Research Policy 45 (2016) 68–80

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Knowledge creation in collaboration networks: Effects of tie configuration

Jian Wang[a,b,c,*]

[a] Georgia Institute of Technology, USA
[b] University of Leuven, Belgium
[c] Institute for Research Information and Quality Assurance (iFQ), Germany

ARTICLE INFO

Article history:
Received 11 November 2013
Received in revised form 12 June 2015
Accepted 22 September 2015
Available online 22 October 2015

Keywords:
Knowledge creation
Organization of science
Research collaboration
Egocentric networks
Tie strength
Tie configuration

ABSTRACT

This paper studies the relationship between egocentric collaboration networks and knowledge creation at the individual level. For egocentric networks we focus on the characteristics of tie strength and tie configuration, and knowledge creation is assessed by the number of citations. Using a panel of 1042 American scientists in five disciplines and fixed effects models, we found an inverted *U*-shaped relationship between network average tie strength and citation impact, because an increase in tie strength on the one hand facilitates the collaborative knowledge creation process and on the other hand decreases cognitive diversity. In addition, when the network average tie strength is high, a more skewed network performs better because it still has a "healthy" mixture of weak and strong ties and a balance between exploration and exploitation. Furthermore, the tie strength skewness moderates the effect of network average tie strength: both the initial positive effect and the later negative effect of an increase in tie strength are smaller in a more skewed network than in a less skewed one.

1. Introduction

In history, scientists were often depicted as lonely wolves, and prominent discoveries were often credited to solitary authors. However, the production of science is increasingly collaborative (Adams et al., 2005; Hicks and Katz, 1996; Price, 1986; Wuchty et al., 2007). By pooling together different expertise and perspectives, collaboration contributes to cross-fertilization of ideas and enables combining different pieces of knowledge to create something novel and useful (Katz and Martin, 1997; Melin, 2000; Page, 2007). The prevalence of collaboration in science has driven science studies to expand from lab benches to collaborative settings at a larger scale (Chompalov et al., 2002; Cummings and Kiesler, 2005; Finholt and Olson, 1997; Shrum et al., 2001) and sparked vigorous studies of collaborative teams (Cummings et al., 2013; Hemlin et al., 2013; Lee et al., 2015; Levine and Moreland, 2004; Murayama et al., 2015; Walsh and Lee, 2015) and networks (Börner et al., 2004; Guimera et al., 2005; Newman, 2004; Sun et al., 2013) in science.

This study investigates the relationship between collaboration networks and knowledge creation at the individual level. Dynamic egocentric collaboration networks are viewed as the venue where scientific knowledge is produced, and the characteristics of egocentric networks shape the process of knowledge creation within the network, which in turn affects the impact or usefulness of the knowledge created from the network. At a fundamental level, knowledge resides within and is created by individuals (Nonaka, 1994). However, the creation of knowledge is also a social process (Latour and Woolgar, 1986; Nonaka, 1994). Therefore, it is important to place a creative individual within a network of interpersonal relationships for a better understanding of knowledge creation (Simonton, 1984). Previous studies have extensively investigated the effect of collaboration networks on research performance at the individual level (Abbasi et al., 2011; Gonzalez-Brambila et al., 2013; Li et al., 2013; McFadyen and Cannella, 2004; McFadyen et al., 2009). These studies typically adopt a *social capital* perspective, where a scientist's egocentric network or his/her position in the global network represents his/her social capital, and social capital affects research performance indirectly, through serving as an input for current knowledge creation. However, this paper studies collaboration networks as organizations of knowledge creation and focuses on how the current network affects knowledge creation directly, via its effect on the creative process and resource mobilization. Specifically, this paper focuses on the effect of tie strength and tie configuration on citation impact at the individual level.

* Correspondence to: Naamsestraat 69, bus 3535, 3000 Leuven, Belgium.
E-mail address: jian.wang@kuleuven.be

http://dx.doi.org/10.1016/j.respol.2015.09.003
0048-7333/

This paper makes the following theoretical contributions. First, it adds to the organization of science literature, studying egocentric collaboration networks as organizations for science production. Second, it explores tie configuration within collaboration networks and contributes to the development of a network theory beyond a simple dichotomy between strong and weak ties.

The rest of the paper is organized as follows. First, we briefly review the literature on collaborative teams and networks in science and discuss the motivation for studying egocentric networks. Second, we develop hypotheses concerning the effect of tie strength and tie configuration on knowledge creation, drawing literature of science studies, social networks, organization theory, and organizational behavior. We use a panel dataset with both survey and bibliometric information for 1042 American scientists in five disciplines (biology, chemistry, computer science, earth and atmospheric sciences, and electrical engineering). We incorporate (1) individual fixed effects to account for unobserved time invariant individual heterogeneity and (2) career age and prior performance to control for time variant individual differences. We found (1) an inverted *U*-shaped relationship between network average tie strength and citation impact, (2) a positive effect of the skewness of tie strength distribution on citation impact, when the network average tie strength is high, and (3) that the effect of network average tie strength is moderated by the level of skewness. We also discuss the implications of these findings.

2. Knowledge creation in science

2.1. Collaborative science: teams and networks

Scientific knowledge is increasingly created collaboratively, as reflected in the rising share of coauthored papers and the growing size of collaborative teams (Adams et al., 2005; Hicks and Katz, 1996; Price, 1986; Wuchty et al., 2007). While earlier science studies focus on individual traits and laboratory settings (Latour and Woolgar, 1986; Simonton, 1999; Zuckerman, 1967), the prevalence of collaboration in science calls for studying the organization of collaborative science, and researchers have extended science studies and laboratory ethnographies from lab benches to collaborative settings at a larger scale (Chompalov et al., 2002; Cummings and Kiesler, 2005; Finholt and Olson, 1997; Shrum et al., 2001).

Recently, there emerges a new body of literature labeled as *science of team science*, which brings in insights from the psychology literature on small groups and the sociology literature on work organizations to study the team production of science (Falk-Krzesinski et al., 2010; Fiore, 2008; Stokols et al., 2008). For example, previous research has investigated the group process (Levine and Moreland, 2004), leadership (Hemlin et al., 2013), and bureaucratization (Walsh and Lee, 2015) in scientific teams, as well as the effects of team characteristics on team productivity (Cummings et al., 2013), creativity (Lee et al., 2015), and the quality of team product (Murayama et al., 2015).

Besides scientific teams, collaboration networks have also been extensively studied at the system level (i.e., all sciences or a particular scientific field), covering topics such as patterns of collaboration networks (Guimera et al., 2005; Newman, 2004), evolution of scientific networks and mechanisms underlying the process (Börner et al., 2004; Sun et al., 2013), and the network effects on research performance (Guimera et al., 2005).

2.2. Individuals and egocentric networks

At the individual level, egocentric network or individual's position in the global network have also been explored to explain the productivity or creativity of individual scientists (Abbasi et al., 2011; Klenk et al., 2010; Li et al., 2013; McFadyen and Cannella, 2004; McFadyen et al., 2009). The individual-level network studies typically adopt a *social capital* perspective; a scientist's egocentric network or position in the global network represents his/her social capital, which in turn affects his/her performance the same way as intellectual and other capital. From this social capital perspective, collaboration networks affect individual performance indirectly through providing social capital as an input but not directly by serving as a work organization. This nuance is more evident when scrutinizing empirical strategies adopted in these studies, which measure social capital based on collaboration networks in previous years and estimate its effect on individual performance in the current year (McFadyen and Cannella, 2004; McFadyen et al., 2009). Preceding network provides social capital as an input for current knowledge creation, but the current network, which is directly responsible for the current science production, is ignored. Furthermore, focusing on the previous but not the current network does not explain how social capital is mobilized for current knowledge creation (Lin, 1999, 2001). Different from the social capital perspective, this paper studies the current collaboration networks as work organizations bearing direct effects on current knowledge creation.

The concept of social capital is evoked as a bridge between egocentric networks and individual performance, presumably because individuals or egocentric networks are not recognized as legitimate forms of organization for scientific production, while teams are. Accordingly, the distinction between our egocentric network approach and the team approach is twofold: in the egocentric network approach, (1) individual is still a relevant unit of analysis for studying knowledge creation in science and (2) egocentric collaboration network is also a legitimate form of organization for knowledge creation.

Science is increasingly performed in teams. However, at a fundamental level, knowledge still resides within and is created by individuals (Nonaka, 1994). Studies of group creativity also emphasize the importance of individuals' abilities, previous experiences, and other resources that they carry with them (Amabile, 1983; Ford, 1996; Woodman et al., 1993). Therefore, research evaluation at the individual level is still a relevant practice, and a scientist endowed with a higher level of intellectual or social capital can contribute his advantages to all his collaborative teams and achieve better performance across all his collaborations.

In addition, although science is increasingly performed in teams, knowledge creation within a team also depends on activities outside the team. One distinct feature of modern science, compared with other systems of work organization, is its autonomy and self-governance (Whitley, 2000). As a result, scientific teams are extremely fluid, with ill-defined and constantly changing boundaries (Borgman, 2007). This fluidness of scientific teams is also reflected in the difficulty of determining authorships (Haeussler and Sauermann, 2013; Laudel, 2002). More importantly, the fluidness of collaborative teams is associated with the interdependence between teams connected by common members. Scientists often participate in multiple teams simultaneously, and these teams may share several common members and also similar research agendas. Under such circumstance, knowledge spillovers across teams are likely to take place. For example, Tang and Hu (2013) showed that scholars pick up new research lines from their international collaborators and further pursue them in their domestic collaborations, and Wang and Hicks (2015) demonstrated knowledge spillovers from a scientist's new collaborators to his/her other teams not involving these new collaborators. Since knowledge creation at the team level also depends on external activities, it is also important to study the open and dynamic egocentric networks, in addition to the closed collaborative teams, in order to better understand knowledge creation in science.

Because of the fluidness of teams and the interdependence between them, the organization of scientific collaboration may be described by a *garbage can model* (Cohen et al., 1972). There are streams of problems, expertise, and collaborators in the network. Problems are searching for relevant expertise, expertise is searching for problems, and collaborators are searching for common research interests (i.e., problems) and complementary expertise. A collaborative team emerges when these three streams converge. However, the emergence of a team is not the end of the story. Instead, the team still interacts with these three streams and co-evolves with them. Different networks may have different problem, expertise, and collaborator streams, and these differences in turn lead to variance in final creative outcomes. Furthermore, network structures determine the opportunities and constraints for (1) assembling differentiated but interdependent teams and (2) balancing explorative and exploitative research activities, and such opportunities and constraints affect knowledge creation at the individual level across teams. Therefore a network approach, which investigates sources of knowledge creation in dynamic networks beyond team boundaries, is meaningful for understanding the production of science.

3. Collaboration networks and knowledge creation

3.1. Tie strength

To study the relationship between egocentric collaboration network and knowledge creation, this paper only investigates direct ties, for two reasons: First, direct ties play a much more central role in knowledge creation (McFadyen and Cannella, 2004). While previous literature suggests the importance of indirect ties for knowledge transfer, McFadyen and Cannella (2004) suggested that direct ties are "absolutely central" for knowledge creation. Second, because of their direct and central role in collaborative knowledge creation, direct ties allow for studying collective action and resource mobilization for knowledge creation. Among various characteristics of the direct tie, this paper further focuses on the strength of ties, for two main reasons: First, there is a rich but divergent body of literature on the strength of tie, calling for testing and reconciling competing theories. Second, the strength of tie only depends on two collaborators but not others, and is therefore easier to be managed by individual scientists, which allows drawing more direct policy and managerial implications.

Before making the network level predictions, we will first discuss the effect of tie strength at the dyadic level. Granovetter (1973) defined the strength of tie as "a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie" (p. 1361). Weak ties are more likely to provide non-redundant information (Burt, 1992; Granovetter, 1973; Uzzi, 1996; Uzzi and Spiro, 2005). People bonded by strong ties are more likely to be similar to each other and connected with similar others. Therefore, information obtained from such networks tends to be redundant. In contrast, weak ties are more likely to bridge *structural holes* between communities that are otherwise unconnected and provide access to information and resources beyond those available in one's own social circles.

Because of the access to non-redundant information, weak-tie-collaborations are more likely to generate novel and useful ideas. One important benefit from collaboration is the cross-fertilization of ideas by pooling together different expertise and perspectives (Hudson, 1996; Katz and Martin, 1997; Melin, 2000; Page, 2007). Many scholars have suggested that one important source of novelty is making unusual but fruitful recombination of preexisting components, such as ideas, principals, and devices (Kuhn, 1970; Mednick, 1962; Nelson and Winter, 1982; Schumpeter, 1939). Therefore, exposure to diverse knowledge and perspectives increases the chance of generating novel ideas. Furthermore, according to Page (2007), scientific research can be understood as a search process in a complex problem space for the best solution, and cognitive diversity contributes to a broader and more thorough search and consequently a higher possibility of finding the global optimum instead of being trapped at a local optimum.

However, the association between tie strength and knowledge creation is not so straightforward, because many other factors may affect the knowledge creation process. Weak ties have low *cognitive capital* (i.e., shared knowledge and understanding) and *relational capital* (i.e., trust, norm, and obligation)(Nahapiet and Ghoshal, 1998). Because of the lack of a common knowledge base, collaborators bonded by weak ties may find significant communicational and epistemological problems in exchanging and integrating different perspective, ideas, and data (Edwards et al., 2011; Kuhn, 1970; Star and Griesemer, 1989). In addition, the lack of mutual trust, obligation, and norm in the collaborative tie may increase opportunistic behavior and impede coordinated actions (Krackhardt, 1992; Lin and Ensel, 1989; Obstfeld, 2005; Podolny and Baron, 1997; Uzzi, 1996). As the strength of tie increases, so does the *cognitive* and *relational capital*, and as a result, the collaboration has a more effective knowledge creation process. Many empirical studies have also shown the advantage of strong ties for knowledge transfer (Hansen, 1999; Reagans and McEvily, 2003) and knowledge creation (McFadyen and Cannella, 2004; McFadyen et al., 2009; Tortoriello and Krackhardt, 2010; Walsh and Maloney, 2002).

However, this effect may turn negative when the strength of tie is too strong. First, cognition of the collaborators becomes similar, so the potential of generating novel and useful ideas is diminished (McFadyen and Cannella, 2004; Uzzi, 1997). Second, common collaboration experience gives birth to shared cognitive structures/routines that govern behavior of the collaborators (Granovetter, 1985). Skilton and Dooley (2010) argued that an enduring *mental model* would emerge from repeated collaboration and would shape not only the way that individuals explain, predict, and describe events, but also the way that the team differentiates roles among members. Furthermore, the mental model is inert and constrains subsequent collaboration. Thereby, repeated collaboration is less able to generate novel ideas. Empirical studies have also found a negative association between repeated collaboration and creativity (Guimera et al., 2005; Porac et al., 2004).

In summary, at the dyadic level, there is an inverted *U*-shaped relationship between tie strength and knowledge creation, specifically, the effect of tie strength is initially positive and turns negative after a threshold. How to translate this to the network-level analysis? If we assume that ties in the same egocentric network are relatively homogeneous, then we can use the network average tie strength to indicate the overall tie strength of the whole network, and then the tie strength effect at the network level is a simple aggregation of effects at the dyadic level. Therefore,

Hypothesis 1. There is an inverted *U*-shaped relationship between network average tie strength and knowledge creation, that is, the effect of network average tie strength is initially positive and turns negative after a threshold.

3.2. Tie configuration: skewness

However, this tie homogeneity assumption might be problematic. Uzzi (1996) found that firms maintain both embedded and arms-length ties, suggesting that the configuration of ties, rather than a simple dichotomy between strong-tie-network and weak-tie-network, should be investigated. Uzzi (1996) used a Herfindahl-type indicator to measure the dominance of strong ties in a network. In addition, some other studies defined a boundary between strong and weak ties, counted the number of weak and strong ties separately, and evaluated their respective effects (Tortoriello and Krackhardt, 2010; Walsh and Maloney, 2002).

This paper investigates tie configuration, specifically the skewness of the tie strength distribution. It is common that a scientist simultaneously has a small group of colleagues with intense interactions on the one hand and a number of loose contacts on the other. In other words, the tie strength distribution of an egocentric network tends to be skewed and different from a normal distribution, so that using the network average tie strength may hide distinct tie configuration characteristics. For example, out of two networks with the same average tie strength, one may have all ties of medium strength, while the other has half strong and half weak ties.

Empirically, egocentric collaboration networks have (positively) skewed tie strength distributions, with a long tail on the right side and the bulk of the values lie to the left of the mean. The small number of strong ties may reflect the constraint of carrying capacity. Scientists have limited amount of time and energy, but maintaining strong relations is costly. Therefore, having too many strong collaborative ties is simply infeasible or inefficient (McFadyen and Cannella, 2004; Perry-Smith and Shalley, 2003). On the other hand, the large number of weak ties may reflect scientists' broad search for diverse and complementary knowledge. A large number of weak ties may augment the scientist's knowledge base about the research domain and also enhance his/her *absorptive capacity* (Perry-Smith and Shalley, 2003; Reagans and McEvily, 2003; Simonton, 1999).

In addition, there is another implicit assumption underlying the simple aggregation approach (i.e., network-level effect is an aggregation of dyadic-level effects): There is no interaction effect between dyads. However, individuals can bring in lessons learned from previous team experiences to new situations (Ancona, 1990; Gino et al., 2010; Reagans et al., 2005), and there are significant knowledge spillover from one team to another connected by shared members (Tang and Hu, 2013; Wang and Hicks, 2015). Therefore, many weak ties augment the knowledge base and enhance the absorptive capacity, which benefits not only weak-tie-collaborations themselves but also strong-tie-collaborations in the same network. In other words, the benefit of knowledge diversity gained from weak ties can also be transferred to other collaborations.

A useful perspective to understand the effect of tie configuration on knowledge creation is the *exploration* vs. *exploitation* literature. March (1991) distinguished between exploitation and exploration in organizational learning: "The essence of exploitation is the refinement and extension of existing competencies, technologies, and paradigms... The essence of exploration is experimentation with new alternatives" (p. 85). This seminal work has trigged a large volume of studies of the tradeoff between exploration and exploitation (e.g., Fleming, 2001; Katila and Ahuja, 2002; Wong, 2004). March (1991) also suggested that both exploration and exploitation are important for the success of the organization, and recent studies suggested that one way of balancing exploration and exploitation is via *ambidexterity*, namely synchronous pursuit of exploration and exploitation through highly differentiated but weakly integrated subunits or individuals, each of which is specialized in either exploration or exploitation (Benner and Tushman, 2003; Fang et al., 2010; Lazer and Friedman, 2007).

In addition, the organizational learning literature also suggests that newcomers are more likely to conduct exploration, while old-timers are more likely to do exploitation (Gupta et al., 2006; March, 1991; Perretti and Negro, 2006). Newcomers are important sources of innovation for an organization because they (1) are more likely to bring in different knowledge and perspectives which are not yet shared in the organization (Gupta et al., 2006; March, 1991; Perretti and Negro, 2006) and (2) loosen the *mental model*, which emerges from previous collaboration experiences and constraints current collaboration (Skilton and Dooley, 2010). Therefore, a balance between exploration and exploitation within a collaborative team can be achieved through a mixture of old-timers and newcomers, and such mixture contributes to better team innovative performance (Chen, 2005; Perretti and Negro, 2006).

Similarly, an egocentric collaboration network with a mixture of weak and strong ties can maintain a balance between the lack of weak ties an therefore can achieve better performance in knowledge creation. Specifically, when the network average tie strength is high, a less skewed network suffers from the lack of weak ties and exploration. In contrast, a more skewed network still has a number of weak ties. Furthermore, gains from explorative activities in weak-tie-collaborations can be transferred to other collaborations and contribute to better knowledge creation of the whole egocentric collaboration network. In other words, when the network average tie strength is high, a more skewed network performs better in knowledge creation because it still has a "healthy" mixture of strong and weak ties and therefore can maintain a balance between exploration and exploitation. When the network average tie strength is low, however, the effect of tie strength skewness is unclear. Therefore, we hypothesize that

Hypothesis 2. A more skewed network performs better in knowledge creation, compared with a less skewed network, when the network average tie strength is high.

Furthermore, ties strength skewness moderates the effect of network average tie strength. Given the heterogeneity of tie strength in a skewed network, the average tie strength is not an accurate indication of the overall tie strength of the whole network, so a more skewed network is less sensitive to the changes in network average tie strength. For example, when the network average tie strength is low, a more skewed network already has some very strong ties. Under such circumstance, if we increase the strength of each tie, the network does benefit from the increase in those very weak ties but not from the increase in those already very strong ones. Therefore, the aggregated positive effect is smaller in a more skewed network than in a less skewed one. Similarly, when the network average tie strength is high, a more skewed network still has many weak ties. Under such circumstance, if we increase the strength of each tie, the network does suffer from the increase in those very strong ties but not as much from the increases in those still very weak ones. Therefore, the aggregated negative effect is also smaller in a more skewed network.

Hypothesis 3. Tie strength skewness moderates the effect of network average tie strength, specifically, both the initial positive effect and the later negative effect caused by an increase in network average tie strength are smaller in a more skewed network than in a less skewed one.

4. Data and methods

4.1. Data

A panel consisting of both survey and bibliometric data for 1042 American scientists (i.e., egos) with 6998 observations (i.e., ego-year) are used for testing our hypotheses. The sample of scientists came from a survey funded by the United States National Science Foundation (NSF). The survey was conducted in 2007 on 3677 stratified randomly sampled American scientists in six disciplines: biology (BIOL), chemistry (CHEM), computer science (CS), earth and atmospheric sciences (EAS), electrical engineering (EE), and physics

(PHYS). The random sample was stratified by sex, rank, and discipline, from the population of academic scientists and engineers in these six disciplines in Carnegie-designated Research I universities (150 universities). The population was constructed by manually retrieving information from the websites of the relevant departments or university directories. Of the 1774 completed surveys, 176 were removed because of ineligible rank or discipline, resulting in a final total sample size of 1598. The overall response rate of the survey, calculated using the RR2 method of the American Association for Public Opinion Research (AAPOR) is 45.8%, and the weighted response rate is 43.0%. The responses' distribution of sex, rank, and discipline are similar to the survey population.

Life-time publications for the survey respondents were subsequently retrieved from Thomson Reuters Web of Science (WoS). The collection of the publication data firstly required an author name and affiliation match, and then cleaned out false papers of homonymous authors following the procedure documented in Wang et al. (2012). Coauthor names were also cleaned and disambiguated to identify unique collaborators. The retrieval of publication records from WoS for each ego was last updated in May 2011. Because of the complex publishing practice in the field of physics (i.e., papers often have hundreds of authors), publication data for physicists were excluded from the data cleaning process, leaving 1323 scientists in the remaining five disciplines available for analysis. In addition, because (1) at the time of final retrieval of publication records for each ego, the database coverage for 2010 was still incomplete, and (2) for papers published in and after, but not before, 1980, we have complete citation information for each paper till the end of 2013. We kept only papers published between 1980 and 2009 for analysis. Out of the 1323 scientists, 1310 published 41,964 journal articles in total.

In addition, papers with a large number of authors may cause problems for this study. The theory of this paper relies on substantial interpersonal interactions in collaborative ties. However, some papers with hundreds of authors were observed in the data, and it is unclear whether coauthors on this type of papers actually had substantial interpersonal interactions when collaborating on these papers. In addition, theoretically, the *hyper-authorship* (Cronin, 2001) is beyond the scope of this study. Therefore, papers with more than 15 authors were not used for constructing the variables, leaving 1310 egos with 41,364 journal articles. Two other thresholds, 10 and 29 were also tried and did not change the conclusions. In addition, we also tried an alternative treatment, where an ego was completely excluded from the analysis if he/she has paper(s) with more than 15 authors. This treatment also yielded consistent results.

Reported regression results are based on a dataset constructed from these 41,364 journal articles. Among them 7 are shared by three egos, and 559 by two egos, while 40,798 (98.6%) involve only one ego in our sample. Therefore, the sampled egos are largely unconnected to each other, and different egos can be treated as independent observations.

A panel dataset for these egos was constructed for the regression analysis, where one observation is one ego in one year. Because the tie strength measure for year t is measured as the number of coauthored papers between year $t-4$ and t, the first four years of observations of each ego were excluded from regressions. In total, we have 6998 observations of 1042 egos.

4.2. Measures

4.2.1. Knowledge creation

To assess knowledge creation in year t, we used the total number of citations received by an ego's papers published in year t. While there are concerns about the validity of citations as a measure of impact (Bornmann and Daniel, 2008; De Bellis, 2009; Martin and Irvine, 1983), citation-based metrics have been widely used in science studies and research evaluations. From Merton's perspective, citation serves as an elementary building block of the scientific reward system. For a paper, the acceptance for publishing indicates the acknowledgment of its original contributions to science from peers in the field. Being cited further indicates the peer-recognition of its value and its impact on the scientific community (De Bellis, 2009; Merton, 1973; Wang, 2014). Empirically, citations have been found positively related to winning Nobel Prize, peer recognition, and novelty (Cole and Cole, 1967; Garfield, 1973; Newman and Cooper, 1993; Uzzi et al., 2013). Therefore, we used citation counts as a measure for the impact or usefulness of a scientist's research. A five-year citation time window was used to count citations, that is, for a paper published in year t, its citations between year t and $t+4$ were counted.

Several treatments were undertaken to address potential issues in the use of citation counts. Martin and Irvine (1983) provided a thorough discussion on problems of using citations for research evaluation. First, citation aging pattern differs across papers; many highly cited papers takes a long time to establish themselves as elite papers, while many others have early citation peaks (Garfield, 1980; Glänzel et al., 2003; Van Raan, 2004). Therefore, a sufficient time window is needed to give reliable citation counts. According to Wang's (2013) calculation on the whole WoS database, the Spearman correlations between five-year citation counts and 31-year citation counts are: 0.810, 0.906, 0.852, 0.888, and 0.792 in fields of biology, biomedical research, chemistry, earth and space, and engineering, respectively. The correlations are sufficiently high for this study. The second issue pertains to "obliteration by incorporation," that is, some fundamental papers become so widely known and integrated into the daily work in the field that they no longer need to be cited explicitly (Latour and Woolgar, 1986; Merton, 1983). This issue does not cause problems in this study because only recent publications were studied. Third, citations are incomparable between fields because of field differences in the volume of publications and the norms of referencing (Moed et al., 1985). Fourth, there is a "halo effect" in citation; a paper of a prestigious author or institution tends to be evaluated more highly and gets more citations than another comparable paper of a less prestigious author or institution (Wang, 2014). To address the third and fourth issue, our regression strategy incorporates ego fixed effects and estimates within-ego effects. Therefore, our analysis does not make between-field or between-individual comparisons. The fifth issue is about self-citations. Some productive scientists may actively cite their own papers, but self-citations do not reflect the recognition from the community (Aksnes, 2003; Glänzel et al., 2006). Therefore, non-self-citation counts were also tried, which gave similar results.

4.2.2. Collaboration network

For an ego, his/her coauthors in year t were identified to construct his/her collaboration network for knowledge creation in year t. *Network size* is the number of coauthors. Previous literature suggested an inverted *U*-shaped relationship between network size and knowledge creation (Lavie and Drori, 2012; McFadyen and Cannella, 2004), because an increase in network size on the one hand increases cognitive diversity but on the other hand may distract scientists from other more productive activities. Therefore, both *network size* and *network size*^{2} were included in the regression models.

Tie strength is operationalized as the frequency of collaboration in a five-year time window, including the current and preceding four years. Specifically, in year t, the strength of a tie between one ego and a coauthor was measured as the number of coauthored papers between them in the period from $t-4$ to t. At the egocentric network level, *tie strength avg* was calculated as the network average tie strength. *Skewness* of the tie strength distribution was calculated using the following formula:

$$Skewness = \frac{n}{(n-1)(n-2)} \cdot \frac{\sum_1^n (x_i - \bar{x})^3}{((1/n-1) \cdot \sum_1^n (x_i - \bar{x})^2)^{3/2}}$$

where n is the number of ties in an egocentric network, and x_i is the tie strength for the i-th tie in the egocentric network. In addition, two other popular skewness formulas, i.e., $(1/n) \cdot \sum_1^n (x_i - \bar{x})^3/((1/n) \cdot \sum_1^2 (x_i - \bar{x})^2)^{3/2}$ and $(1/n) \cdot \sum_1^n (x_i - \bar{x})^3/((1/(n-1)) \cdot \sum_1^n (x_i - \bar{x})^2)^{3/2}$, were also tried, and all three skewness measures were highly correlated and yielded similar regression results.

4.2.3. Control variables

The number of papers published in year t (*pubs*) was controlled, given that more papers may result in higher total number of citations. In addition, having *ln*(*citations*) and *ln*(*pubs*) on the two sides of the equation is the appropriate way of modeling the power law scaling relationship between them (Katz, 1999, 2000).

We incorporated ego fixed effects to control for unobserved and time-invariant individual heterogeneities. In addition, we adopted the following variables to control for time-variant individual characteristics. Both the number of citations (the dependent variables) and the collaboration network (the explanatory variables) are likely to be correlated with the ego's previous performance. For example, successful history breeds further success and also attracts new collaborators. Therefore, we control for ego's citation performance in year $t-1$ (*citations lag*).

Furthermore, age and experience are important factors of research collaboration and performance (Lee and Bozeman, 2005; Levin and Stephan, 1991; van Rijnsoever and Hessels, 2011; van Rijnsoever et al., 2008), so *career age* was included to control for both effects, following Lee and Bozeman (2005). To account for the nonlinear trajectory of research performance over the life cycle (Cole, 1979; Levin and Stephan, 1991), both *career age* and *career age*^{2} were included in the regression models. The list of variables is provided in Table 1.

4.3. Methods

The total number of citations is a non-negative count variable with over-dispersion, so the Poisson model with robust standard errors was adopted, following previous literature (Hall and Ziedonis, 2001; Hottenrott and Lopes-Bento, 2015; Somaya et al., 2007; Wang et al., 2015). An alternative is the negative binomial model. However, because the Poisson model is in the linear exponential class, Gourieroux et al. (1984) have shown that the Poisson estimator and the robust standard errors are consistent so long as the mean is correctly specified even under misspecification of the distribution, but the negative binomial estimator is inconsistent if the true underlying distribution is not negative binomial. Therefore, we report the Poisson model with robust standard errors in the paper, and use negative binomial models as a robustness check.

Furthermore, we incorporated individual fixed effects to account for unobserved and time-invariant individual heterogeneities, so that within-ego effects were estimated. Such fixed effects Poisson models can be fitted by conditioning out the individual fixed effects (Hausman et al., 1984). Specifically, we used the *xtpoisson* command in *STATA* (StataCorp, 2013c), which implements the formula as presented in Wooldridge (1999).

Hausman et al. (1984) also developed a conditional maximum likelihood strategy for negative binomial models, which is implemented in the *xtnbreg* function in *STATA* (StataCorp, 2013a). However, this method allows for individual-specific variation in the dispersion parameter rather than in the conditional mean, and therefore does not qualify as a true fixed effects method (Allison and Waterman, 2002; Greene, 2005; Guimarães, 2008). To the best of our knowledge, we are not aware of any statistics software providing a true fixed effect negative binomial solution. For testing robustness of our findings, we also fitted the *xtnbreg* models and got consistent results.

Poisson models predict the natural logarithm of the dependent variable with a linear combination of the independent variables. Therefore, the natural logarithm of *citations lag* was used in the regression model. In addition, Katz (1999) suggested a power law scaling relationship between the number of citations and publications and suggested to use a log-log model for data analysis, so the number of publication (*pubs*) was also natural logarithm transformed.

Table 1
Variable descriptions.

Variables	Descriptions
Citations	The total number of citation received by a scientist's papers published in year t. A five-year citation time window is used to count the citations for each paper, i.e., for each paper published in year t, its citations between t and $t+4$ are counted.
Citations lag	The total number of citation received by a scientist's papers published in year $t-1$. A five-year citation time window is used to count the citations for each paper.
Career age	Year t minus the year receiving the PhD degree.
Pubs	The number of publications of a scientist in year t.
Network size	The number of coauthors of a scientist in year t.
Tie strength avg	The average tie strength between a scientist and his/her coauthors of year t. A five-year time window is used for measuring tie strength between an ego and his/her coauthor in year t, specifically the number of times that they coauthored between year $t-4$ and t.
Skewness	The skewness of a scientist's tie strength distribution in year t.

5. Results

5.1. Descriptive statistics

Descriptive statistics and correlations are reported in Table 2. On average, the total number of citations is 58.44, the number of publications is 3.86, and career age is 16.79. The average network size is 9.96, ranging from 3 to 109. Network average tie strength has mean 2.46 and ranges from 1.05 to 36. Skewness has mean 1.38 and ranges from −3 to 5.71. The number of citations is highly correlated with citations in the previous year, the number of publications, and network size, all above 0.5. The number of citations is also positively correlated with network average tie strength (0.13) and skewness (0.20). The focal variable *tie strength avg* has the highest correlation with *citation lag* (0.25), and *skewness* has the highest correlation with *network size* (0.36). We are not concerned about the multicollinearity issue with these levels of correlations.

5.2. Regression results

Fixed effects Poisson models are reported in Table 3, which estimate within-individual effects. From column 1 to 6, variables of interest are added sequentially, for appropriately testing the quadric effect of network average tie strength and the moderating effect of tie strength skewness. Wald tests, $\Delta\chi^2$(model i vs $i-1$), are also reported to test whether each sequentially added variable is significant. The second set of Wald tests, $\Delta\chi^2$(model i vs. $i-2$), test the added linear and quadric terms together. For example, $\Delta\chi^2$(model i vs. $i-2$) in column 3 compares model 3 against

Table 2
Descriptive statistics.

	Variables	Mean	SD	Min	Max	Spearman correlations					
						1	2	3	4	5	6
1	Citations	58.44	80.17	0	2067						
2	Citations lag	54.12	78.89	0	2067	**.52**					
3	Career age	16.79	9.48	−6	56	−.02	**.03**				
4	Pubs	3.86	2.74	1	25	**.62**	**.33**	**.10**			
5	Network size	9.96	8.32	3	109	**.56**	**.34**	**.10**	**.65**		
6	Tie strength avg	2.46	1.42	1.05	36	**.13**	**.25**	**.11**	**.21**	**.02**	
7	Skewness	1.38	1.14	−3	5.71	**.20**	**.12**	.00	**.20**	**.36**	**−.22**

Number of observations: 6998.
Number of egos: 1042.
Correlations with bold numbers are significant at $p < .05$.

model 1 to test the significance of *tie strength avg* and *tie strength avg*^{2} together.

To better illustrate the relationship between network average tie strength and the number of citations, at different levels of tie strength skewness, Fig. 1 plots the estimated citations against *tie strength avg* in low-, median-, and high-skew (i.e., first, second, and third skewness quartile) networks separately. Estimates are based on the regression result in Table 3 column 6.

In terms of the effect of network average tie strength on citations, without accounting for skewness, neither the linear nor the quadratic terms of *tie strength avg* is significant (column 3). However, after adding the skewness and interaction terms to the model, *tie strength avg* has a significantly positively effect, while *tie strength avg*^{2} has a significantly negative effect, suggesting an inverted *U*-shaped relationship between network average tie strength and the number of citations (Fig. 1). To be more specific, for the same individual, with the same career age, number of publications, and prior citation performance, an increase in network average tie strength first has a positive and latter a negative effect on the number of citations for the currently produced papers. Furthermore, the fact that the effect of network average tie strength is insignificant when the tie strength skewness is not appropriately accounted for also confirms our argument that the effect of network average tie strength is only pronounced when the network has a homogeneous tie strength distribution.

We argued that, when the network average tie strength is high, a network with skewed tie strength distribution will achieve better performance, because it has a balance between exploration and exploitation. The regression result (Table 3 column 6) suggests a significantly positive effect of skewness when network average tie

Table 3
Fixed effects Poisson models.

	Citations					
	(1)	(2)	(3)	(4)	(5)	(6)
Career age	−0.0318***	−0.0317***	−0.0321***	−0.0320***	−0.0326***	−0.0324***
	(0.0081)	(0.0082)	(0.0083)	(0.0083)	(0.0083)	(0.0082)
Career age^2	0.0006***	0.0006***	0.0006***	0.0006***	0.0006***	0.0006***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Pubs (ln)	0.7442***	0.7512***	0.7485***	0.7485***	0.7517***	0.7565***
	(0.0401)	(0.0401)	(0.0396)	(0.0396)	(0.0391)	(0.0388)
Citations lag (ln)	0.0394***	0.0421***	0.0409**	0.0409**	0.0427***	0.0429***
	(0.0149)	(0.0161)	(0.0161)	(0.0161)	(0.0161)	(0.0161)
Network size	0.0311***	0.0301***	0.0305***	0.0305***	0.0305***	0.0302***
	(0.0041)	(0.0043)	(0.0043)	(0.0041)	(0.0040)	(0.0040)
Network size^2	−0.0003***	−0.0003***	−0.0003***	−0.0003***	−0.0003***	−0.0003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0000)
Tie strength avg		−0.0138	0.0177	0.0177	0.0572	0.0988**
		(0.0147)	(0.0327)	(0.0342)	(0.0354)	(0.0443)
Tie strength avg^2			−0.0034	−0.0035	−0.0047*	−0.0090**
			(0.0027)	(0.0028)	(0.0027)	(0.0038)
Skewness				0.0001	0.0530**	0.1034***
				(0.0138)	(0.0221)	(0.0365)
Tie Strength avg * Skewness					−0.0215***	−0.0527***
					(0.0067)	(0.0186)
Tie Strength avg^2 * Skewness						0.0034**
						(0.0017)
Ego fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Log pseudolikelihood	−74,680	−74,651	−74,617	−74,617	−74,468	−74,404
Wald χ^2	1312***	1337***	1345***	1439***	1507***	1473***
$\Delta\chi^2$(model i vs $i-1$)		0.88	1.62	0.00	10.15***	4.10**
$\Delta\chi^2$(model i vs $i-2$)			3.92			10.36***

Number of observations: 6852.
Number of egos: 896.
146 observations/egos dropped because single observation per ego.
Cluster-robust standard errors in parentheses.
* $p < .10$
** $p < .05$
*** $p < .01$.

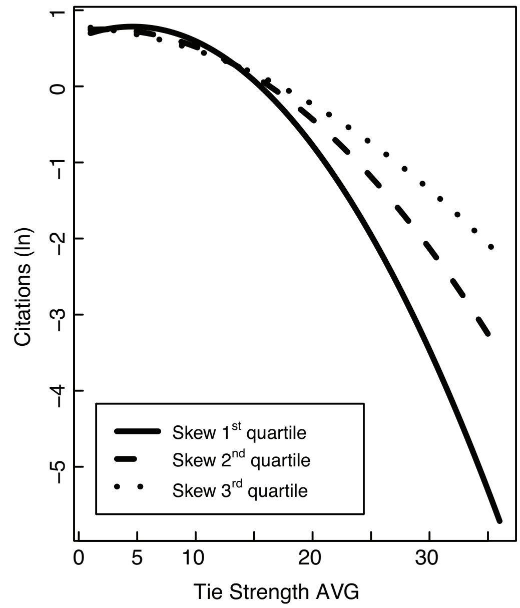

Fig. 1. Tie strength effect on citations.

strength is 0. Fig. 1 further demonstrates that the positive effect of skewness is larger when the network average tie strength is higher.

Regression results also confirm the moderating effect of skewness on network average tie strength. The interaction effect between network average tie strength and skewness is negative, indicating that the first-order positive effect of network average tie strength is smaller in more skewed networks. This negative interaction effect between network average tie strength and skewness is significantly positive, indicating that the negative effect of tie strength increases is also smaller in more skewed networks.

Effects of control variables might also be worth noting. Within the observed interval of $[-6, 56]$, career age has a continuously positive effect on citations. This is different from previously observed inverted *U*-shaped relationship between age and publication productivity (Cole, 1979; Levin and Stephan, 1991). This suggests divergent life-cycle dynamics in publication and citation performance. As expected, the number of publications is strongly related to citations; as the number of publications increase by 1%, the number of total citations increases by 0.75%. However, the prediction power of prior citations on current citations is not so strong; an 1% increases in citations received by papers published in the preceding year is associated with 0.04% increase in citations received by papers published in the present year. On the other hand, the 0-order correlation between *citations* and *citations lag* is 0.52. This suggests that the inertia or persistence in performance is more pronounced when making between-individual comparisons. In other words, scientist A's current papers are much more likely to be highly cited than scientist B's, if A's previous papers have more citations than B's. However, when comparing A's performance with him/herself over time, prior success has a lower prediction power. In line with previous literature (Lavie and Drori, 2012; McFadyen and Cannella, 2004), an inverted *U*-shaped relationship between network size and citations is observed. An increase in network size has a positive effect on citations at first, but after a size of around 58 collaborators, a further increase in network size has a negative effect on citations.

5.3. Robustness tests

There are remarkable differences between different scientific disciplines in terms of how science is produced and how the scientific work is organized (Whitley, 2000). Therefore, one question is whether our findings are field-specific or generalizable across fields. To address this concern, we run regressions for five fields separately (Table 4 column 1–5). Coefficients on *tie strength avg*, *tie strength avg*^{2}, *skewness*, and interaction terms between them all have the same direction as in Table 3, where all fields are pooled together. However, they are insignificant in most fields, except in chemistry. This is probably because we do not have enough data for individual field analysis. Note that chemistry has the largest number of observations in our sample. Given that all coefficients have the same direction, and no significant coefficients have the opposite direction, we cautiously conclude that there is no evidence that our findings are not generalizable cross field, but more data need to be collected to further test this in the future.

Like most network studies, this paper focuses on the structural aspect of the network does but not account for differences between network nodes, such as whether the collaborators are prestigious or peripheral researchers, and whether the ego has higher power in choosing potential collaborators. With individual fixed effects and prior performance to account for both time-invariant and time-variant individual heterogeneities, we believe that our model can appropriately control for differences in egos and their coauthors. Nevertheless, we run separate regressions for senior and junior researchers. The idea is that the differentiation between senior and junior should to some extent capture the difference in ego status and coauthor quality. Specifically, we partition our data into two sets with the same number of observations, one with career age above the population median and the other below. Regression results are reported in Table 4 column 6 and 7. After splitting the sample, the effect of career age disappeared. For the senior group, coefficients on focal explanatory variables are all significant and have the same direction as in Table 3. However, for the junior group, coefficients still have the same direction but none is significant. A further scrutiny of the data shows that although two groups have the same number of observations, the junior group has more individuals. Because each ego has a smaller number of observations in the junior group, there is not much within-individual variance left after controlling for individual fixed effects. Again, since we do not observe significant but opposing results, we cautiously conclude that there is no strong evidence against our findings.

Furthermore, we have done a number of robustness checks. As reported in the data section, we excluded papers with more than 15 authors for constructing the dataset, in order to address the *hyper-authorship* problem. In addition to the threshold of 15, we also tried two other thresholds, 10 and 29, and obtained similar results. We also tried an alternative strategy, excluding all the observations of an ego if he/she has a paper with more than 15 authors, and got consistent results. The reported results used one formula of skewness, using two other popular formulas yields robust results. We used Poisson model, using the negative binomial model also led to similar results. Reported results used a panel strategy with ego fixed effects, we also tried a cross-sectional strategy and got robust results, where we used the period between 2005 and 2007 for each ego, without ego fixed effects but with additional field and demographic control variables.

5.4. Alternative explanations

For the observed relatively poor performance of networks with very high tie strength, our explanation is that networks dominated by strong ties have low cognitive diversity and therefore are less likely to generate novel ideas. One alternative explanation is that it is not really because of cognitive diversity but because of network constraints. Networks present not only opportunities but also constraints (Gabbay, 1997), and this constraint effect can happen through the following two mechanisms. First, strong

Table 4
Fixed effects Poisson models by field and seniority.

	Citations						
	(1) BIOL	(2) CHEM	(3) CS	(4) EAS	(5) EE	(6) Junior	(7) Senior
Career age	−0.0394**	−0.0370***	0.0079	0.0036	−0.0474	−0.0269	−0.0153
	(0.0155)	(0.0123)	(0.0323)	(0.0114)	(0.0301)	(0.0282)	(0.0260)
Career age^2	0.0005	0.0007***	0.0004	−0.0002	0.0021***	0.0002	0.0003
	(0.0004)	(0.0003)	(0.0007)	(0.0003)	(0.0007)	(0.0014)	(0.0004)
Pubs (ln)	0.6664***	0.7854***	0.7093***	0.6271***	0.9675***	0.7505***	0.7676***
	(0.0584)	(0.0605)	(0.1496)	(0.0778)	(0.1083)	(0.0510)	(0.0564)
Citations lag (ln)	0.0431**	0.0764**	−0.1064***	0.0040	0.0754*	0.0000	0.0353*
	(0.0215)	(0.0295)	(0.0374)	(0.0206)	(0.0423)	(0.0257)	(0.0188)
Network size	0.0348***	0.0311***	0.0419*	0.0450***	0.0124	0.0293***	0.0332***
	(0.0052)	(0.0076)	(0.0248)	(0.0107)	(0.0080)	(0.0061)	(0.0061)
Network size^2	−0.0003***	−0.0004***	0.0003	−0.0004***	−0.0001	−0.0003***	−0.0003***
	(0.0000)	(0.0001)	(0.0005)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Tie strength avg	0.0088	0.1416	0.2338	0.1507	0.0717	0.0683	0.1328**
	(0.0730)	(0.0958)	(0.2663)	(0.0964)	(0.0980)	(0.0580)	(0.0627)
Tie strength avg^2	−0.0062	−0.0157**	−0.0164	−0.0152*	−0.0037	−0.0059	−0.0166***
	(0.0060)	(0.0080)	(0.0275)	(0.0083)	(0.0067)	(0.0050)	(0.0059)
Skewness	−0.0202	0.1670**	0.0616	0.1863*	0.1413	0.0747	0.1463**
	(0.0729)	(0.0693)	(0.2181)	(0.0986)	(0.1029)	(0.0466)	(0.0662)
Tie Strength avg * Skewness	−0.0062	−0.0958***	−0.0962	−0.0838	−0.0701*	−0.0376	−0.0921**
	(0.0449)	(0.0345)	(0.1222)	(0.0558)	(0.0386)	(0.0232)	(0.0364)
Tie Strength avg^2 * Skewness	0.0000	0.0099***	0.0156	0.0071	0.0027	0.0020	0.0095**
	(0.0046)	(0.0034)	(0.0130)	(0.0054)	(0.0028)	(0.0020)	(0.0037)
Ego fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N obs	1585	2207	727	1446	887	3357	3355
N egos	201	236	135	197	127	666	486
Log pseudolikelihood	−19,122	−24,791	−6106	−13,335	−8664	−34,642	−32,597
Wald χ^2	944***	847***	142***	788***	917***	770***	999***

Cluster-robust standard errors in parentheses.
* $p < .10$
** $p < .05$
*** $p < .01$.

relationships are binding and powerful to impose obligations to cooperate (Hansen, 1999; Weick, 1976). This strong binding effect may help performance of the group but is not necessarily optimal for the individual personally, because it reduces individual's autonomy and prevents him from strategically allocating energy and efforts across different collaborations in order to maximize his personal gains. Second, strong relationships further prevents individuals from altering current network structure or establishing new and more efficient networks (Gabbay, 1997). To assess this alternative explanation, the Spearman correlation between the number of new collaborators in 2008 and the network average tie strength between 2005 and 2007 was calculated, which is 0.03 and insignificant ($p = 0.34$). Therefore, we did not find evidence that strong ties would restrict developing new collaborative relations.

The second alternative explanation is that it is not really because strong-tie-collaborations are less likely to be creative but because they choose to produce many small papers which are expected to have narrow applications and small impact. Scientists can be modeled as rational agents and only pursue the collaborative project when the expected payoffs are greater than the costs. Since strong-tie-collaborations have lower costs because of high *cognitive* and *relational capitals*, many small projects with low payoffs become “profitable.” Under such circumstance, the observed average citations of strong-tie-collaborations will be pulled down by these small papers. The other side of the costs story is that experimental projects are also only “profitable” for strong-tie-collaborations (Aghion et al., 2008; Catalini, 2012). An experimental project has a high payoff once it reaches the final success. However, it takes a long trial-and-error process and may fail at any stage in the process. After accounting for the high probability of failure, the expected payoff of this experimental project is very low. Therefore, an experimental project would be pursued only when the costs are very low, such as in strong-tie-collaborations. In summary, our explanation about the declined creative capacity predicts that strong-tie-collaborations have both low average and maximum citations, while the alternative costs theory predicts that strong-tie-collaborations have low average but high maximum citations. To test these two competing theories, we classified each scientist's coauthors, in each year, into two types: *new* (not coauthored in the last three years) and *repeated* (coauthored at least once in the last three years). Subsequently, we compared a scientist's *new* collaboration papers (i.e., papers with only *new* collaborators) and his *repeated* collaboration papers (i.e., papers with only *repeated* collaborators). This comparison focuses on purely new versus purely repeated collaborations and therefore excluded solo authored papers and papers with both new and repeated coauthors. Paired Wilcoxon Signed-Rank tests suggested that repeated-collaborations have significantly lower average and maximum citations. This result supported our creative capacity theory and rejected the costs theory.

6. Conclusions

This paper investigated the relationship between egocentric collaboration networks and knowledge creation at the individual level. For egocentric collaboration networks, this paper focused on network characteristics in terms of tie strength and strong/weak tie configuration. Knowledge creation was evaluated by the number of citations. An inverted *U*-shaped relationship was found between network average tie strength and knowledge creation. An increase in tie strength (1) on the one hand increases the cognitive and relational capital in the network and therefore facilitates the collaborative knowledge creation process, and (2) on the other hand decreases the cognitive diversity and therefore impedes the generation of novel and useful ideas. Taking two mechanisms together, an increase in tie strength initially has a positive effect on knowledge creation, but the effect turns into negative after the tie strength reaches a threshold. Furthermore, when the network average tie strength is high, a more skewed network can achieve better performance than a less skewed one, because a more skewed network still has a "healthy" mixture of strong and weak ties and a balance between exploration and exploitation. In addition, the skewness of tie strength distribution moderates the effect of network average tie strength. In a more skewed network, both the initial positive effect and the later negative effect of network average tie strength are smaller than in a less skewed network.

There are several limitations of this study. First, because this study relied heavily on the bibliometric data, we could not avoid the issues pertaining to the use of citations for evaluating knowledge creation (Martin and Irvine, 1983) and the use of coauthorships for measuring collaborations (Katz and Martin, 1997; Laudel, 2002), although we took various procedures to account for those potential issues and various robustness tests to ensure the reliability of our findings. It may be helpful for future research to use different kinds of data and measures to validate these findings. In addition, disambiguating author names is a difficult obstacle for individual-level analysis. We have invested a lot of effort in collecting and cleaning publication record for individual scientists. However, physics was excluded from the data cleaning process, because it is much more challenging to disambiguate physicists whose papers often have hundreds of coauthors. However, leaving one important field out is a big limitation of this study. Furthermore, this paper studied only the egocentric networks, direct ties, and the strength of direct ties, but not the global networks, indirect ties, or other aspects of network structure. Therefore, it does not answer questions concerning these other aspects. Different from previous individual level network studies following the social capital perspective and studying the indirect effect of previous network on current knowledge creation, this paper studied the current network as work organizations and its direct effect on knowledge creation. Our approach contributes to understanding the organization of science and social capital mobilization, but also has a price. Specifically, we cannot study the effect of current network on productivity with our data, because the construction of current network and the assessment of current productivity will be like counting the same set of papers in different ways, which would have serious endogeneity issues. Therefore, we only focused on the effect of current network on citations, controlling for productivity, where citations occur temporally after the network and is out of the hands of egos.

This paper makes the following theoretical contributions. First, it contributes to the organization of science literature. The production of scientific knowledge is increasingly collaborative. On the one hand, it is important to study collaborative teams which are the actual "factories" producing science. On the other hand, because collaborative teams in science are fluid and interdependent, it is also important to look beyond team boundaries and account for external activities, in other words, it is also important to adopt the network approach for studying the organization of collaborative science. Second, it contributes to the studies of individual scientists. Previous studies adopt the social capital perspective, studying the effect of previous network on current performance. However, this paper follows a work organization perspective, studying the effect of current network on current knowledge creation. This alternative approach allows for studying the knowledge creation process and resource mobilization in the current network. Third, this study also demonstrates the importance of investigating the configuration of ties rather than adopting a simple dichotomy between weak and strong ties. In social network studies, it might not always be appropriate to assume that network ties are homogeneous or view the network level effect as a simple aggregation of dyadic level effects. For example, in the world of science, scientists' egocentric collaboration networks have a mixture of strong and weak ties, and the network ties are heterogeneous. This complicated configuration characteristics should be studied, instead of simply studying the overall tie strength of the network. In this paper, we studied the skewness of tie strength distribution and demonstrated the positive and moderating effects associated with this tie configuration characteristic.

This paper also has implications for science funding and research management. Given the widely accepted notion that collaboration is good for productivity and creativity, many funding agencies have established special programs supporting collaborative research. However, not all collaborations are equally beneficial. Therefore, it is important to distinguish between different types of collaborations and design more targeted funding polices. In addition, there is increasing concern in the United States that the current competitive project-based funding model may impede path-breaking discoveries, because it favors investigators with established successful records and favors projects confirming rather than challenging current norms (Alberts, 2010; NPR, 2013; Petsko, 2012; Walsh, 2013). This paper also contributes to this discussion; suggesting another risk in the current funding model, which may encourage repeated collaborations with prior success but discourage the exploration of new collaborative relations. Consequently, this may drive up the tie strength of scientists' collaboration networks and impede creativity at both the individual level and the level of the whole science system. In addition, findings of this paper suggest that individual scientists can benefit from a balance between explorative and exploitative collaborations and a mixture of both weak and strong ties. It is a good strategy for individual scientists to maintain a small number of very close collaborators but at the same time expand personal networks and explore new collaborative opportunities.

Dourish (2006) suggested that while empirical contributions provide more specific implications for practice, analytical contributions lead to more profound implications in terms of new ways to approach the problem. Following this idea, this paper also contributes to the economics of science. Economics studies have modeled the behavior of scientists in choosing collaborative partners or projects, where the goal of an scientist is to maximize his/her personal payoffs from one specific collaboration (Banal-Estañol et al., 2014; Carayol, 2003; Gans and Murray, 2013). However, this paper may suggest take into account the interdependence between different collaborations and consider the goal of a scientist as maximizing payoffs from his portfolio of collaborations instead of a single collaboration. In addition, this study also suggests new ways for designing the funding systems. There is a remarkable transition toward competitive project-based funding systems (Stephan, 2013) and a long-standing interest in estimating the funding effect on research productivity and impact (Arora and Gambardella, 2005; Azoulay et al., 2011; Jacob and Lefgren, 2011). On the other hand, there is also debate on whether project- or individual-based funding system is more efficient (Ioannidis, 2011). This paper contributes to the design of funding programs by suggesting alternative strategies. For example, set up lab-based funding programs with a proportion of funds reserved for outreaching activities. On the one hand, this funding strategy facilitates the development of strong ties between lab members. On the other hand, it creates opportunities for lab members to establish weak ties outside the lab. One essential component of this funding strategy is to encourage active interactions between lab members, instead of building a virtual lab pooling researchers' profiles online without substantial collaborations between them. In addition, compared with project-based funding strategy, this lab-based funding strategy provides researchers with the flexibility to allocate the funds, which is found to be beneficial to creativity (Heinze et al., 2009). Another essential component is the reserved funds to encourage researchers to establish weak ties through visiting other institutions, hosting visiting scholars, and organizing workshops and conferences.

Acknowledgements

Earlier versions of this paper were presented at EuSPRI Early Career Research Conference, Milan; Atlanta Conference on Science and Innovation Policy; and Sunbelt XXXIV, St. Pete Beach. The author thanks the conference participants, the anonymous reviewers, Diana Hicks, Julia Melkers, John Walsh, Juan Rogers, Yajun Mei, Sybille Hinze, Stefan Hornbostel, and Cristina Rossi-Lamastra for very helpful comments and suggestions on the paper. The survey and publication data are from the NETWISE-I survey funded by the United States National Science Foundation grant, "Women in Science and Engineering: Network Access, Participation, and Career Outcomes" (NSF Grant # REC-0529642, PI Julia Melkers). The citation data are from a bibliometrics database developed by the Competence Center for Bibliometrics for the German Science System (KB) and derived from the 1980 to 2012 Science Citation Index Expanded (SCI-E), Social Sciences Citation Index (SSCI), Arts and Humanities Citation Index (AHCI), Conference Proceedings Citation Index-Science (CPCI-S), and Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH) prepared by Thomson Reuters (Scientific) Inc. (TR®), Philadelphia, Pennsylvania, USA: ©Copyright Thomson Reuters (Scientific) 2013. KB is funded by the German Federal Ministry of Education and Research(BMBF, project number: 01PQ08004A). The author gratefully acknowledges a postdoctoral fellowship from the Research Foundation – Flanders (FWO).

References

Abbasi, A., Altmann, J., Hossain, L., 2011. Identifying the effects of co-authorship networks on the performance of scholars: a correlation and regression analysis of performance measures and social network analysis measures. J. Informetr. 5 (4), 594–607, http://dx.doi.org/10.1016/j.joi.2011.05.007.

Adams, J.D., Black, G.C., Clemmons, J.R., Stephan, P.E., 2005. Scientific teams and institutional collaborations: Evidence from US universities, 1981–1999. Res. Policy 34 (3), 259–285, http://dx.doi.org/10.1016/j.respol.2005.01.014.

Aghion, P., Dewatripont, M., Stein, J.C., 2008. Academic freedom, private-sector focus, and the process of innovation. RAND J. Econ. 39 (3), 617–635, http://dx.doi.org/10.1111/j.1756-2171.2008.00031.x.

Aksnes, D.W., 2003. A macro study of self-citation. Scientometrics 56 (2), 235–246, http://dx.doi.org/10.1023/a:1021919228368.

Alberts, B., 2010. Overbuilding research capacity. Science 329 (5997), 1257, http://dx.doi.org/10.1126/science.1197077.

Allison, P.D., Waterman, R.P., 2002. Fixed-effects negative binomial regression models. Sociol. Methodol. 32 (1), 247–265, http://dx.doi.org/10.1111/1467-9531.00117.

Amabile, T.M., 1983. The social psychology of creativity: a componential conceptualization. J. Pers. Soc. Psychol. 45 (2), 357–376, http://dx.doi.org/10.1037/0022-3514.45.2.357.

Ancona, D.G., 1990. Outward bound: strategies for team survival in an organization. Acad. Manage. J. 33 (2), 334–365, http://dx.doi.org/10.2307/256328.

Arora, A., Gambardella, A., 2005. The impact of NSF support for basic research in economics. Annales d'Économie et de Statistique (79/80), 91–117, http://dx.doi.org/10.2307/20777571.

Azoulay, P., Graff Zivin, J.S., Manso, G., 2011. Incentives and creativity: evidence from the academic life sciences. RAND J. Econ. 42 (3), 527–554, http://dx.doi.org/10.1111/j.1756-2171.2011.00140.x.

Banal-Estañol, A., Macho-Stadler, I., Pérez-Castrillo, D., 2014. Endogeneous Matching in University-Industry Collaboration: Theory and Empirical Evidence from the UK. Universitat Pompeu Fabra Working paper.

Benner, M.J., Tushman, M.L., 2003. Exploitation, exploration, and process management: the productivity dilemma revisited. Acad. Manage. Rev. 28 (2), 238–256, http://dx.doi.org/10.5465/amr.2003.9416096.

Borgman, C.L., 2007. Scholarship in the Digital Age: Information, Infrastructure, and The Internet. MIT Press, Cambridge, MA.

Börner, K., Maru, J.T., Goldstone, R.L., 2004. The simultaneous evolution of author and paper networks. Proc. Natl. Acad. Sci. U. S. A. 101 (suppl. 1), 5266–5273, http://dx.doi.org/10.1073/pnas.0307625100.

Bornmann, L., Daniel, H.D., 2008. What do citation counts measure? A review of studies on citing behavior. J. Doc. 64 (1), 45–80, http://dx.doi.org/10.1108/00220410810844150.

Burt, R.S., 1992. Structural Holes: The Social Structure of Competition. Harvard University Press, Cambridge, MA.

Carayol, N., 2003. Objectives, agreements and matching in science-industry collaborations: reassembling the pieces of the puzzle. Res. Policy 32 (6), 887–908, http://dx.doi.org/10.1016/S0048-7333(02)00108-7.

Catalini, C., 2012. Microgeography and the Direction of Inventive Activity. Available at SSRN 2126890.

Chen, G., 2005. Newcomer adaptation in teams: multilevel antecedents and outcomes. Acad. Manage. J. 48 (1), 101–116, http://dx.doi.org/10.5465/amj.2005.15993147.

Chompalov, I., Genuth, J., Shrum, W., 2002. The organization of scientific collaborations. Res. Policy 31 (5), 749–767, http://dx.doi.org/10.1016/S0048-7333(01)00145-7.

Cohen, M.D., March, J.G., Olsen, J.P., 1972. A garbage can model of organizational choice. Adm. Sci. Q. 17 (1), 1–25, http://dx.doi.org/10.2307/2392088.

Cole, S., 1979. Age and scientific performance. Am. J. Sociol. 84 (4), 958–977, http://dx.doi.org/10.2307/2778031.

Cole, S., Cole, J.R., 1967. Scientific output and recognition. Am. Sociol. Rev. 32 (3), 377–390, http://dx.doi.org/10.2307/2091085.

Cronin, B., 2001. Hyperauthorship: a postmodern perversion or evidence of a structural shift in scholarly communication practices? J. Am. Soc. Inf. Sci. Technol. 52 (7), 558–569, http://dx.doi.org/10.1002/asi.1097.

Cummings, J.N., Kiesler, S., 2005. Collaborative research across disciplinary and organizational boundaries. Soc. Stud. Sci. 35 (5), 703–722, http://dx.doi.org/10.1177/03061270505535.

Cummings, J.N., Kiesler, S., Zadeh, R.B., Balakrishnan, A.D., 2013. Group heterogeneity increases the risks of large group size: a longitudinal study of productivity in research groups. Psychol. Sci. 24 (6), 880–890.

De Bellis, N., 2009. Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics. Scarecrow Press, Lanham, MD.

Dourish, P., 2006. Implications for design. In: Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montréal, Québec, Canada.

Edwards, P., Mayernik, M.S., Batcheller, A., Bowker, G., Borgman, C., 2011. Science friction: data, metadata, and collaboration. Soc. Stud. Sci., http://dx.doi.org/10.1177/0306312711413314.

Falk-Krzesinski, H.J., Börner, K., Contractor, N., Fiore, S.M., Hall, K.L., Keyton, J., Spring, B., Stokols, D., Trochim, W., Uzzi, B., 2010. Advancing the science of team science. Clin. Transl. Sci. 3 (5), 263–266, http://dx.doi.org/10.1111/j.1752-8062.2010.00223.x.

Fang, C., Lee, J., Schilling, M.A., 2010. Balancing exploration and exploitation through structural design: the isolation of subgroups and organizational learning. Organ. Sci. 21 (3), 625–642, http://dx.doi.org/10.1287/orsc.1090.0468.

Finholt, T.A., Olson, G.M., 1997. From laboratories to collaboratories: a new organizational form for scientific collaboration. Psychol. Sci. 8 (1), 28–36, http://dx.doi.org/10.1111/j.1467-9280.1997.tb00540.x.

Fiore, S.M., 2008. Interdisciplinarity as teamwork: how the science of teams can inform team science. Small Group Res. 39 (3), 251–277, http://dx.doi.org/10.1177/1046496408317797.

Fleming, L., 2001. Recombinant uncertainty in technological search. Manage. Sci. 47 (1), 117–132, http://dx.doi.org/10.1287/mnsc.47.1.117.10671.

Ford, C.M., 1996. Theory of individual creative action in multiple social domains. Acad. Manage. Rev. 21 (4), 1112–1142, http://dx.doi.org/10.2307/259166.

Gabbay, S.M., 1997. Social Capital in the Creation of Financial Capital: The Case of Network Marketing. Stipes Publishing, Champaign, IL.

Gans, J.S., Murray, F., 2013. Credit History: The Changing Nature of Scientific Credit. National Bureau of Economic Research.

Garfield, E., 1973. Citation and distinction. Nature 242 (5398), 485, http://dx.doi.org/10.1038/242485a0.

Garfield, E., 1980. Premature discovery or delayed recognition – why. Curr. Contents 21, 5–10.

Gino, F., Argote, L., Miron-Spektor, E., Todorova, G., 2010. First, get your feet wet: the effects of learning from direct and indirect experience on team creativity. Organ. Behav. Hum. Decis. Process. 111 (2), 102–115, http://dx.doi.org/10.1016/j.obhdp.2009.11.002.

Glänzel, W., Debackere, K., Thijs, B., Schubert, A., 2006. A concise review on the role of author self-citations in information science, bibliometrics and science policy. Scientometrics 67 (2), 263–277, http://dx.doi.org/10.1007/s11192-006-0098-9.

Glänzel, W., Schlemmer, B., Thijs, B., 2003. Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon. Scientometrics 58 (3), 571–586.

Gonzalez-Brambila, C.N., Veloso, F.M., Krackhardt, D., 2013. The impact of network embeddedness on research output. Res. Policy 42 (9), 1555–1567, http://dx.doi.org/10.1016/j.respol.2013.07.008.

Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: applications to Poisson models. Econometrica 52 (3), 701–720, http://dx.doi.org/10.2307/1913472.

Granovetter, M., 1973. Strength of weak ties. Am. J. Sociol. 78 (6), 1360–1380.

Granovetter, M., 1985. Economic action and social structure: the problem of embeddedness. Am. J. Sociol. 91 (3), 481–510.

Greene, W., 2005. Functional form and heterogeneity in models for count data. Found. Trends Econom. 1 (2), 113–218.

Guimarães, P., 2008. The fixed effects negative binomial model revisited. Econ. Lett. 99 (1), 63–66, http://dx.doi.org/10.1016/j.econlet.2007.05.030.

Guimera, R., Uzzi, B., Spiro, J., Amaral, L.A.N., 2005. Team assembly mechanisms determine collaboration network structure and team performance. Science 308 (5722), 697–702, http://dx.doi.org/10.1126/science.1106340.
Gupta, A.K., Smith, K.G., Shalley, C.E., 2006. The interplay between exploration and exploitation. Acad. Manage. J. 49 (4), 693–706, http://dx.doi.org/10.2307/20159793.
Haeussler, C., Sauermann, H., 2013. Credit where credit is due? The impact of project contributions and social factors on authorship and inventorship. Res. Policy 42 (3), 688–703, http://dx.doi.org/10.1016/j.respol.2012.09.009.
Hall, B.H., Ziedonis, R.H., 2001. The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1979–1995. Rand J. Econ. 32 (1), 101–128, http://dx.doi.org/10.2307/2696400.
Hansen, M.T., 1999. The search-transfer problem: the role of weak ties in sharing knowledge across organization subunits. Adm. Sci. Q. 44 (1), 82–111, http://dx.doi.org/10.2307/2667032.
Hausman, J., Hall, B.H., Griliches, Z., 1984. Econometric models for count data with an application to the patents R&D relationship. Econometrica 52 (4), 909–938, http://dx.doi.org/10.2307/1911191.
Heinze, T., Shapira, P., Rogers, J.D., Senker, J.M., 2009. Organizational and institutional influences on creativity in scientific research. Res. Policy 38 (4), 610–623, http://dx.doi.org/10.1016/j.respol.2009.01.014.
Hemlin, S., Allwood, C.M., Martin, B.R., Mumford, M.D., 2013. Creativity and Leadership in Science, Technology, and Innovation. Routledge, New York, NY.
Hicks, D., Katz, J.S., 1996. Where is science going? Science Technol. Hum. Values 21 (4), 379–406, http://dx.doi.org/10.2307/690087.
Hottenrott, H., Lopes-Bento, C., 2015. Quantity or quality? Knowledge alliances and their effects on patenting. Ind. Corp. Change, http://dx.doi.org/10.1093/icc/dtu019.
Hudson, J., 1996. Trends in multi-authored papers in economics. J. Econ. Perspect. 10 (3), 153–158.
Ioannidis, J.P.A., 2011. More time for research: fund people not projects. Nature 477 (7366), 529–531.
Jacob, B.A., Lefgren, L., 2011. The impact of research grant funding on scientific productivity. J. Public Econ. 95 (9–10), 1168–1177, http://dx.doi.org/10.1016/j.jpubeco.2011.05.005.
Katila, R., Ahuja, G., 2002. Something old, something new: a longitudinal study of search behavior and new product introduction. Acad. Manage. J. 45 (6), 1183–1194, http://dx.doi.org/10.2307/3069433.
Katz, J.S., 1999. The self-similar science system. Res. Policy 28 (5), 501–517, http://dx.doi.org/10.1016/S0048-7333(99)00010-4.
Katz, J.S., 2000. Scale-independent indicators and research evaluation. Sci. Public Policy 27 (1), 23–36, http://dx.doi.org/10.3152/147154300781782156.
Katz, J.S., Martin, B.R., 1997. What is research collaboration? Res. Policy 26 (1), 1–18.
Klenk, N.L., Hickey, G.M., MacLellan, J.I., 2010. Evaluating the social capital accrued in large research networks: the case of the Sustainable Forest Management Network (1995–2009). Soc. Stud. Sci. 40 (6), 931–960, http://dx.doi.org/10.1177/0306312710374130.
Krackhardt, D., 1992. The strength of strong ties: the importance of philos in organizations. In: Nohria, N., Eccles, R.G. (Eds.), Networks and Organizations: Structure, Form, and Action. Harvard Business School Press, Boston, MA, pp. 216–239.
Kuhn, T.S., 1970. The Structure of Scienti C Revolutions. Chicago, IL.
Latour, B., Woolgar, S., 1986. Laboratory Life: The Construction of Scientific Facts. Princeton University Press, Princeton, NJ.
Laudel, G., 2002. What do we measure by co-authorships? Res. Eval. 11 (1), 3–15.
Lavie, D., Drori, I., 2012. Collaborating for knowledge creation and application: the case of nanotechnology research programs. Organ. Sci. 23 (3), 704–724, http://dx.doi.org/10.1287/orsc.1110.0656.
Lazer, D., Friedman, A., 2007. The network structure of exploration and exploitation. Adm. Sci. Q. 52 (4), 667–694, http://dx.doi.org/10.2189/asqu.52.4.667.
Lee, S., Bozeman, B., 2005. The impact of research collaboration on scientific productivity. Soc. Stud. Sci. 35 (5), 673–702, http://dx.doi.org/10.1177/0306312705052359.
Lee, Y.-N., Walsh, J.P., Wang, J., 2015. Creativity in scientific teams: unpacking novelty and impact. Res. Policy 44 (3), 684–697, http://dx.doi.org/10.1016/j.respol.2014.10.007.
Levin, S.G., Stephan, P.E., 1991. Research productivity over the life-cycle – evidence for academic scientists. Am. Econ. Rev. 81 (1), 114–132.
Levine, J.M., Moreland, R.L., 2004. Collaboration: the social context of theory development. Pers. Soc. Psychol. Rev. 8 (2), 164–172, http://dx.doi.org/10.1207/s15327957pspr0802_10.
Li, E.Y., Liao, C.H., Yen, H.R., 2013. Co-authorship networks and research impact: a social capital perspective. Res. Policy 42 (9), 1515–1530, http://dx.doi.org/10.1016/j.respol.2013.06.012.
Lin, N., 1999. Social networks and status attainment. Ann. Rev. Sociol. 25, 467–487, http://dx.doi.org/10.2307/223513.
Lin, N., 2001. Social Capital: A Theory of Social Structure and Action. Cambridge University Press, Cambridge, UK/New York, NY.
Lin, N., Ensel, W.M., 1989. Life stress and health: stressors and resources. Am. Sociol. Rev. 54 (3), 382.
March, J.G., 1991. Exploration and exploitation in organizational learning. Organ. Sci. 2 (1), 71–87, http://dx.doi.org/10.1287/orsc.2.1.71.
Martin, B.R., Irvine, J., 1983. Assessing basic research: some partial indivators of scientific progress in radio astronomy. Res. Policy 12 (2), 61–90, http://dx.doi.org/10.1016/0048-7333(83)90005-7.
McFadyen, M.A., Cannella, A.A., 2004. Social capital and knowledge creation: diminishing returns of the number and strength of exchange relationships. Acad. Manage. J. 47 (5), 735–746.
McFadyen, M.A., Semadeni, M., Cannella, A.A., 2009. Value of strong ties to disconnected others: examining knowledge creation in biomedicine. Organ. Sci. 20 (3), 552–564, http://dx.doi.org/10.1287/orsc.1080.0388.
Mednick, S.A., 1962. The associative basis of the creative process. Psychol. Rev. 69 (3), 220–232, http://dx.doi.org/10.1037/h0048850.
Melin, G., 2000. Pragmatism and self-organization: research collaboration on the individual level. Res. Policy 29 (1), 31–40.
Merton, R.K., 1973. The Sociology of Science. University of Chicago Press, Chicago, IL.
Merton, R.K., 1983. Foreword. In: Garfield, E. (Ed.), Citation Indexing, Its Theory and Application in Science, Technology, and Humanities (pp. xiii, 274 p.). ISI Press, Philadelphia, PA.
Moed, H.F., Burger, W., Frankfort, J., van Raan, A., 1985. The application of bibliometric indicators: important field- and time-dependent factors to be considered. Scientometrics 8 (3), 177–203, http://dx.doi.org/10.1007/bf02016935.
Murayama, K., Nirei, M., Shimizu, H., 2015. Management of science, serendipity, and research performance: evidence from a survey of scientists in Japan and the U.S. Res. Policy 44 (4), 862–873, http://dx.doi.org/10.1016/j.respol.2015.01.018.
Nahapiet, J., Ghoshal, S., 1998. Social capital, intellectual capital, and the organizational advantage. Acad. Manage. Rev. 23 (2), 242–266.
Nelson, R.R., Winter, S.G., 1982. An Evolutionary Theory of Economic Change. Belknap Press of Harvard University Press, Cambridge, MA.
Newman, J.M., Cooper, E., 1993. Determinants of academic recognition: the case of the journal of applied psychology. J. Appl. Psychol. 78 (3), 518–526, http://dx.doi.org/10.1037/0021-9010.78.3.518.
Newman, M.E.J., 2004. Coauthorship networks and patterns of scientific collaboration. Proc. Natl. Acad. Sci. 101 (suppl. 1), 5200–5205, http://dx.doi.org/10.1073/pnas.0307545100.
Nonaka, I., 1994. A dynamic theory of organizational knowledge creation. Organ. Sci. 5 (1), 14–37.
NPR, 2013. Scientists win Nobel for work on how cells communicate, http://www.npr.org/templates/story/story.php?storyId=230192033 (retrieved 29.04.15).
Obstfeld, D., 2005. Social networks, the tertius iungens orientation, and involvement in innovation. Adm. Sci. Q. 50 (1), 100–130, http://dx.doi.org/10.2189/asqu.2005.50.1.100.
Page, S.E., 2007. The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies. Princeton University Press, Princeton, NJ.
Perretti, F., Negro, G., 2006. Filling empty seats: how status and organizational hierarchies affect exploration versus exploitation in team design. Acad. Manage. J. 49 (4), 759–777, http://dx.doi.org/10.2307/20159797.
Perry-Smith, J.E., Shalley, C.E., 2003. The social side of creativity: a static and dynamic social network perspective. Acad. Manage. Rev. 28 (1), 89–106.
Petsko, G.A., 2012. Goodbye, Columbus. Genome Biol. 13 (5), http://dx.doi.org/10.1186/gb-2012-13-5-155.
Podolny, J.M., Baron, J.N., 1997. Resources and relationships: social networks and mobility in the workplace. Am. Sociol. Rev. 62 (5), 673–693.
Porac, J.F., Wade, J.B., Fischer, H.M., Brown, J., Kanfer, A., Bowker, G., 2004. Human capital heterogeneity, collaborative relationships, and publication patterns in a multidisciplinary scientific alliance: a comparative case study of two scientific teams. Res. Policy 33 (4), 661–678, http://dx.doi.org/10.1016/j.respol.2004.01.007.
Price, D.J.D., 1986. Little Science, Big Science. Columbia University Press, New York, NY.
Reagans, R., Argote, L., Brooks, D., 2005. Individual experience and experience working together: predicting learning rates from knowing who knows what and knowing how to work together. Manage. Sci. 51 (6), 869–881, http://dx.doi.org/10.1287/mnsc.1050.0366.
Reagans, R., McEvily, B., 2003. Network structure and knowledge transfer: the effects of cohesion and range. Adm. Sci. Q. 48 (2), 240–267, http://dx.doi.org/10.2307/3556658.
Schumpeter, J.A., 1939. Business Cycles; A Theoretical, Historical, and Statistical Analysis of the Capitalist Process, 1st ed. McGraw-Hill Book Company, Inc., New York, London.
Shrum, W., Chompalov, I., Genuth, J., 2001. Trust, conflict and performance in scientific collaborations. Soc. Stud. Sci. 31 (5), 681–730, http://dx.doi.org/10.1177/030631201031005002.
Simonton, D.K., 1984. Artistic creativity and interpersonal relationships across and within generations. J. Pers. Soc. Psychol. 46 (6), 1273–1286, http://dx.doi.org/10.1037/0022-3514.46.6.1273.
Simonton, D.K., 1999. Origins of Genius: Darwinian Perspectives on Creativity. Oxford University Press, New York.
Skilton, P.F., Dooley, K.J., 2010. The effects of repeated collaboration on creative abrasion. Acad. Manage. Rev. 35 (1), 118–134.
Somaya, D., Williamson, I.O., Zhang, X.M., 2007. Combining patent law expertise with R&D for patenting performance. Organ. Sci. 18 (6), 922–937, http://dx.doi.org/10.1287/orsc.1070.0292.

Star, S.L., Griesemer, J.R., 1989. Institutional ecology, 'translations' and boundary objects: amateurs and professionals in Berkeley's museum of vertebrate zoology, 1907–39. Soc. Stud. Sci. 19 (3), 387–420, http://dx.doi.org/10.1177/030631289019003001.

StataCorp, 2013a. xtnbreg, http://www.stata.com/manuals13/xtxtnbreg.pdf (retrieved 18.04.15).

StataCorp, 2013c. xtpoisson, http://www.stata.com/manuals13/xtxtpoisson.pdf (retrieved 18.04.15).

Stephan, P.E., 2013. The Endless Frontier: Reaping what Bush Sowed? National Bureau of Economic Research.

Stokols, D., Hall, K.L., Taylor, B.K., Moser, R.P., 2008. The science of team science: overview of the field and introduction to the supplement. Am. J. Prev. Med. 35 (suppl. 2), S77–S89, http://dx.doi.org/10.1016/j.amepre.2008.05.002.

Sun, X., Kaur, J., Milojevic, S., Flammini, A., Menczer, F., 2013. Social dynamics of science. Sci. Rep., 3.

Tang, L., Hu, G., 2013. Tracing the footprint of knowledge spillover: evidence from U.S.–China collaboration in nanotechnology. J. Am. Soc. Inf. Sci. Technol. 64 (9), 1791–1801, http://dx.doi.org/10.1002/asi.22873.

Tortoriello, M., Krackhardt, D., 2010. Activating cross-boundary knowledge: the role of Simmelian ties in the generation of innovations. Acad. Manage. J. 53 (1), 167–181.

Uzzi, B., 1996. The sources and consequences of embeddedness for the economic performance of organizations: the network effect. Am. Sociol. Rev. 61 (4), 674–698.

Uzzi, B., 1997. Social structure and competition in interfirm networks: the paradox of embeddedness. Adm. Sci. Q. 42 (1), 35–67.

Uzzi, B., Mukherjee, S., Stringer, M., Jones, B., 2013. Atypical combinations and scientific impact. Science 342 (6157), 468–472, http://dx.doi.org/10.1126/science.1240474.

Uzzi, B., Spiro, J., 2005. Collaboration and creativity: the small world problem. Am. J. Sociol. 111 (2), 447–504, http://dx.doi.org/10.1086/432782.

Van Raan, A., 2004. Sleeping beauties in science. Scientometrics 59 (3), 467–472.

van Rijnsoever, F.J., Hessels, L.K., 2011. Factors associated with disciplinary and interdisciplinary research collaboration. Res. Policy 40 (3), 463–472, http://dx.doi.org/10.1016/j.respol.2010.11.001.

van Rijnsoever, F.J., Hessels, L.K., Vandeberg, R.L.J., 2008. A resource-based view on the interactions of university researchers. Res. Policy 37 (8), 1255–1266, http://dx.doi.org/10.1016/j.respol.2008.04.020.

Walsh, D., 2013. Not safe for funding: the N.S.F. and the economics of science. New Yorker, http://www.newyorker.com/tech/elements/not-safe-for-funding-the-n-s-f-and-the-economics-of-science (retrieved 29.04.15).

Walsh, J.P., Lee, Y.-N., 2015. The bureaucratization of science. Res. Policy 44 (8), 1584–1600, http://dx.doi.org/10.1016/j.respol.2015.04.010.

Walsh, J.P., Maloney, N.G., 2002. Computer network use, collaboration structures and productivity. In: Hinds, P., Kiesler, S. (Eds.), Distributed Work. MIT Press, Cambridge, MA, pp. 433–458.

Wang, J., 2013. Citation time window choice for research impact evaluation. Scientometrics 94 (3), 851–872, http://dx.doi.org/10.1007/s11192-012-0775-9.

Wang, J., 2014. Unpacking the Matthew effect in citations. J. Informetr. 8 (2), 329–339, http://dx.doi.org/10.1016/j.joi.2014.01.006.

Wang, J., Berzins, K., Hicks, D., Melkers, J., Xiao, F., Pinheiro, D., 2012. A boosted-trees method for name disambiguation. Scientometrics 93 (2), 391–411, http://dx.doi.org/10.1007/s11192-012-0681-1.

Wang, J., Hicks, D., 2015. Scientific teams: self-assembly, fluidness, and interdependence. J. Informetr. 9 (1), 197–207, http://dx.doi.org/10.1016/j.joi.2014.12.006.

Wang, J., Thijs, B., Glänzel, W., 2015. Interdisciplinarity and impact: distinct effects of variety, balance, and disparity. PLoS ONE 10 (5), e0127298, http://dx.doi.org/10.1371/journal.pone.0127298.

Weick, K.E., 1976. Educational organizations as loosely coupled systems. Adm. Sci. Q. 21 (1), 1–19, http://dx.doi.org/10.2307/2391875.

Whitley, R., 2000. The Intellectual and Social Organization of the Sciences, 2nd ed. Oxford University Press, Oxford, UK/New York, NY.

Wong, S.S., 2004. Distal and local group learning: performance trade-offs and tensions. Organ. Sci. 15 (6), 645–656, http://dx.doi.org/10.1287/orsc.1040.0080.

Woodman, R.W., Sawyer, J.E., Griffin, R.W., 1993. Toward a theory of organizational creativity. Acad. Manage. Rev. 18 (2), 293–321, http://dx.doi.org/10.2307/258761.

Wooldridge, J.M., 1999. Distribution-free estimation of some nonlinear panel data models. J. Econom. 90 (1), 77–97, http://dx.doi.org/10.1016/s0304-4076(98)00033-5.

Wuchty, S., Jones, B.F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. Science 316 (5827), 1036–1039, http://dx.doi.org/10.1126/science.1136099.

Zuckerman, H., 1967. Nobel laureates in science: patterns of productivity, collaboration, and authorship. Am. Sociol. Rev. 32 (3), 391–403.