Reaching for Unique Resources: Structural Holes and Specialization in Scientific Collaboration Networks

Michał Bojanowski[a] and Dominika Czerniawska[b]

Abstract

On some fundamental level, we can think of scholars as actors possessing, or controlling, various types of resources. Collaboration in science is understood here as a process of pooling and exchanging such resources. We show how diversity of resources engaged in scientific collaboration is related to the structure of collaboration networks. We demonstrate that scholars within their personal networks simultaneously (1) diversify resources in collaboration ties surrounded by structural holes and (2) specialize resources in collaboration ties embedded in dense collaboration groups. These complementary mechanisms decrease individual efforts required to maintain effective collaborations in complex social settings. To this end, we develop a concept of "pairwise redundancy" capturing structural redundancy of ego's neighbors *vis-à-vis* each other.

_____

a. Corresponding author: Kozminski University, Warsaw, Poland
b. University of Manchester, Manchester, UK and University of Warsaw, Warsaw, Poland

1. Introduction

Scientists form collaboration ties with others because, among other things, thanks to pooling resources they can jointly benefit from rewards associated with a created outcome, e.g. a scientific publication. We can think of this kind of collaboration as a "co-production" of an outcome. Resources needed to get ahead in science are unequally distributed across the scientific community. For example, scientists at one laboratory specialize in field work and have collected research samples while scientists at some other laboratory might have access to sophisticated equipment that is needed to analyze these samples. Such unequal distribution of resources creates extra incentives to form collaborations and can be linked to decreasing popularity of individual, as opposed to collective, creation of scientific outcomes in contemporary science. An independent mode of work has become less effective for many scientists including those working in disciplines, which were traditionally more individualistic (Moody 2004). The process of formation of collaborative relations involves the matching of resources required for "creating an outcome." It has many features of a market-like mechanism. To attract desired resources controlled by others, a scientist himself has to offer resources desired by potential collaborators. Simultaneously, scientists face constraints (such as limited time), so there exists some form of competition for access to more desirable resources and more attractive collaborators. We propose a novel approach explaining diversity of resources conveyed in collaboration ties based on complementary mechanisms of structural holes and specialization.

There are a variety of resources relevant for doing science. We can roughly divide them into two categories: (1) resources which are directly engaged in collaboration such as expertise in a particular research topic or a research method, and (2) resources of a more "social" kind, e.g. contacts in academia or prestige. Actors control sets of resources and decide whether to engage them in collaborations according to the demands of collaborators, the desirability of rewards, and time and energy they have. To understand how an actor might engage resources in collaborations let us consider Figure 1 with a simplistic graph of four collaborating scientists. Five types of ties connecting the scientists correspond to different kinds of resources they contribute when collaborating with others. For example, Scientists A and B collaborate and in this collaboration, B conceptualized the research idea while A performed data analysis. At the same time, Scientist A contributes different resources to his other collaborations with D and C. Actors can contribute multiple types of resources in the same dyad, as is the case in pairs A-D or A-C – A contributes a single resource while D and C contribute two types. We may say that Scientist *A* contributes different *bundles* of resources to B and C. In its entirety, it is a multiplex network or a *multigraph* (Wasserman and Faust 1994:Ch. 4.6). The interdependencies between different types of ties (here resources) can be quite complex and resulting from, among other things, different availability of, and demands for, different resources among the scientists.
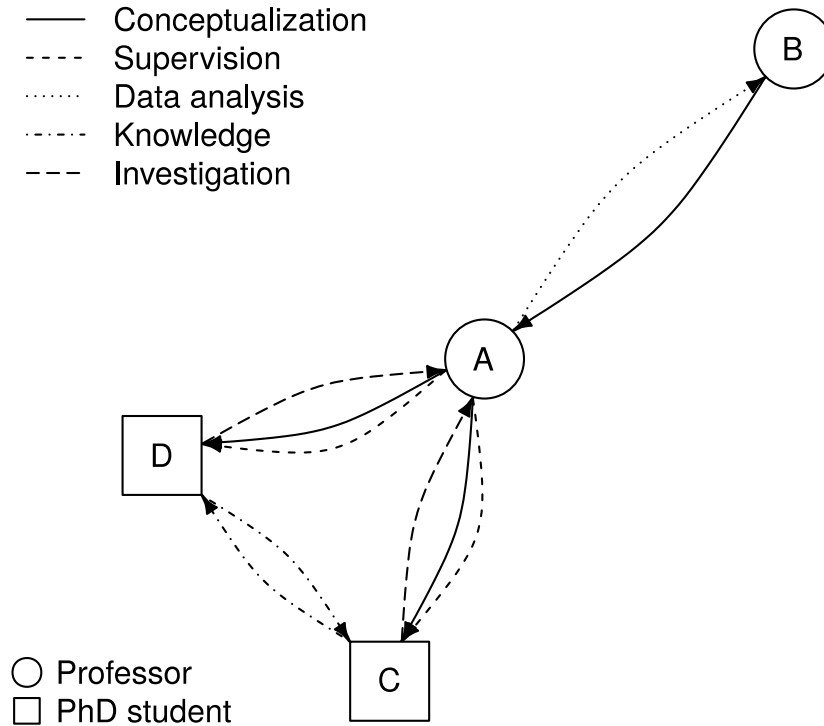
Figure 1: Collaboration as a process of pooling resources.

It often happens that scientists have different responsibilities or "roles"[1] in different scientific projects. Scientist A in Figure 1 is responsible for data analysis in his collaboration with B, but plays a different "role" in his collaborations with C and D by conceptualizing research ideas and providing supervision. From A's perspective there is a *diversity* in the sets of resources he *contributes to others* – his contributions to C and to D are more similar to each other, but both are different from his contribution to B. There is also a diversity in the sets of resources *others contribute to* A. Contributions of C to A and of D to A are more similar to each other, but different from contributions of B to A. Such defined diversity of contributed resources constitutes an aspect of the above mentioned multiplexity. On the one hand, if an actor is characterized with a low diversity of sets of resources contributed to others, he will have *similar* types of outgoing ties in all his collaborations. Certain types of ties will co-exist in all the non-empty dyads he is involved in. On the other hand, for an actor characterized with a high diversity, the sets of resources contributed in different dyads will be *dissimilar*. The general question we ask is:

- What social mechanisms are responsible for the diversity of resources engaged in different collaborations?

---

[1] Here we use the term "role" in a more casual sense of the word than the strict meaning in the social networks literature (e.g. Wasserman and Faust 1994:Ch. 9).

Assuming that collaboration ties are created purposefully, as a result of goal-directed behavior (Coleman 1994), we may expect actors to seek collaborators that will be "good matches" for the set of resources they possess. A good match here means that it promises a successful collaboration. However, we argue, the formation of collaborative relations is more complex than dyadic resource match-making because, among other things: (1) actors face constraints on the use of resources (e.g. there is a limited number of PhD students one can supervise); (2) the ability to provide certain resources might be related to specific attributes of actors (e.g. only professors can supervise PhD students); and (3) the demand for a particular resource among potential collaborators depends on what other types of resources these collaborators contribute to other collaborations (e.g. a data analyst may need somebody to do a data analysis because he is busy providing similar service to somebody else). In other words, we can expect that mechanisms of diversification (or specialization) of resources contributed to/by others might depend on properties of the actors involved as well as on the patterns of connections among those actors. Therefore, we ask a more specific question:

- How the diversity of resource bundles engaged by actors in different collaborations depends on the broader structure of the collaboration network?

For example, in the Figure 1, Actor A contributes similar bundles of resources to collaborations with C and D, who also collaborate with one another. At the same time, A contributes a different resource bundle to B, who does not collaborate with any other collaborators of A.

Section 2 starts with providing more background and motivations for our approach. We propose candidate explanations and hypotheses how network structure can influence the diversity of resources contributed to collaborations in Sections 2.1 and 2.2. In particular, we formulate hypotheses built upon concepts of structural holes and specialization. We confront the developed explanations with a small but rich data set based on a qualitative study. Section 3 describes the process of data collection, measurement and data set construction. In Section 4, we introduce the concept of *pairwise redundancy* and its measure which operationalizes the hypothesized role of structural holes in differentiating the resource bundles contributed to different collaborations. The section also describes the multilevel statistical model we apply to the data. The results are presented in Section 5. The paper is concluded with the discussion in Section 6 and auxiliary details were put into Appendices A1, A2, and A3.

2 Resources and relations

Scientists undertake many kinds of activities, many of which have a collaborative aspect (Boyer 1997). Yet a substantial part of research on scientific collaboration is based on co-authorship data. It brought many insights, such as evidence of different research paths scientists might take corresponding to three basic collaboration network substructures formed by scientists (Moody 2004). Co-authorship data also allow for the analysis of the growth of collaboration structures over time (Wagner 2009), the emergence of new scientific disciplines similar to classical bibliometric studies (Nobre and Tavares 2017; Terekhov 2017), factors improving scientific productiveness (Albarrán, Carrasco, and Ruiz-Castillo 2017), gender inequalities (Hildrun, Alexander, and Johannes 2012), and more. However, scholars indicate that co-authorship data represent only a certain fraction of collaboration activities (Sonnenwald 2007) that are of a very particular kind (Lewis, Ross, and Holden 2012). While co-authorship studies allow for addressing mentioned

research questions, and often at a considerable scale, their ability to explain why some scientists collaborate and some other not is limited. The limitation comes, among other things, from the character of bibliographical data which is rather scanty in information relevant for explaining collaboration.

One approach to understand why scientists collaborate is to think about *incentives* that might lead them to do so. According to Lewis et al. (2012), collaboration in science can be of two types: (1) tangible, concrete, and instrumental, e.g.: designing and conducting a study together, jointly creating specific scientific artifacts such as publications, technological prototypes, production processes, or product designs, and (2) more fluid, relying on discussion, feedback, and commentary. In both cases, the incentives to form a new collaborative relation come from *resources* scientists possess or control and the *interests* they might have in resources possessed or controlled by others. For example, an experimentalist might be interested in competencies of a theorist and the resources would correspond to abilities to conduct experiments and develop a theory in a particular research problem. As theorized by Coleman (1994:Ch. 2), actors are interested in resources and pursue these interests through, e.g. engaging in exchanges with other actors. Such exchanges probably take place among scientists too.

Alternatively, collaboration in science might be conceptualized not so much as an exchange, but rather as a process of collaborative creation (Ridley 2011). Scientists' primary interests are not in the resources themselves, but rather in the outcomes of scientific work that needs these resources as inputs. For example, researchers are interested in publishing an experimental research article. To achieve that goal, it is necessary to design the experiment, conduct it, analyze the data, and write the article. Resources would correspond to abilities to provide or accomplish these smaller tasks efficiently, e.g. specific skills, access to equipment, and so on. To some extent the incentive structure in such a "co-creation" setting seems similar to exchange. By contributing different resources to a common endeavor, actors "exchange" time spent on different tasks as if agreeing to arrangements such as "I'm better at data analysis, so you do the theory."

Such a resource-based perspective has been elaborated and applied in various settings across the social sciences (e.g. Ekeh 1974; Cook 1987; Lazega and Pattison 1999; Bearman 1997). It has also been applied to the analysis of co-authorship data. However, proper operationalization and measurement of resource contributions become a challenge. For example, Schummer (2004) use departmental affiliation as an indicator of expertise and knowledge bound in scientific disciplines interpreting inter-departmental collaboration as bringing different types of expertise (resources) together. A similar logic was undertaken by, e.g. Bordons et al. (1999) and Qin (1994). It may be debated if the fact of a different departmental affiliation is precise enough as evidence of collaboration through pooling diverse knowledge or methods. Unless additional data on individual responsibilities are available (e.g. Corrêa Jr et al. 2017), co-authorship analysis usually relies on similar assumptions. To overcome such limitations, scientists turn to data of different nature including computational methods such as text mining (e.g. Wang, Notten, and Surpatean 2013; Cheng et al. 2015) and sociological methods, e.g. surveys and interviews (e.g. Youtie and Bozeman 2014; Lazega et al. 2008; Jian and Xiaoli 2013; Laudel 2001). We follow the latter approach as, at the expense of scale, it allows for reconstructing individual collaborations and resources involved in a much greater detail, as we describe later in Section 3.

Convenient abstraction for the approach we advocate in this paper is to represent contributions of different resources between the scientists as a multiplex network. As we have signaled in the

Introduction and Figure 1, types of resources contributed are represented with different types of directed ties in a multiplex network of scientists. Study of multiplex social networks is an established area of research (see Kuwabara, Luo, and Sheldon 2010 for a review), especially applied to collective settings such as teams, firms, and other types of organizations. Multiplexity has been used as an explanatory factor. For example, Lazega et al. (2008) investigated how academic success depends on, among other things, position in different types of interpersonal networks. Similar approach was used by Podolny and Baron (1997) to explain intra-organizational mobility (grade advancement).

In contrast to taking multiplexity as given our goal is to explain certain aspects of it. In that sense, our research questions are similar to those of Lazega and Pattison (1999), who analyze how different types of ties (co-working, friendship, and advice) co-exist in a law firm. Their approach was fitting Markov Exponential Random Graph Models (Pattison and Wasserman 1999) to a multivariate data set corresponding to the multiplex network. Coefficients of the estimated models give insights about the interdependencies between different types of ties. For example, they find that it is likely that lawyers seek advice from coworkers of their coworkers, but it is not likely to seek advice of advisors of their coworkers or coworkers of their advisors (Lazega and Pattison 1999:84). Many other equally or more complex interdependencies have been identified. One drawback of their approach is that it is difficult to formulate hypotheses for and interpret models containing many (in the order of dozens) parameters. The other is that it has been since identified that Markovian ERGMs are very often characterized with model degeneracy (Schweinberger 2011) which makes the estimates unreliable.

As will become more evident in the coming sections, we postulate somewhat simpler hypotheses for the dependency structure between ties corresponding to different resources. Implications of our hypotheses can be assessed with a simpler method of looking at, on the one hand, similarity of resource bundles contributed to different collaborations (Section 4.1) and, on the other hand, a measure of relative redundancy of alters in a collaboration network (Section 4.2). When looking at multiplexity from that perspective, there are two relevant concepts of how the structure of collaboration network might influence the diversity in resource bundles contributed to and received from collaborators: Ronald Burt's concept of structural holes, which we elaborate in Section 2.1 and the concept of specialization which we elaborate in Section 2.2.

2.1 Resources and structural holes

The concept of structural holes introduced by Burt (1995) has become an important approach in social capital research (Crossley et al. 2015). Burt showed how network structure improves access to diverse resources and in return increases individual output like creativity and good ideas. The ideas are built upon Simmels idea of "tertius gaudens" (Simmel 1972). Burt focused on a general setting in which social networks are a source of benefits that can be accessed directly from network peers, but also indirectly – through peers from other members of the network. The theory is well illustrated when we think of the benefits related to resources circulating in the network. In general, it is beneficial to have as many different connections as possible because they provide multiple channels the resources can come to an actor and also the resources can be sent by an actor to others. Access to more groups translates to access to more novel ideas, which can be then adapted in different social circles.

Because maintaining social ties is costly, it becomes necessary to economize and maintain the relations that are *efficient* and not *redundant*. Some ties are "redundant to the extent that they lead to the same people, and so provide the same benefits" (Burt 1995:17). Burt distinguishes *redundancy by cohesion* from the *redundancy by equivalence* to separate the case in which alters are redundant because they are directly connected to each other from the situation in which alters are redundant because they broke relations to the same set of others (Burt 1995:Ch.1). Figure 2 illustrates this point.
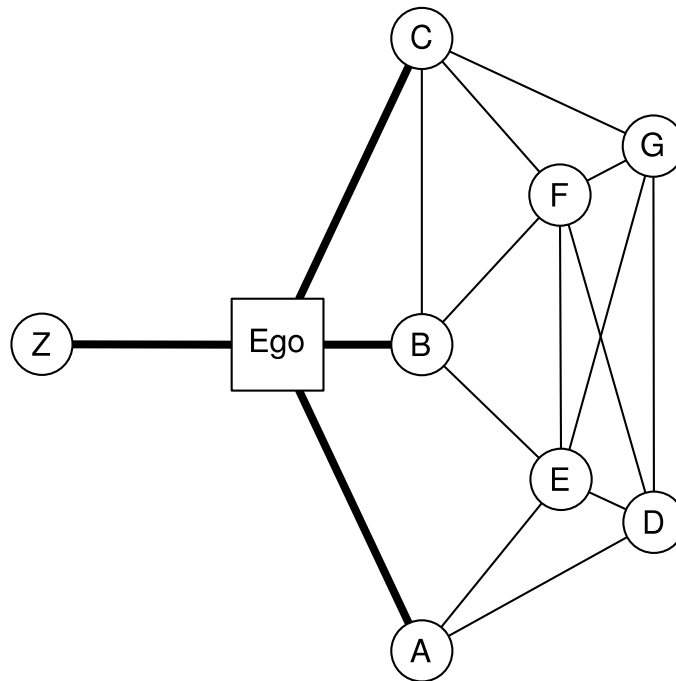

Figure 2: Burt's redundancy in egocentric networks.

Actors B and C are redundant to Ego by cohesion because they are directly connected to each other. Simultaneously, actors A, B, and C are redundant by equivalence because they are all connected to the group D-E-F-G. Should an important resource originate from that group, the ego will learn it through any of A, B, or C. Actor Z is a non-redundant contact because he is not related directly or indirectly to other direct contacts of Ego. Following Burt's argument we can expect that a purposeful collaborator might form non-redundant collaborations because these ties will bring different resources from those that he can acquire elsewhere[2].

---

[2] From the empirical point of view it is important to recognize that only the redundancy by cohesion can be reliably measured in an ego-centric network study. Consider again the illustration above. In an ego-centric study, we capture the ties of Ego to his alters A, B, C, and Z and connections among them. We can learn that B and C are redundant by cohesion, but we have no empirical information that will tell us that A, B, and C are also equivalence-redundant.

Given the above, we can expect that bridging structural holes by maintaining structurally non-redundant collaborations will be associated with relatively unique sets of resources contributed in those collaborations. Putting it succinctly:

H1. Egos will acquire more similar sets of resources from alters who are redundant by cohesion.

In the ego-network in Figure 2, Actors B and C are redundant, so according to *H1* we would expect that their contributions will be more similar to each other than contributions of Actors C and Z to each other as Actors C and Z are not redundant. Do note that the presented argument and hypothesis require a somewhat non-standard operationalization of redundancy. We are not considering redundancy as a property of an alter, but rather as a property of *pairs of alters*. This concept is further illustrated qualitatively in Section 2.3 and fully developed in Section 4.2.

The structural holes mechanism has been intensively investigated in business environments (e.g. Allen 1977; Katz and Tushman 1981; Burt 2004; Zaheer and Bell 2005; Tiwana and Keil 2007). The proximity of structural holes can also improve access to valuable resources in science. Recently, Bellotti (2012) showed that occupying brokerage position in a scientific community is more important for getting funded than a prestigious position in a scientific field or a recognized affiliation. Lopaciuk-Gonczaryk (2016) indicated that having collaborators who do not collaborate with each other is related to increased publishing productivity. A handful of research shows that sparse networks rich in structural holes result in better scientific output (Abramo, D'Angelo, and Solazzi 2010; Andrade, Los Reyes Lopez, and Martin 2009). Merton and Barber (2006) highlight the role of serendipity in scientific discoveries, which is possible only in sparse, diverse, and "accidental" networks.

2.2 Resources and specialization

According to Burt, redundant ties are a source of more similar resources. We propose that similarity of resources could also be decreased through specialization. On the individual level, a mechanism of specialization takes place when an actor provides all his collaborators with a similar set of resources but resources provided by collaborators differ from each other. Only through pooling resources a desired outcome can be achieved. Individuals specialize to reduce effort, which is required to maintain a larger number of ties.

Specialization as a mechanism which decreases similarity of resources seems to be contradictory to the concept of structural holes. Specialization in science has, however, become one of the most recognized phenomena over the last years. There are two general explanations for the increasing specialization. The first one is the growing complexity of scientific endeavors (Leahey 2016). The classical studies connect it with the processes of centralization and increasing importance of technology (Hagstrom 1964). The second reason is professionalization of science (Beaver and Rosen 1979) when scientists have become experts in particular fields (Gibbons et al. 1994). Freeman, Ganguli, and Murciano-Goroff (2015) identified empirically that it is "access to specialized human capital" that pulls scientists into collaboration. In general, specialization sustains an uneven distribution of research-relevant resources in the scientific community.

We argue that specialization is a mechanism complementary to structural holes and it takes place only in densely collaborating groups. If we assume that this process is unconditional in the sense of not depending on the collaboration network in any way, we would expect that an ego will

specialize in a certain set of resources and contribute only those to his collaborations. In other words, the sets of contributed resources will be similar in all his collaborations and the pattern of collaborations among alters should not matter. Scientists operate in various formal and informal research teams. The composition of formal, institutionalized teams is often influenced by external factors such as institutional hiring procedures. Apart from such factors, scholars might look for collaborators with complementary sets of resources elsewhere. Research teams would span institutional boundaries (Jones, Wuchty, and Uzzi 2008) and specialization would occur within these extended research groups. Therefore, specialization is present in ties redundant by cohesion that share multiple other collaborators but not in ties with a single redundant tie regardless of institutional affiliation.

H2. Egos collaborating with alters belonging to a densely connected group acquire dissimilar sets of resources from these alters.

2.3 Structural holes and specialization: a rejoinder

Let us bring together and summarize the considerations above. Figure 3 shows two ego-networks of a Scientist 0 in two ideal-case situations corresponding to our hypotheses from Sections 2.1 and 2.2 above. For simplification, but without the loss of generality, we reduced the complexity of the pictures by showing only a single resource type being contributed in a particular direction in every dyad. If the same type of resource is being acquired by Actor 0 in different dyads we expect not necessarily identical resources, but a higher *similarity of resource bundles* in these dyads.

(A)
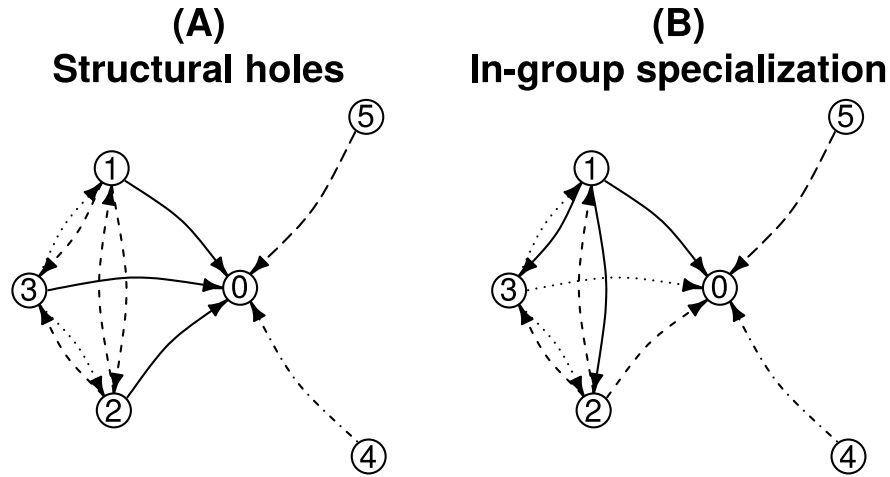**Structural holes**

(B)
In-group specialization

Figure 3: Implications of structural holes and specialization hypotheses. (A) The more redundant the alters are, the more similar resources they contribute. Ego acquires similar resources from Actors 1, 2, and 3 because they are redundant (B) specialization only takes place in a tightly collaborating group (0-1-2-3). Ego acquires different resources from actors 1, 2, and 3 because they specialize within the group of 0, 1, 2, and 3.

Network in Panel A of Figure 3 shows idealized situation in which resources are contributed according to the structural holes argument. In line with our *H1*, alters who are redundant, here 1, 2, and 3, contribute similar resources to ego 0. Alters who are non-redundant to 0, namely 4 and 5, are more likely to contribute different resources than anybody else.

9

Network in Panel B corresponds to a situation of perfect within-group specialization. Scientists 0, 1, 2, and 3 form a tightly collaborating group. Should the specialization take place only within such groups, we see, in line with *H2*, each member of the group to contribute the same type of resources to all other members. In particular, the ego 0 receives different resources from each of the members 1, 2, and 3. Alters 4 and 5 who are not members of that group may contribute still other types of resources.

Table 1 summarizes these hypotheses. According to structural holes argument, we should expect similar bundles of resources contributed by redundant alters (1 and 2) – effect of redundancy on similarity is positive. Specialization argument implies that these contributions will be dissimilar – effect of redundancy on resource similarity is negative.

Table 1: Directions of effects of redundancy on similarity of acquired resources under *H1* and *H2*.

Hypothesis	Effect of redundancy on similarity of resources acquired by ego
H1: Structural holes	+
H2: In-group specialization	–

We can think of specialization as a process of reducing costs of structural redundancy. Redundant ties may be worth maintaining, but only if they convey resources unique for ego within the group.

2.4 Other considerations

Science is a highly institutionalized social setting (Whitley 2000) with degrees and hierarchies which together determine various formal and informal aspects of scientific collaborations. Academia was historically built upon the hierarchical master-student relation and for centuries it was a dominant mode of work (Perkin 2007). We expect the theory laid out in Sections 2.1 and 2.2 to operate independently of other processes. In particular, we may expect that the similarity of resources acquired by ego from alters may be affected by certain attributes of egos as well as by general "similarity" of alters to each other on various dimensions. In the analyses presented below, we control for three such variables:

(1) Ego's scientific degree – following the specialization argument we might expect that the higher the scientific degree the more specialized ego is the more similar resources he will acquire from alters.

(2) Alters' institutional similarity – following the logic from Section 2.2, we may expect pairs of alters affiliated with the same department, but not necessarily the same as ego, to contribute more similar resources.

(3) Alters' career similarity – some of the resources can be more specific to certain career stages. Supervision or providing academic contexts are examples of such resources. As a consequence, we may expect that pairs of alters who are at the same stage of scientific career (measured by their scientific degree) are more likely to provide ego with more similar resources.

Developing full theory with respect to the above variables is beyond the scope of the presented paper. The results presented in Section 5 include them as control variables.

3 Data collection and measurement

Data consists of 40 individual in-depth interviews conducted between April and August 2016 by two interviewers. The interviewees mentioned 333 collaborators in total. The sample consists of 20 female and 20 male scientists from six Polish cities. Respondents represented a broad range of disciplines: natural sciences, social sciences, life sciences, the humanities, engineering, and technology on different levels of career from PhD candidates to professors. The detailed description of the sample can be found in Appendix A2 and in Bojanowski, Czerniawska, and Fenrich (2020). The data set is available online[3].

Each interview consisted of four parts. After the initial introduction and a short description of respondent's professional interests (part one), respondents were asked to name up to 10 most important collaborators during last 5 years (part two). A collaboration might have already ended at the time the interview was conducted, but at least part of it have had to be undertaken in the indicated period. Scientific collaboration was defined broadly as a shared process of creating new knowledge and mutual help in intellectual endeavors to include less standardized collaboration practices. The diversity in scientific collaboration practices is often underlined in the literature (Beaver 2001; Katz and Martin 1997). The definition used during the interviews was built upon definition from Lewis et al. (2012), which includes two types of collaboration: "collaboration" and "Collaboration" (with capital "C"). The first term describes situations in which the relation is fluid, mostly relying on a discussion, feedback, and commentary while the second is more tangible, concrete, and instrumental, including designing and conducting a study together as well as later publications.

Respondents were asked to mention up to 10 "most important" collaborators to reduce the possibility of pointing collaborators not relevant for respondents' work. The interviewees were asked about collaborators' gender, scientific degree, and institutional affiliation. If the respondent felt uncomfortable with revealing full names of collaborators, s/he was only asked for unique nicknames. All collaborations were discussed separately (part three). Respondents were asked about the history of collaborations, the merits for collaboration, resources each party engaged in collaboration, and rewards gained from collaboration. Respondents were provided with leading questions about resources which might be engaged and gained from a collaboration such as knowledge and skills, contacts, funding opportunities, equipment, prestige, or joint publications. Respondents were also asked if there were any negotiations of the terms of collaboration. Names of all collaborators were attached to a cork board with respondent in the central point (part four). Respondents were asked to indicate all collaborations among her and her collaborators. Collaborators on the cork board were represented with pins and collaborations were represented with rubber bands (see Figure 4). There were two follow-up questions about mutual dependencies

---

[3] Available at: https://recon-icm.github.io/reconqdata/

of collaborations and about possible collaborators crucial for any tie, which were not included in the interview.
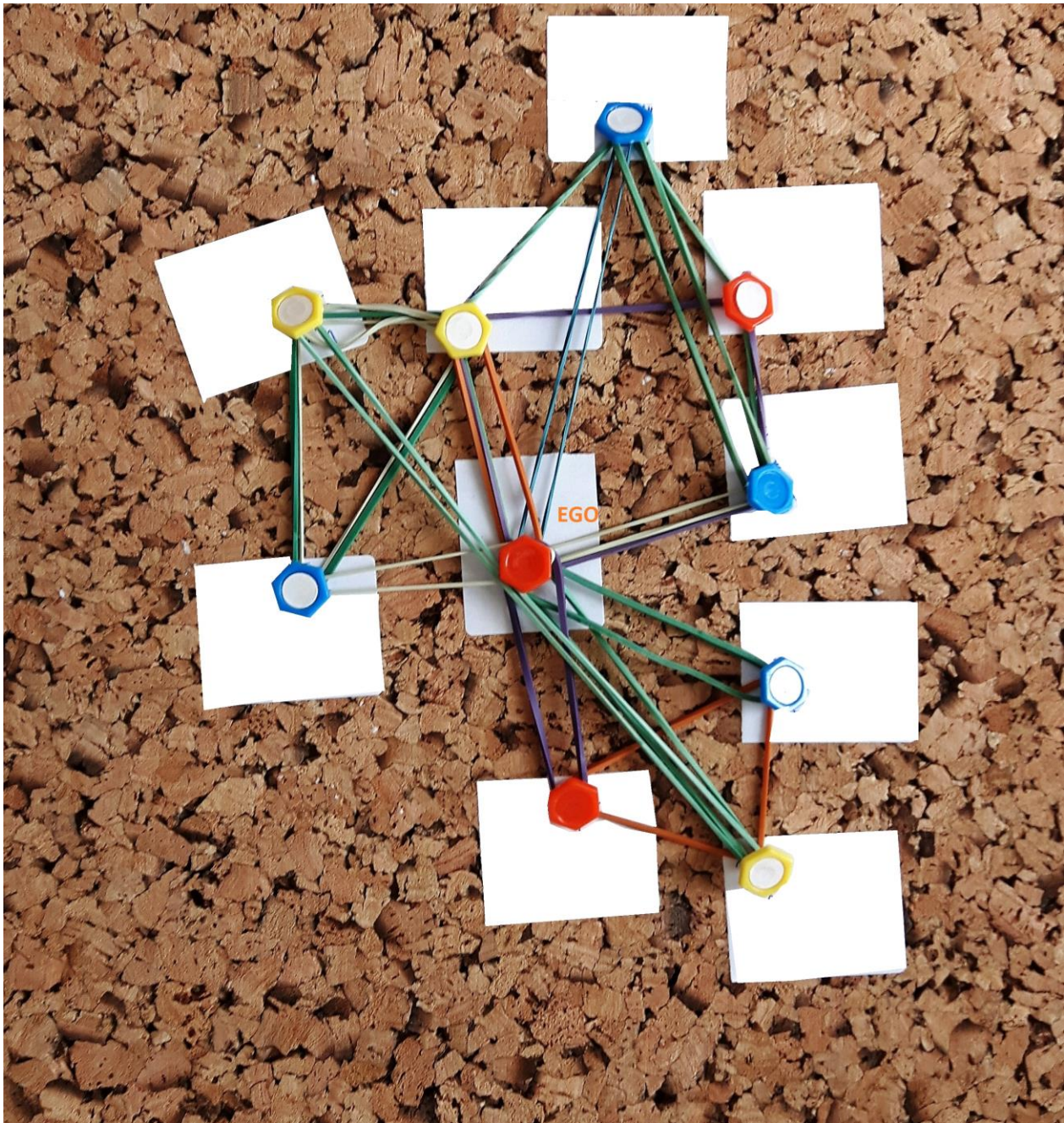

Figure 4: Collecting data on collaboration networks.

Interviews, lasting from 24 to 90 min, were recorded and transcribed. The cork boards with information about collaborations were photographed.

Information about collaboration ties was recovered from photographs of cork boards prepared during the interviews, such as the one presented in Figure 4. Collaborators and collaborations were

labeled with unique identifiers and assembled into a two-mode network data set with modes corresponding to persons (pins) and collaborations (rubber bands), respectively.

Information about collaborators was coded based on transcripts. Respondents and collaborators were described with information about gender, scientific degree, scientific discipline, department (if possible), university, city, and country. Some collaborators had more than one affiliation or discipline. The primary discipline and institutional affiliation were chosen based on Polish Science Database. Missing pieces of information were retrieved from the internet.

3.1 Measuring resources

Data about resources engaged by respondents (egos) and their collaborators (alters) to every collaboration were coded based on transcripts. The coding was done with QDA Miner Lite software[4] and conducted by two persons. Random subsample of the interviews was double-checked by different researchers to ensure reliability.

Resources engaged in collaborations were coded with a coding scheme covering different elements of a research process in different disciplines. The coding scheme consisted of the following categories:

- "Conceptualisation" – coming up with an idea for a study, providing general theoretical framework, designing a general framework for a study;

- "Methodology" – designing methodology for a study;

- "Investigation" – conducting research, gathering data;

- "Data analysis" – data analysis, quantitative as well as qualitative;

- "Data curation" – managing and archiving data;

- "Software creation" – writing software for research process;

- "Prototype construction" – building a prototype that is used in research process; and

- "Knowledge" – knowledge oriented help in research process but not falling into any of the above categories.

Coding scheme also includes different tangible and intangible resources, which might be controlled by scientists. The list of resources was built upon a literature review.

- "Funding acquisition" – the increasing role of ability to secure funding was underlined in many studies on contemporary science (Resnik 2006; Mirowski 2011).

---

[4] Available at: https://provalisresearch.com/products/qualitative-data-analysis-software/

- "Writing" and "Proofreading" – the role of written forms of knowledge was summarized by Bazerman (1983). It is one of the most important activities in science (Popper 1972; Merton 1973).

- Administration – it is one of the main factors interfering with scholarship (Blau 1994). It can be grouped into:

    - "Project administration" – administrative work over a project; and

    - "Formal administration" – bureaucratic work is a result of managing complex scientific institutions.

- "Data" – a large part of scientific work is organized around tangible resources such as data or documents (Latour and Woolgar 2013). The category consists of different types of data which can be used in scientific work: qualitative, quantitative, literature reviews.

- "Equipment" – Hagstrom (1964) and Knorr-Cetina (2009) indicated the crucial role of technology and scientific equipment in shaping scientific collaborations and scientific practices including centralization of collaboration in some disciplines.

- Contacts – the role of social contacts surrounding in the production of knowledge has been underlined in vast literature (e.g. Collins 1974). According to literature, we can group contacts into two categories:

    - "Contacts in academia" (Bellotti 2012); and

    - "Contacts outside academia" (Powell and Owen-Smith 2012).

- "Supervision" – a master-student relation is the most traditional collaboration in academia. It has a significant impact on a career in academia (Wagner 2011; Zuckerman 1967).

- Position in academia:

    - "Prestige" – Bourdieu (1988) indicated that the symbolic power was the main driver for an accumulation of different goods in academia. Some collaboration might be attractive because they are seen as prestigious.

    - "Formal achievements" – contemporary science has developed many forms of formal accountability, where achievements are measured according to designed indicators such as a list of publications.

- "Character traits" – Scientific collaboration like any other teamwork is affected by collaboration skills and traits of character of all parties engaged. The literature on individual traits of character and scientific collaboration is extremely limited except some research on the role of collaborative skills in academia-industry collaboration (Siegel et al. 2003). "Character traits", which include different aspects of collaboration like being agreeable, reliable, or organized, might be an important characteristic of a potential collaborator.

- "Motivation" – one of the character traits, which does not affect collaboration directly but is of great importance in academic setting (Gatfield 2005).

- "Career development" – the studies of scientific biographies also raise questions about breakthrough moments in scientific careers. Research on contemporary Polish science indicates that for many scientists it was exposure to international science. It was usually enabled by collaborators who helped them to get international scholarships, gave access to some rare data or training or wrote a recommendation letter (Lazarowicz-Kowalik 2015).

- "Other input" – many scientific collaborations have a unique character. As a result, some resources are very specific to the local background. To avoid excessive fragmentation of the coding scheme, we have decided to introduce a category that will encounter for the resources unique for particular resources across all interviews.

Several examples of coded interview fragments are presented in Appendix A3.

4 Methods

We have a set of N actors. Let us define the collaboration network as an undirected graph $X = [x_{ij}]_{N \times N}$ where $x_{ij} = 1$ if i collaborates with j and $x_{ij} = 0$ otherwise. No self ties are allowed so $x_{ii} = 0$ for all i. Example of such a network is shown on Figure 5.
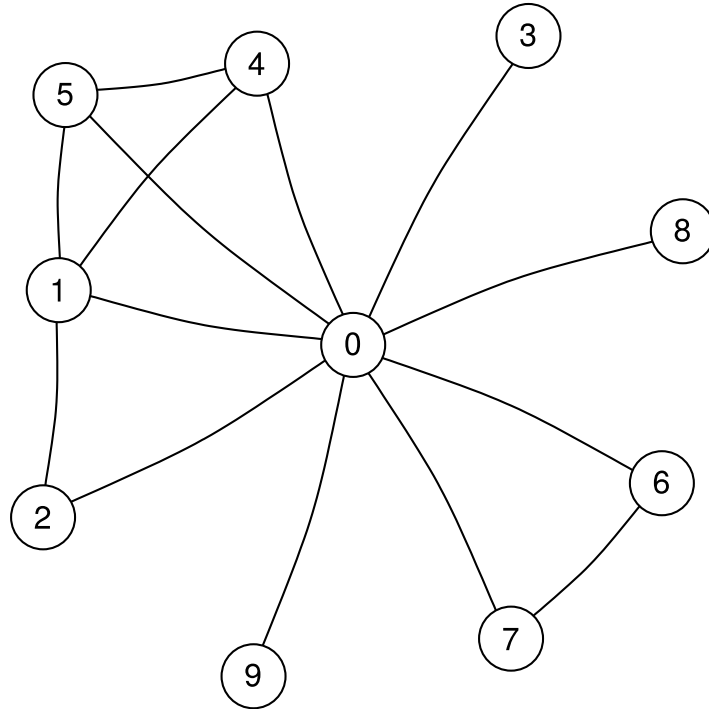

Figure 5: Collaboration network.

Actors can engage various types of resources when collaborating with others. Let us have R types of resources. The engagement of resources in collaborations can be represented as a directed resource flow network $Y = [y_{ijr}]_{N \times N \times R}$, an array in which $y_{ijr} = 1$ if resource r is engaged by actor i in her collaboration with actor j. In other words, resource r "flows" from actor i to actor j. As with the collaboration network, no self ties are allowed so $y_{iir} = 0$ for all i and r. Let us use $*$ for denoting all elements along the specific dimension, for example y_{0*1} would be a binary vector indicating collaborators of actor 0 with whom she engaged the resource 1. Example of a resource flow network is shown in Figure 6. As the data come from an egocentric study, we impose the convention that ego (the respondent) has an index value of 0.
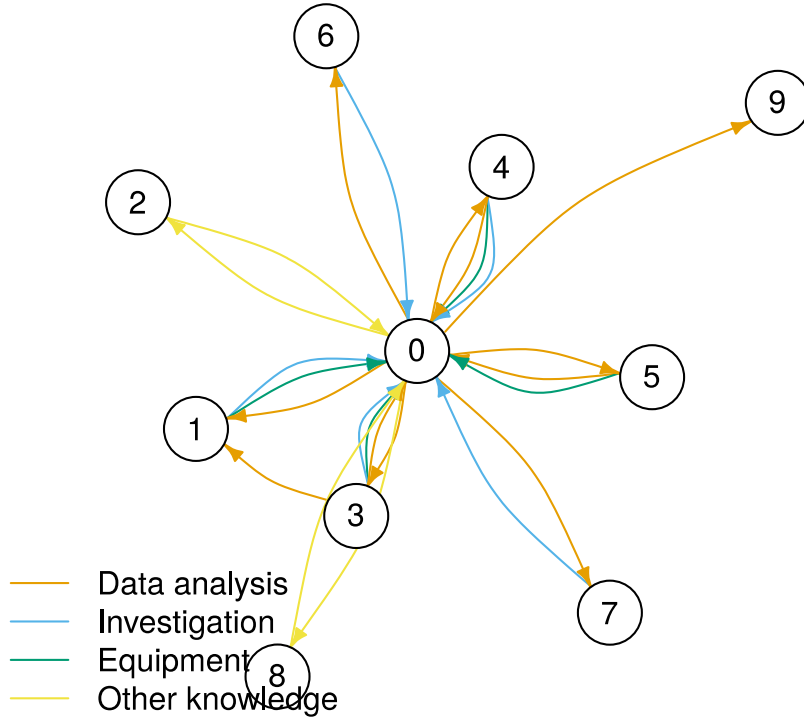
Figure 6: Network of resource flows.

To test our hypotheses we need to compare different collaborations in terms of (1) similarity of resources contributed by scientists in the resource flow network; and (2) the extent of structural redundancy in collaboration network. We propose how to measure these concepts below.

4.1 Resource similarity

To measure our dependent variable – the similarity of resources across different collaborations – we focus on R-tuples of resources. There are two aspects that need to be differentiated:

(1) Resources contributed by alters to their collaborations with ego; and
(2) Resources contributed by ego to his collaborations with different alters.

In both the cases, we would like to assess the extent to which tuples of resources are different across different collaborations. Focusing on case (1) above we are comparing a binary vector y_{i0*} with binary vector y_{j0*}. Let $\delta_{ij} = f(y_{i0*}, y_{j0*})$ be some measure of similarity. Choices for function f are plenty, see Choi, Cha, and Tappert (2010) for a review. We have chosen Jaccard coefficient (Levandowsky and Winter 1971) which in our context can be defined as:

$$f_{\text{Jaccard}}(y_{i0*}, y_{j0*}) = \frac{\sum y_{i0*} y_{j0*}}{\sum y_{j0*} + \sum y_{j0*} - \sum y_{i0*} y_{j0*}}.$$

The coefficient varies between 0 and 1. It is 0 if there is no single common resource contributed by i and j to actor 0 (the ego). It is 1 if the sets of resources contributed by i and j are identical.

The case (2) above is analogous, but we need to compare vectors y_{0i*} (contribution of ego to i) with y_{0j*} (contributions of ego to j).

In the example flow network above, a vector of resources contributed by 4 to 0 is $y_{40*} = (1,1,1,0)$ while contributed by 7 to 0 is $y_{70*} = (0,0,1,0)$. The value of Jaccard similarity is 0.3333333 as 1/3 of types of resources ever used in those two collaborations are common to both.

4.2 Pairwise redundancy

To test *H1* and *H2*, we need a measure of redundancy. The literature provides several options. For example, effective size, efficiency and constraint characterize ego-networks with respect of tie redundancy (Hanneman and Riddle 2005). Dyadic redundancy introduced by Burt (1995) measures the redundancy of each ego-alter tie within one ego-network. It increases if the alter has many ties to other alters of ego. From the perspective of our research questions and hypotheses, unfortunately none of those measures are directly applicable. Consider again the collaboration network from Figure 5. Ego ties to alters 2 and 6 will have the same dyadic redundancy scores as both 2 and 6 have one tie to others in ego's neighborhood. This suggests that these two alters are redundant "to the same extent." However, alters 2 and 6 are likely to be sources of different resources/information because each one is connected to different group of alters (1-2-4-5 and 6-7). In that sense we could say that 2 and 6 are "not" redundant "*vis-à-vis* each other." To capture such redundancy we need to look at "pairs" of alters and assess whether they belong to different parts of ego's neighborhood. Hence, we propose a measure of "pairwise redundancy."

To measure the extent to which collaborations of ego with i and j are pairwise redundant by cohesion, we can use several approaches. The approaches differ in terms of configurations of alter-alter ties that we are willing to interpret as necessary for making ego's ties to i and j redundant.

1. We may only look at whether the direct relationship between i and j exist. If it does, i and j are redundant for ego. They are not if the relation is absent. On the collaboration graph presented above, pairs of redundant alters for ego include (1, 2), (1, 4), (6, 7). While alters 3, 8, and 9 are non-redundant *vis-à-vis* all others. Note that according to this approach pairs (2,4) and (2,5) are "non-redundant" because they are connected, but only indirectly. This approach seems to be justified in contexts where network benefits can only travel for a distance of 2.

2. Should the network benefits travel further than 2 steps, we may want to treat alters further apart as redundant as well. A different approach would be to treat i and j as redundant as soon

as there is a path from i to j when we exclude ego's ties. In the example network all pairs of alters from the group (1, 2, 4, 5) are redundant and so is the pair (6,7).

Approaches (1) and (2) are binary measures. Such simplicity is also a limitation. We may want to measure the extent of redundancy to more closely represent possible complexities in the alter-alter network.

3. To differentiate between pairs of indirectly connected alters, we may measure the inverse of the shortest path between i and j in the network with ego's ties removed. In the example network such measure would be 1 for the pair (4,5) and 0.5 for the pair (2,5).

We may also further differentiate alter-alter pairs who are directly connected:

4. We may argue that i and j, who are directly connected, are more redundant if they belong to the same densely connected subgroup. In this sense (4, 5) are more redundant than (1, 2) because the former have alter 5 in common while the latter have no common alters apart from ego. Consequently, we may measure the redundancy of i and j by counting closed triplets involving i and j.

While there are different arguments for approaches (1-4), it is worth noting that they are complementary and can be combined into a single numerical index that takes values from the interval $[0, \infty)$ such that:

- It is 0 if i and j are not connected directly or indirectly, e.g. alter 3 versus others (option 2 above);

- It is in (0, 1) if i and j are connected only indirectly, e.g. alters 2 and 4. It is the inverse of the shortest path between i and j (option 3 above);

- It is 1 if i and j are connected directly with no shared partners, e.g. alters 1 and 2 (option 1 above); and

- It is in $(1, \infty)$ if i and j are connected directly and have common collaborators, e.g. alters 1 and 4 (option 4 above).

For the Figure 5 collaboration network, the pairwise redundancy scores are calculated in Table 2.

Table 2: Pairwise redundancy scores for the example collaboration network.

	2	3	4	5	6	7	8	9
1	1	0	2	2	0	0	0	0
2		0	0.5	0.5	0	0	0	0
3			0	0	0	0	0	0
4				2	0	0	0	0
5					0	0	0	0
6						1	0	0
7							0	0
8								0

4.3 Other independent variables

To control for similarity of resources resulting from institutional setting, we include other independent variables.

First, we measure alter-alter similarities with respect to their institutional affiliation. Alters affiliated with the same department are expected to be more likely to contribute similar resources to collaborations with ego. This should hold irrespectively of the extent of their structural redundancy for ego.

Second, we measure scientific degree of ego and alters. The variables have four levels: MA, PhD, habilitated PhD, and professor. This information will be used for two different purposes: (1) as the main effect on the ego-level to investigate how similarity of resources changes on different stages of academic career; and (2) as a basis for control of alter-alter similarity of scientific degree.

Table 3 summarizes the variables used in subsequent analyses.

Table 3: Descriptive statistics of variables used in the analysis.

Variable	Min.	Mean	SD	Max.
Level: Alter-alter				
Alters same degree	0	0.306	0.461	1
Alters same department	0	0.233	0.423	1
Pairwise redundancy	0	1.829	2.408	8
Similarity of alter contributions	0	0.161	0.246	1
Level: Ego				
Egos degree: MA	0	0.100	0.304	1
Egos degree: PhD	0	0.400	0.496	1
Egos degree: PhD hab.	0	0.225	0.423	1
Egos degree: professor	0	0.275	0.452	1
Network size	5	9.350	2.293	16

4.4 The model

To test our hypotheses, we estimate random intercept linear models in which level one corresponds to alter-alter comparisons nested in level two, the egos. The complete specification is:

$$s_{eij} = \sum_g \gamma_g \, f_g(P_{ij}) + \sum_h \gamma_h \, X_{hij} + \sum_q \gamma_q \, Z_h e + U_e + R_{eij}, \qquad \text{where } i < j.$$

The dependent variable on the left-hand side is similarity of resource bundles (s_{eij}) of alter i and alter j acquired by ego e, measured by the Jaccard coefficient. On the right-hand side we have the following independent variables:

- Pairwise redundancy of alters i and j (P_{ij}) in the collaboration network of ego e is modeled with fixed-effect linear splines $f_g()$. See below for details.

- Alter-alter similarities on other characteristics (X_{ij}). These are effects of variables capturing the same scientific degree, and the same institution (department) of alters. They are also modeled as fixed effects.

- Ego characteristics (Z_e): scientific degree.

- Level two (ego-specific) residual U_e with variance σ_U^2.

- Level one residual R_{eij} with variance σ_R^2.

Our main independent variable of interest is pairwise redundancy. It varies across the two neighboring intervals: $[0; 1]$ and $[1; \infty]$. The first one describes pairs alters who are not directly connected, the second one describes alters who do have a direct connection. We decided to model the effect of this variable with a linear spline having a knot at point 1. In other words, the effect of pairwise redundancy will be represented by two line segments $f_1(P_{ij})$ for $P_{ij} \in (0; 1)$ and $f_2(P_{ij})$ for $P_{ij} \in (1; \infty)$. These two segments might have different slopes (each with own coefficient in our models), but meet at $P_{ij} = 1$.

We performed data analysis using R (R Core Team 2017) and package `igraph` (Csardi and Nepusz 2006). The models were fitted using package `lme4` (Bates et al. 2015). Linear splines are fit using package `lspline` (Bojanowski 2017).

5 Results

The data have a nested structure: for each respondent, the ego, the data contains information about her individual characteristics (Level 2) and information about characteristics of all alter-alter comparisons among his network peers (Level 1). For example, if a particular ego has 8 alters, this is represented with (8×7)/2=28 level-1 observations. We estimate the following specifications:

1. Null model with no explanatory variables. Summarized in Appendix A1;

2. Model with fixed effects for all explanatory variables apart from pairwise redundancy (Model 1);

3. Model with fixed effects for all explanatory variables and pairwise redundancy as a linear effect (Model 2); and

4. Model above in which the linear effect of pairwise redundancy is replaced with a linear spline effect (Model 3).

Estimates and confidence intervals for fixed effects are presented in Figure 7. Detailed model results are available in Appendix A1.

Overall, the null model statistics indicate that percentage of variation in dependent variable that can be attributed to between-ego variation is equal to 16.1. AIC values indicate that models are improved by adding level-one and level-two variables (c.f. Table 5).
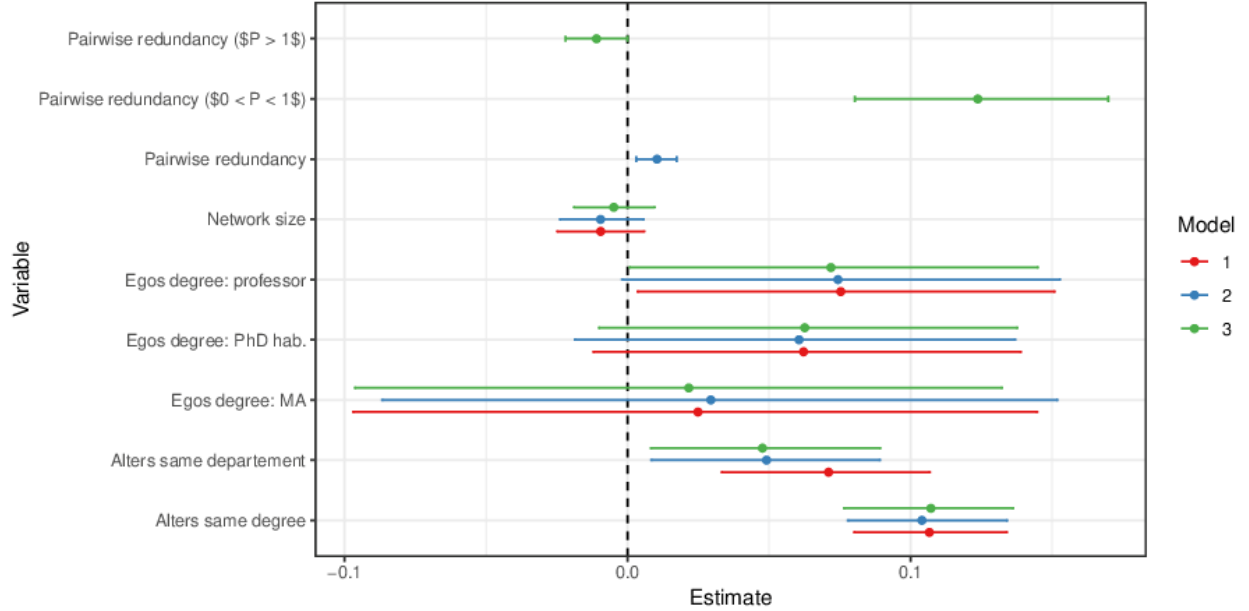

Figure 7: Estimates of fixed effects (dots) and Bootstrap confidence intervals (bars) from random intercept models for similarity of resources acquired by ego. Reference category for ego's degree: PhD. Confidence intervals are based on 1,000 draws. Detailed results can be found in Appendix A1 in Tables A.1 and A.2.

Let us start with our main variable of interest – pairwise redundancy. Including it as a predictor with a linear effect next to the other explanatory variables significantly improves the fit (Model 2 vs Model 1). The effect is positive and significant, albeit small. This implies that, on average, the more redundant the alters are to ego, the more similar are the resources contributed by those alters. As we have explained in Section 2.3, the structural holes and specialization arguments imply partially conflicting predictions with respect to the effects of redundancy of collaboration ties on similarity of resources contributed. Assuming that, in the structural holes argument, the effect of pairwise redundancy on resource similarity is constant across the redundancy scale, this result is in favor of our *H1*.

The specialization argument and *H2* formulated in Section 2.2 assumed that specialization takes place in densely collaborating groups. This corresponds only to the interval $[1; \infty]$ of pairwise redundancy. As a consequence, a proper test of these hypotheses is included in Model 3 in which pairwise redundancy is represented with a two-segment linear spline with a knot at 1. The effect is positive in the interval $[0; 1]$ and slightly negative in the interval $[1; \infty]$. This is consistent with *H2* implying that in densely collaborating groups alters are more likely to contribute dissimilar resources to ego. It should be noted that the negative effect in the $[1; \infty]$ interval is rather small (although significant).
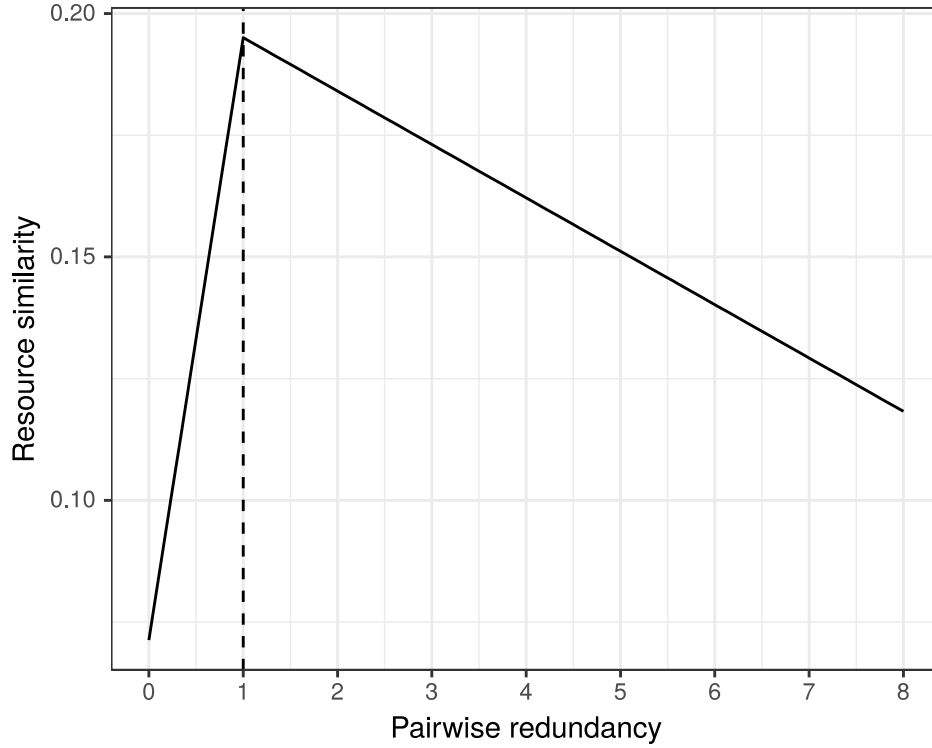
Figure 8: Effect plot of pairwise redundancy on resource similarity based on Model 3 assuming average ego-network size, alters from different departments, of different degree, and ego with a PhD degree.

Next to the specialization in "informal" dense groups we also expected that shared institutional affiliation would foster similarity in sets of resources contributed in collaboration with ego, namely alters from the same scientific institution having access to similar resources and contribute similar resources in collaborations with ego. Alters sharing institutional affiliation are more likely to contribute more similar resources, which is inline with our expectations. The effect persists even when pairwise redundancy is included in the model.

Turning to the career-related effects, the data do not provide enough power to estimate the effects of scientific degree of ego with good precision. The only tendency we can observe shows that if ego is a habilitated PhD or a professor, she is more likely to contribute more similar resources to all her collaborations (as compared to PhDs) irrespective of the properties of the alters and redundancy characteristics. Such trend is not present when analyzing alter contributions. We expected more similar sets of incoming and outgoing resources for alters with the same scientific degree. The coefficients are positive and significant in models for both dependent variables implying that alters with the same scientific degree indeed provide ego with relatively similar sets of resources and vice versa, as we expected.

6 Summary and discussion

Collaboration between scientists can be understood as a multiplex network of resource exchanges. We proposed to analyze this multiplexity by looking at the diversity of resource bundles

contributed to/by collaborators. We provided two hypotheses referring to two different mechanisms, which regulate the diversity of resources exchanged in collaboration ties between alters and egos: structural holes and specialization. To this end we developed a concept and measurement of "pairwise redundancy" designed to capture relative structural redundancy of alters "*vis-à-vis* each other." We also investigated how the diversity of resources depends on institutional co-affiliation and career-related characteristics.

According to the structural holes argument, unique resources should be accessible through collaboration ties bridging structural holes. Our results confirm that collaboration ties with alters who are pairwise redundant are more likely to convey similar sets of resources.

Additional findings indicate that collaboration with scientists sharing institutional affiliation resulted in increased similarity of incoming resources even after controlling for pairwise redundancy. We can conclude that both structural redundancy and institutional factors limit the novelty of sets of resources conveyed in collaboration ties.

The non-linear effect of pairwise redundancy in our models leads to interesting implications for redundancy and brokerage in scientific collaboration networks. The shape of the effect shows that the strongest effect of pairwise redundancy on diversity of resources contributed is observed when comparing a pair of alters who are completely disconnected in ego's neighborhood to a pair of alters who are directly connected. This corresponds to the $[0; 1]$ interval of our pairwise redundancy measure. Directly connected alters who do not share any other common collaborators with ego constitute a configuration in which the resources contributed by alters to ego are most likely to be similar. Further embedding of the said alter-alter relationship into the collaboration network of ego by adding more shared collaborators stops this tendency or even reverses it – in densely connected groups we find evidence of specialization.

We believe there are several avenues originating from the presented results along which further research can proceed. First, our arguments about the pairwise redundancy and specialization in collaboration networks deserve a more unified theoretical treatment. For example, it is somewhat unclear how these two mechanisms reconcile within, e.g.: (a) institutional research teams that are not very cohesive; and (b) cohesive collaboration groups that span across different institutions.

Second, our approach focuses on a collaboration network as a one mode network. We believe a promising extension, both theoretically and methodologically, would be to approach the problem with the formalism of a two-mode collaboration network. In such a network, scientists (Mode 1) are involved in "projects" (Mode 2). Such an approach should better handle possible multi-party collaborations, which often take place in science. It will also require a different design of data collection.

Appendix

A1 Detailed modeling results

Table 4 shows random intercept models described in Section 5. Model comparison statistics are presented in Table 5.

Table 4: Random intercept models for alter contributions. Null model: σ_U^2=0.01, σ_R^2=0.053, AIC=−56.91.

Variable	Model 1		Model 2		Model 3	
	Estimate	SE	Estimate	SE	Estimate	SE
Random effects						
Ego	0.08	–	0.08	–	0.08	–
Residual	0.05	–	0.05	–	0.05	–
Fixed effects						
Intercept	0.17	0.08	0.15	0.08	0.07	0.08
Egos degree: PhD hab.	0.06	0.04	0.06	0.04	0.06	0.04
Egos degree: MA	0.03	0.06	0.03	0.06	0.02	0.06
Egos degree: professor	0.07	0.04	0.07	0.04	0.07	0.04
Network size	−0.01	0.01	−0.01	0.01	−0.01	0.01
Alters same degree	0.11	0.01	0.10	0.01	0.11	0.01
Alters same department	0.07	0.02	0.05	0.02	0.05	0.02
Pairwise redundancy	–	–	0.01	0.00	–	–
Pairwise redundancy ($0<P<1$)	–	–	–	–	0.12	0.02
Pairwise redundancy ($P>1$)	–	–	–	–	−0.01	0.01
Model statistics						
AIC	−121.51	–	−127.58	–	−150.47	–
df	9.00	–	10.00	–	11.00	–

Table 5: Model comparison statistics for alter contributions.

Model	Df	AIC	Deviance	χ^2	χ^2 df	*p*-value
Null model	3	−56.91438	−62.91438	–	–	–
Model 1	9	−121.50963	−139.50963	76.595259	6	0.000
Model 2	10	−127.57490	−147.57490	8.065262	1	0.005
Model 3	11	−150.46507	−172.46507	24.890172	1	0.000

A2 Sample composition

Table 6 presents the sample composition with information about city (anonymized), gender, scientific degree, branch of science, and network size. Branches of science are categorized according to "Bill of Higher Education" introduced in Poland in 2011. Network size equals egos' degree.

Table 6: Basic information about the respondents in the sample

ID	City	Gender	Scientific degree	Branch of science	Network size (degree)
201	A	Male	PhD	Technical sciences	10
202	A	Male	PhD	Social sciences	7
203	A	Male	Professor	Technical sciences	9
204	A	Male	PhD	Technical sciences	8
205	A	Female	PhD	Life sciences	4
206	A	Male	PhD	Natural sciences	4
207	A	Female	PhD	Social sciences	10
208	A	Female	PhD	Humanities	5
209	B	Female	Habilitated PhD	Life sciences	9
210	B	Male	PhD	Humanities	8
211	B	Male	Habilitated PhD	Life sciences	10
212	B	Female	PhD	Social sciences	10
213	B	Male	Professor	Natural sciences	7
214	B	Female	PhD student	Natural sciences/Engineering	4
215	B	Female	Habilitated PhD	Life sciences	10
216	B	Male	Habilitated PhD	Natural sciences	10
217	B	Male	Professor	Technical sciences	10
218	C	Female	Habilitated PhD	Humanities	10
219	C	Male	PhD	Engineering	10
220	C	Male	Habilitated PhD	Natural sciences	6
221	D	Male	PhD	Engineering	9
222	D	Female	Habilitated PhD	Life sciences	10
223	D	Female	Habilitated PhD	Humanities	7
224	D	Male	PhD	Humanities	10
225	D	Male	PhD	Natural sciences	10
226	D	Female	Habilitated PhD	Natural sciences	10
227	D	Male	Professor	Humanities	9
228	D	Male	Habilitated PhD	Life sciences	7
229	D	Female	PhD student	Social sciences	6

230	D	Female	PhD student	Technical sciences	7
231	E	Female	Habilitated PhD	Social sciences	7
232	E	Female	PhD	Social sciences	11
233	E	Female	PhD student	Technical sciences	6
234	E	Female	PhD	Social sciences	10
235	E	Male	PhD	Natural sciences/Engineering	6
236	E	Female	Habilitated PhD	Natural science	9
237	E	Female	PhD	Natural science	7
238	F	Male	Professor	Technical sciences	15
239	F	Male	Professor	Technical sciences	9
240	D	Male	PhD student	Humanities	7

A3 Examples from interview coding

Conceptualization:

- We write equations together. I give him ideas. (Interview 217)

- He brings ideas. He has a lot of them and they are fascinating. (Interview 213)

Methodology:

- I have all the optical measurements from them. […] We complement each other in research methods that are accessible in our institution. My team can do some microscopic measurements. (Interview 226)

- He gave me some tips when I was taking the measurements like what will match, how I could find something, how it will match. He also taught me one method. (Interview 214)

Investigation:

- His input is significant. I cannot collect samples without him. (Interview 211)

- I usually ask for observations, which I handle myself later on. (Interview 216)

Data analysis:

- They provide me with data and I analyse it. (Interview 216)

- Sometimes I don't have time to look at charts and so forth. I collect the "life" data and ask my colleague to compute it and to check the results. (Interview 219)

Data curation:

- She wanted me to manage the whole process from building a data set to analysing it. (Interview 212)

- I was looking for someone to get the files in order to digitalise them. The files have over 3 million records, 3 million card records so obviously she did not do it herself. (Interview 227)

Software creation:

- He was a specialist in computer stuff. Yes? All things like that, software etc. I managed to do it myself when it was relatively easy. I'm rather a loner – the majority of my work I did alone. I started collaboration before habilitation and I invited him. It was the right choice as I've recently learned. We were invited to take part in an international contest. (Interview 220)

- He did the "programming machinery". Thanks to that our system works. (Interview 227)

Prototype construction:

- They are strong in design. […] in this project. (Interview 221)

Knowledge:

- We talk, you know. Sometimes it is a brain storm. For example, if there is a call for abstract, we consult it. [I think] I don't fit in here, because I have nothing to say in this area. At least I think so. [She says:] Oh, look! Do this and this. And I will focus on something else. We meet half way. Or when we host a conference […] I give her suggestions to analyse something. Maybe she would have interesting results. These are creative, inspiring meetings. (Interview 218)

- He is an older gentlemen with very broad knowledge about the field when it comes to methods, ideas […] one could address. (Interview 225)

Funding acquisition:

- It was my grant and I invited them. Now it is the other way round. They have a grant and I was invited. (Interview 224)

- I paid him something. He analysed things for me. (Interview 217)

Data and other sources:

- It is appreciated to have an extensive collection of samples that could be used in experiments. I mentioned once during a conversation that I had a lot of samples, which could be utilized. (Interview 211)

- It was material of great value that I had access to and the aspects I learned about and investigated as well. (Interview 222)

Equipment:

- So I came here, because I was needed. There was a laboratory but there was no one, who could take it over and do this kind of research. (Interview 205)

- I don't have access here to the equipment I use. I work in conditions like this, so I very much need it. (Interview 203)

Supervision:

- I am a supervisor of Mr. A and Ms. B. (Interview 209)

- He is my PhD student. I met him when he was a student, and then he came to me. (Interview 217)

Drafts:

- If he brings something intellectual, I can't see a reason why I should buy it and assume it is mine. Right? Only because I paid for it. So I added him as a co-author. (Interview 211)

- The things we prepared, the drafts […] We sent them to him. (Interview 226)

Proofreading:

- So with these books, my input is proofreading. (Interview 223)

- So he tells me what he can do, what I should do, what I should correct. (Interview 214)

Contacts in academia:

- If we want to use some equipment and he knows someone who has it, he recommends us. If he wants something, I recommend him as well so he can use it. (Interview 211)

- He promised at once to give me some names, emails, and phone numbers […] and I invited these people. (Interview 223)

- We collaborate internationally very intensively. We do visit to many places. We host people and these are really the top people in our field in the world. I think they would not get it at some other place. (Interview 224)

Contacts outside of academia:

- Thanks to her we have many contacts among practitioners. (Interview 212)

- She gives us mostly a lot of recognition in this environment and her name opens many doors. Thanks to her we found all those people, we would not have found otherwise. She was our bridge to this group. (Interview 224)

Project administration:

- He manages a project. I smile because one has to do some paper work. (Interview 203)

- We finished a project, where she was one of a leaders managing the project. (Interview 209)

Formal administration:

- She is our unit leader. She initiates many projects […]. She informs us about conferences and she encourages us to take part. She initiates research […] (Interview 207)

- Professor is our boss so all financial matters. She is the unite leader. (Interview 208)

Motivation:

- I met him when I was finishing my master thesis. He encouraged me or motivated me to peruse academic career further. (Interview 225)

- I had a student from University A. She did her engineering diploma here. She is very gifted. During one conversation she asked what should she have done later on. I said that she could apply for Master's degree and I would be her supervisor or she could try to go abroad and […] apply for PhDs degree. (Interview 226)

Traits of character:

- She is on the one hand very responsible, she works very well. On the other hand, her work is excellent on the basis of merits. (Interview 215)

- Professor is very communicative person. He is very friendly and his traits of character encourage collaboration. (Interview 221)

Prestige:

- He is the most recognised scientist in the world. He is well known scientist but older. He has some contacts and it helps in being pulled into the scientific world. (Interview 213)

- They have a few good papers with names of those people. It just increases your chances. (Interview 217)

Professional achievements:

- At first I helped him. I gave him some contacts […] but at some point he did it by himself […] without my help, name and so forth. (Interview 213)

- He was aboard for long time so he also brings academic achievements. (Interview 205)

- When we apply for funding, he brings himself. (Interview 206)

Other characteristics:

- I don't know a lot about the issue but she said that she needed someone from the other side of the country to conduct the research in companies so the results could be compared. (Interview 202)

- It was funding for Polish-German partnership. I was looking for a partner from Germany. He is interested in these issues. Since he is one of the most recognised specialists in the field, I wrote to him asking if he would be interested in this funding. (Interview 224)

Career development:

- I do a lot of things to make her my successor […]. I see her as someone, who can take it further, better. She is the youngest in the team and with no seniority but undoubtedly she has her wits gathered and she is the best based on merits. (Interview 227)

- We had the discussion in March and I had due date in May. [He said] You know what? I will send you the syllabus. We have a statistical module at out department and I am sending you information how you can start improving your knowledge. You can learn here for free so you will not have to pay there. (Interview 212)

Other input:

- Our supervisor, PhD, deals with substantial things. He doesn't tackle any technical issues of his students or some other collaborators. He just doesn't do it. He has more important things to do. We do it instead. (Interview 230)

- I bring some fresh ideas. When one has a lot of experience, she expects some things not to work out. When you don't know that it might not work out, you have this innocence, childlikeness. (Interview 214)

Acknowledgments

References

Abramo, Giovanni, Ciriaco Andrea D'Angelo, and Marco Solazzi. 2010. "The Relationship Between Scientists' Research Performance and the Degree of Internationalization of Their Research." *Scientometrics* 86 (3): 629–43.

Albarrán, Pedro, Raquel Carrasco, and Javier Ruiz-Castillo. 2017. "Geographic Mobility and Research Productivity in a Selection of Top World Economics Departments." *Scientometrics* 111 (1): 241–65.

Allen, Thomas J. 1977. *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information Within the R&D Organization*. Cambridge, MA: MIT Press.

Andrade, Helga Bermeo, Ernesto de Los Reyes Lopez, and Tomas Bonavia Martin. 2009. "Dimensions of Scientific Collaboration and Its Contribution to the Academic Research Groups' Scientific Quality." *Research Evaluation* 18 (4): 301–11.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Bazerman, Charles. 1983. "Scientific Writing as a Social Act: A Review of the Literature of the Sociology of Science." In *New Essays in Technical and Scientific Communication: Research, Theory, Practice*, edited by Paul V. Anderson, John R. Brockmann, and Carolyn R. Miller, 156–84. New York: Routledge.

Bearman, Peter. 1997. "Generalized Exchange." *American Journal of Sociology* 102 (5): 1383–1415.

Beaver, Donald DeB. 2001. "Reflections on Scientific Collaboration (and Its Study): Past, Present, and Future." *Scientometrics* 52 (3): 365–77.

Beaver, Donald DeB., and Richard Rosen. 1979. "Studies in Scientific Collaboration Part III. Professionalization and the Natural History of Modern Scientific Co-Authorship." *Scientometrics* 1 (3): 231–45.

Bellotti, Elisa. 2012. "Getting Funded. Multi-Level Network of Physicists in Italy." *Social Networks* 34 (2): 215–29.

Blau, Peter M. 1994. *The Organization of Academic Work*. New York, NY: Routledges.

Bojanowski, Michal, Dominika Czerniawska, and Wojciech Fenrich. 2020. "Academic Collaboration via Resource Contributions: An Egocentric Dataset." *Connections* 40 (1): 25–30.

Bojanowski, Michał. 2017. *Lspline: Linear Splines with Convenient Parametrisations*. https://cran.r-project.org/package=lspline.

Bordons, Maria, M Zulueta, F Romero, and Santos Barrigón. 1999. "Measuring Interdisciplinary Collaboration Within a University: The Effects of the Multidisciplinary Research Programme." *Scientometrics* 46 (3): 383–98.

Bourdieu, Pierre. 1988. *Homo Academicus*. Stanford, CA: Stanford University Press.

Boyer, Ernest L. 1997. *Scholarship Reconsidered: Priorities of the Professoriate*. Hoboken, NJ: Jossey-Bass.

Burt, Ronald S. 1995. *Structural Holes. The Social Structure of Competition*. Cambridge: Harvard University Press.

———. 2004. "Structural Holes and Good Ideas." *American Journal of Sociology* 110 (2): 349–99.

Cheng, Qing, Xin Lu, Zhong Liu, and Jincai Huang. 2015. "Mining Research Trends with Anomaly Detection Models: The Case of Social Computing Research." *Scientometrics* 103 (2): 453–69.

Choi, Seung-Seok, Sung-Hyuk Cha, and Charles C Tappert. 2010. "A Survey of Binary Similarity and Distance Measures." *Journal of Systemics, Cybernetics and Informatics* 8 (1): 43–48.

Coleman, James S. 1994. *Foundations of Social Theory*. Harvard, MA: Harvard University Press.

Collins, Harry M. 1974. "The Tea Set: Tacit Knowledge and Scientific Networks." *Social Studies of Science* 4 (2): 165–85.

Cook, Karen S, ed. 1987. *Social Exchange Theory*. London: Sage.

Corrêa Jr, Edilson A., Filipi N. Silva, Luciano da F. Costa, and Diego R. Amancio. 2017. "Patterns of Authors Contribution in Scientific Manuscripts." *Journal of Informetrics* 11 (2): 498–510.

Crossley, Nick, Elisa Bellotti, Gemma Edwards, Martin G Everett, Johan Koskinen, and Mark Tranmer. 2015. *Social Network Analysis for Ego-Nets: Social Network Analysis for Actor-Centred Networks*. London: Sage.

Csardi, Gabor, and Tamas Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal Complex Systems*, 1695. http://igraph.org.

Ekeh, Peter Palmer. 1974. *Social Exchange Theory: The Two Traditions*. Harvard, MA: Harvard University Press.

Freeman, Richard B, Ina Ganguli, and Raviv Murciano-Goroff. 2015. "Why and Wherefore of Increased Scientific Collaboration." In *The Changing Frontier: Rethinking Science and Innovation Policy*, edited by Adam Jaffe and Benjamin Jones. National Bureau of Economic Research; University of Chicago Press.

Gatfield, Terry. 2005. "An Investigation into Phd Supervisory Management Styles: Development of a Dynamic Conceptual Model and Its Managerial Implications." *Journal of Higher Education Policy and Management* 27 (3): 311–25.

Gibbons, Michael, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott, and Martin Trow. 1994. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: Sage.

Hagstrom, Warren O. 1964. "Traditional and Modern Forms of Scientific Teamwork." *Administrative Science Quarterly*, 241–63.

Hanneman, Robert A, and Mark Riddle. 2005. *Introduction to Social Network Methods*. Riverside, CA: University of California Riverside.

Hildrun, Kretschmer, Pudovkin Alexander, and Stegmann Johannes. 2012. "Research Evaluation. Part II: Gender Effects of Evaluation: Are Men More Productive and More Cited Than Women?" *Scientometrics* 93 (1): 17–30.

Jian, Du, and Tang Xiaoli. 2013. "Perceptions of Author Order Versus Contribution Among Researchers with Different Professional Ranks and the Potential of Harmonic Counts for Encouraging Ethical Co-Authorship Practices." *Scientometrics* 96 (1): 277–95.

Jones, Benjamin F., Stefan Wuchty, and Brian Uzzi. 2008. "Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science." *Science* 322 (5905): 1259–62.

Katz, J Sylvan, and Ben R Martin. 1997. "What Is Research Collaboration?" *Research Policy* 26 (1): 1–18.

Katz, Ralph, and Michael Tushman. 1981. "An Investigation into the Managerial Roles and Career Paths of Gatekeepers and Project Supervisors in a Major R & d Facility." *R&D Management* 11 (3): 103–10.

Knorr-Cetina, Karin. 2009. *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard, MA: Harvard University Press.

Kuwabara, Ko, Jiao Luo, and Oliver Sheldon. 2010. "Multiplex Exchange Relations." *Advances in Group Processes* 27: 239–68.

Latour, Bruno, and Steve Woolgar. 2013. *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

Laudel, Grit. 2001. "Collaboration, Creativity and Rewards: Why and How Scientists Collaborate." *International Journal of Technology Management* 22 (7-8): 762–81.

Lazarowicz-Kowalik, Marta. 2015. "Anatomia Sukcesu W Nauce. Studium Socjologiczne Laureatów Konkursów Fundacji na rzecz Nauki Polskiej." PhD thesis, Warsaw: University of Warsaw.

Lazega, Emmanuel, Marie-Therese Jourda, Lise Mounier, and Rafael Stofer. 2008. "Catching up with Big Fish in the Big Pond? Multi-Level Network Analysis Through Linked Design." *Social Networks* 30 (2): 159–76.

Lazega, Emmanuel, and Philippa E. Pattison. 1999. "Multiplexity, Generalized Exchange and Cooperation in Organizations: A Case Study." *Social Networks* 21 (1): 67–90.

Leahey, Erin. 2016. "From Sole Investigator to Team Scientist: Trends in the Practice and Study of Research Collaboration." *Annual Review of Sociology* 42: 81–100.

Levandowsky, Michael, and David Winter. 1971. "Distance Between Sets." *Nature* 234 (5323): 34–35.

Lewis, Jenny M, Sandy Ross, and Thomas Holden. 2012. "The How and Why of Academic Collaboration: Disciplinary Differences and Policy Implications." *Higher Education* 64 (5): 693–708.

Lopaciuk-Gonczaryk, Beata. 2016. "Collaboration Strategies for Publishing Articles in International Journals–a Study of Polish Scientists in Economics." *Social Networks* 44: 50–63.

Merton, Robert K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, IL: University of Chicago Press.

Merton, Robert K, and Elinor Barber. 2006. *The Travels and Adventures of Serendipity: A Study in Sociological Semantics and the Sociology of Science*. Princeton, NJ: Princeton University Press.

Mirowski, Philip. 2011. *Science-Mart*. Harvard, MA: Harvard University Press.

Moody, James. 2004. "The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999." *American Sociological Review* 69 (2): 213–38.

Nobre, Gustavo Cattelan, and Elaine Tavares. 2017. "Scientific Literature Analysis on Big Data and Internet of Things Applications on Circular Economy: A Bibliometric Study." *Scientometrics* 111 (1): 463–92.

Pattison, Philippa, and Stanley Wasserman. 1999. "Logit Models and Logistic Regressions for Social Networks: II. Multivariate Relations." *British Journal of Mathematical and Statistical Psychology* 52 (2): 169–93.

Perkin, Harold. 2007. "History of Universities." In *International Handbook of Higher Education*, edited by James J. F. Forest and Philip G. Altbach, 159–205. New York, NY: Springer.

Podolny, Joel M, and James N Baron. 1997. "Resources and Relationships: Social Networks and Mobility in the Workplace." *American Sociological Review* 62 (5): 673–93.

Popper, Karl R. 1972. *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press Oxford.

Powell, Walter W, and Jason Owen-Smith. 2012. "An Open Elite: Arbiters, Catalysts, or Gatekeepers in the Dynamics of Industry Evolution?" In *The Emergence of Organizations and Markets*, edited by John F Padgett and Walter W Powell, 466–95. Princeton, NJ: Princeton University Press.

Qin, Jian. 1994. "An Investigation of Research Collaboration in the Sciences Through the Philosophical Transactions 1901–1991." *Scientometrics* 29 (2): 219–38.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Resnik, David B. 2006. *The Price of Truth: How Money Affects the Norms of Science*. Oxford: Oxford University Press.

Ridley, Matt. 2011. *The Rational Optimist: How Prosperity Evolves*. New York: Harper Perennial.

Schummer, Joachim. 2004. "Multidisciplinarity, Interdisciplinarity, and Patterns of Research Collaboration in Nanoscience and Nanotechnology." *Scientometrics* 59 (3): 425–65.

Schweinberger, Michael. 2011. "Instability, Sensitivity, and Degeneracy of Discrete Exponential Families." *Journal of the American Statistical Association* 106 (496): 1361–70.

Siegel, Donald S, David A Waldman, Leanne E Atwater, and Albert N Link. 2003. "Commercial Knowledge Transfers from Universities to Firms: Improving the Effectiveness of University–Industry Collaboration." *The Journal of High Technology Management Research* 14 (1): 111–33.

Simmel, Georg. 1972. *Sociology: On Individuality and Social Forms*. Chicago, IL: University of Chicago Press.

Sonnenwald, Diane H. 2007. "Scientific Collaboration." *Annual Review of Information Science and Technology* 41 (1): 643–81.

Terekhov, Alexander I. 2017. "Bibliometric Spectroscopy of Russia's Nanotechnology: 2000–2014." *Scientometrics* 110: 1217–42.

Tiwana, Amrit, and Mark Keil. 2007. "Does Peripheral Knowledge Complement Control? An Empirical Test in Technology Outsourcing Alliances." *Strategic Management Journal* 28 (6): 623–34.

Wagner, Caroline S. 2009. *The New Invisible College: Science for Development*. Washington, DC: Brookings Institution Press.

Wagner, Izabela. 2011. *Becoming Transnational Professional: Kariery I Mobilność Polskich Elit Naukowych*. Warsaw: Wydawnictwo Naukowe Scholar.

Wang, Lili, Ad Notten, and Alexandru Surpatean. 2013. "Interdisciplinarity of Nano Research Fields: A Keyword Mining Approach." *Scientometrics* 94 (3): 877–92.

Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Whitley, Richard. 2000. *The Intellectual and Social Organization of the Sciences*. Oxford: Oxford University Press.

Youtie, Jan, and Barry Bozeman. 2014. "Social Dynamics of Research Collaboration: Norms, Practices, and Ethical Issues in Determining Co-Authorship Rights." *Scientometrics* 101 (2): 953–62.

Zaheer, Akbar, and Geoffrey G Bell. 2005. "Benefiting from Network Position: Firm Capabilities, Structural Holes, and Performance." *Strategic Management Journal* 26 (9): 809–25.

Zuckerman, Harriet. 1967. "Nobel Laureates in Science: Patterns of Productivity, Collaboration, and Authorship." *American Sociological Review* 32 (3): 391–403.