P., 2017. Dynamic network actor models: Investigating coordination ties through time. *Sociol. Methodol.* 47 (1), 1–40.
- Stark, T.H., Rambaran, J.A., McFarland, D.A., 2020. The meeting of minds: Forging social and intellectual networks within universities. *Sociol. Sci.* 7, 433–464.
- Wang, F., Qiu, J., Yu, H., 2012. Research on the cross-citation relationship of core authors in scientometrics. *Scientometrics* 91 (3), 1011–1033.
- White, H.D., 2011. Scientific and Scholarly Networks. In: Scott, J., Carrington, P. (Eds.), *The SAGE Handbook of Social Network Analysis*. Sage, pp. 271–285.
- White, H.D., Griffith, B.C., 1981. Author cocitation: A literature measure of intellectual structure. *J. Am. Soc. Inf. Sci.* 32 (3), 163–171.
- White, H.D., Wellman, B., Nazer, N., 2004. Does citation reflect social structure?: Longitudinal evidence from the “Globenet” interdisciplinary research group. *J. Am. Soc. Inf. Sci. Technol.* 55 (2), 111–126.
- Witteck, M., Bartenhagen, C., Berthold, F., 2023. The development of stratification and segregation in a new scientific field: A study of collaboration among scientists in neuroblastoma research between 1975 and 2016. *Soc. Netw.* 72, 80–107.
- Wuchty, S., Jones, B.F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. *Science* 316 (5827), 1036–1039.
- Zhao, D., Strotmann, A., 2008. Author bibliographic coupling: Another approach to citation-based author knowledge network analysis. *Proc. Am. Soc. Inf. Sci. Technol.* 45 (1), 1–10.
- Zuccala, A., 2006. Modeling the invisible college. *J. Am. Soc. Inf. Sci. Technol.* 57 (2), 152–168.
- Zuckerman, H., 1967. Nobel laureates in science: Patterns of productivity, collaboration, and authorship. *Am. Sociol. Rev.* 391–403.