

Scientific Teams: Self-assembly, Fluidness, and Interdependence*

Jian Wang^{1,2} & Diana Hicks³

¹University of Leuven

²Institute for Research Information and Quality Assurance (iFQ)

³Georgia Institute of Technology;

Email: jian.wang@kuleuven.be, dhicks@gatech.edu

December 18, 2014

Abstract

Science is increasingly produced in collaborative teams, but collaborative teams in science are self-assembled and fluid. Such characteristics call for a network approach to account for external activities responsible for team product but taking place beyond closed team boundaries in the open network. Given such characteristics of collaborative teams in science, we empirically test the interdependence between collaborative teams in the same network. Specifically, using fixed effects Poisson models and panel data of 1,310 American scientists' life-time publication histories, we demonstrate knowledge spillovers from new collaborators to other teams not involving these new collaborators. Our findings have important implications for studying the organization of science.

Keywords: organization of science; team science; social networks; research collaboration

*Jian Wang and Diana Hicks. (2015). Scientific teams: self-assembly, fluidness, and interdependence. *Journal of Informetrics*, 9(1), 197–207. <http://dx.doi.org/10.1016/j.joi.2014.12.006>
©2014 Elsevier Ltd.

The authors thank Julia Melkers, John Walsh, Juan Rogers, Yajun Mei, Sybille Hinze, and two anonymous reviewers for their very helpful comments and suggestions. 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 bibliometric 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).

1 Introduction

Scientific knowledge is increasingly created collaboratively, as reflected in the increasing share of coauthored papers and the growing size of collaborative teams (de Solla Price, 1986; Hicks and Katz, 1996; Wuchty et al., 2007). This increase in collaboration is driven by a variety of factors, such as the importance of interdisciplinary research questions, growing specialization and the consequent gains from division of labor, new information and communication technologies, and the need to develop and access large shared equipment and large databases (de Solla Price, 1986; Katz and Martin, 1997; Stephan, 2012). Because of the prevalence of collaboration in science, the collaborative team has become an important unit of analysis for studying the organization of science. Accordingly, scholars have extended science studies and laboratory ethnographies from lab benches to *collaboratories* (Chompalov et al., 2002; Cummings and Kiesler, 2005; Shrum et al., 2001). In addition, by bringing in insights from the psychology literature on small groups and the sociology literature on work organizations, scholars have made significant contributions to the *science of team science* (Falk-Krzesinski et al., 2010; Fiore, 2008; Stokols et al., 2008). For example, scholars have investigated the structure and process of collaborative teams, and their effects on team productivity and creativity (Cummings et al., 2013; Hemlin et al., 2013; Lee et al., 2014; Levine and Moreland, 2004).

Before this recent emergence of the science of team science, research on teams has a long history in the social psychology literature (Guzzo and Dickson, 1996; Levine and Moreland, 1990; McGrath and Kravitz, 1982). The management literature has also investigated teams extensively in the corporate world, as cross-functional project teams are increasingly adopted by companies to couple with the rapidly changing and competitive environment (Denison et al., 1996; Keller, 2001; Pearce and Ensley, 2004).

Teams in the real-world operate in various settings and undertake different structures and dynamics. Therefore, scholars have updated the traditional view of teams. For example, the *external approach* emphasizes team behavior directed toward the ex-

ternal environment, rather than assuming teams as isolated and independent from the external environment (Ancona, 1990; Ancona and Caldwell, 1988, 1992). The *virtual team* literature investigates teams with geographically dispersed members and predominantly coordinated via electronic media (Hertel et al., 2005; Lipnack and Stamps, 1997; Townsend et al., 1998), as opposed to the earlier view of teams as small, collocated, and cohesive entities with intense face-to-face interactions. Research on open source software teams highlights that these teams are voluntary-based and fluid, different from the traditional teams with relatively clear and stable boundaries, functions, roles, and norms (Aime et al., 2014; Bagozzi and Dholakia, 2006; Hertel et al., 2003).

Wellman (1997) argued that some teams are tightly-bound, clearly-delimited and densely-knit, while other teams ramify like an expanding spider’s web. The latter is more suitably studied as an open network instead of as a traditional team with clear and stable boundaries. Collaborative teams in science clearly fall in the second category. Following Wellman (1997), this paper argues that the network approach is especially important for studying the organization of science. Specifically, this paper first discusses special characteristics of teams in science, that is, teams in science are self-assembled and fluid, and therefore it’s important to incorporate a network approach when studying scientific teams. Contributing to the understanding of the complexity in scientific teams, we develop empirical strategy to test the dependence of teams on networks, specifically, the knowledge spillover across coauthor teams linked by a single scientist. The *organizational learning* literature has long emphasized newcomers as important sources of innovation in organizations (Gupta et al., 2006; March, 1991; Perretti and Negro, 2006), we further hypothesize that a scientist’s new collaborators would also contribute to the citation performance of his/her other papers not coauthored with these new collaborators. The confirmation of this hypothesis provides empirical evidence of the dependence of collaborative teams on other scientists and teams, and such dependence has important implications for investigating the organization of science.

2 Teams in Science: Definition and Characteristics

The literature has defined teams in various ways. Following Guzzo and Dickson (1996), we define a *team* as “made up of individuals who see themselves and who are seen by others as a social entity, who are interdependent because of the tasks they perform as members of a group, who are embedded in one or more larger social systems (e.g. community, organization), and who perform tasks that affect others (such as customers or coworkers).” (p. 308-309). More specifically, we focus on the ad hoc teams in science assembled for a specific project, with members from different organizations and locations. Such teams differ from formal academic organizations such as departments or laboratories, which are sometimes referred to as teams. Ad hoc project teams in science have unstable memberships and ill-defined boundaries, causing a lot of problems for those attempting to operationalize teams for empirical studies (Haeussler and Sauermann, 2013; Katz and Martin, 1997; Laudel, 2002). Therefore, we adopt two definitions in this paper: (1) a *collaborative team* defined broadly as the group of researchers contributing to a scientific project, which is often hard to identify since some members may not be listed as authors of the final publication, and (2) a *coauthor team* defined narrowly as the group of authors taking credits for the publication. We will discuss the gap between these two definitions and how a network approach is important to mitigate such problems in team studies.

Science has many distinct features compared with other systems of work organization (Whitley, 2000), and collaborative teams in science have many special characteristics which may challenge the traditional view of teams. One important characteristic distinguishing science from other systems of work organization is its autonomy and self-governance. Collaborative teams in science are largely voluntary and based on mutual interests, and scientists have substantial autonomy to create, maintain, restructure, and dissolve their collaborative teams. This paper focuses on two characteristics: self-assembly and fluidness. First, collaborative teams in science are self-assembled. In

traditional team studies, teams are assembled first (and most likely assembled by managers or researchers rather than by team members themselves) and then undertake the complete group process which results in the final team outcome, so researchers can treat team characteristics as exogenous variables and investigate their effects on team performance (Cohen and Bailey, 1997; Hulsheger et al., 2009; Stewart, 2006). However, for self-assembled teams nothing is exogenous (Contractor, 2013; Lungeanu et al., 2014; Zhu et al., 2013). Projects are both the result and the cause of the team (Zhu et al., 2013). In other words, the process of idea generation precedes or co-evolves with the process of team assembly, and the creative process underlying the final team product starts in the open network before the team is assembled.

Second, collaborative teams in science are fluid, that is, with ill-defined boundaries and unstable memberships. Because of the autonomy in organizing teams, people constantly come and go. For example, collaborative teams may acquire new members when new expertise is needed, and team members may leave the team as they no longer share the common interest with other teammates. To some extent, a collaborative team is co-evolving with the project, and there is rarely a stable team seeing through the whole creative process, such as idea generation, convergence, and implementation (Hackman and Morris, 1975; Levine and Moreland, 2004; Skilton and Dooley, 2010). This fluidness is also magnified by another feature that distinguishes science from other types of working environment, that is, high task uncertainty and continual novel production (Whitley, 2000). The course of the project may change at any point, and collaborative teams may reorganize themselves accordingly.

Temporal instability of memberships creates fuzzy team boundaries. At any point in time, there might be a group of scientists with a mutual understanding that they are working together for a specific project, so there is a team boundary, and the project responsibility resides within the team. However, collaborative teams are rarely closed or exclusive. Instead, they interact with the external environment seeking advice from other

scientists and remain open to the possibility of reorganizing the team. Some scientists with indispensable expertise might be invited to join the team, and some enthusiastic external advisors may have made such significant contributions to the project that they are admitted as official team members. On the other hand, some members who left the team in the process and external advisors who haven't made sufficient contribution to become an official member are not listed as authors taking credit for the final output. In other words, the actual organization (i.e., the collaborative team) responsible for the final output is an open network consisting of researchers with different levels of engagements and contributions.

Therefore, it's not always easy to cut out a closed group (i.e., the coauthor team) to take credit and responsibility for the final project outcome. This issue is evidently reflected in the difficulty of determining authorships for scientific publications and operationalizing collaborative teams as coauthor teams (Haeussler and Sauermann, 2013; Katz and Martin, 1997; Laudel, 2002; Melin and Persson, 1996). Because collaborative teams in science have unstable and ill-defined boundaries, and the team production of science is not isolated from, but constantly exchanges with, the external network of other scientists and teams, so these external activities in the open network provide an additional explanation for team performance, in addition to within-team factors.

3 Interdependence between Teams

The main message from the preceding section is that both team assembly and team operation are deeply embedded in open networks, and such dependence on the network should be accounted for when studying teams in science. In this section, we develop empirical strategies to directly test the dependence of teams on networks, specifically, the knowledge spillover across coauthor teams linked by a single scientist.

In addition to the fact that collaborative teams in science are self-assembled and fluid,

scientists also simultaneously participate in multiple collaborative teams which may share multiple common members and similar research agendas. Therefore, something learned in one project might be applied soon after in another project by a person participating in two projects, in other words, teams are interdependent. Since teams in science are not independent or isolated, such interdependence with the external environment and the external activities contributing to the team output should be taken into account in order to better understand the team production of science.

Intellectual and other capital carried by members are important inputs to the team (Amabile, 1983; Ford, 1996; Woodman et al., 1993), and undoubtedly team members can bring in lessons learned from previous team experiences to new situations (Ancona, 1990; Gino et al., 2010; Nonaka, 1994; Reagans et al., 2005). For example, a scientist may take a novel idea from one team and implement it in another team, use methods developed in one team to help problem-solving in another, and start new lines of research with a new team based on knowledge learned from an old one. By tracking the turnover of keywords, Tang and Hu (2013) showed that scholars pick up new research streams from their international collaborators and further pursue these new streams in their domestic collaborations.

In the literature on small groups, some scholars have long advocated an external approach to study team behavior directed toward the external environment and its impact on group performance (Ancona, 1990; Ancona and Caldwell, 1988, 1992). In these studies, the focus of external activities is about managing external dependence and obtaining critical resources (Pfeffer and Salancik, 1978), for example, obtain technical information, map resources, support, and trends in organizations, to influence those individuals with key resources, and to synchronize work flow (Ancona and Caldwell, 1988). In contrast, Mathieu et al. (2001) proposed the concept of multiteam systems (MTSs) defined as “two or more teams that interface directly and interdependently in response to environmental contingencies toward the accomplishment of collective goals” (p. 289). This

MTSs perspective emphasizes that teams are embedded in a network of teams and operate collectively for their common goal. Some studies further showed that inter-team processes explain MTS performance in addition to intra-team processes (DeChurch and Marks, 2006; Marks et al., 2005). In the world of scientific collaboration, it is possible that many teams form a multi-team system in the pursuit of a large scale project, and each team is specialized and responsible for a sub-project, but this is not always the case. In other words, the interdependency between collaborative teams in science might not always be as high as described in the MTSs approach but is clearly much higher than acknowledged in the external approach. The preceding section has discussed the idea that teams co-evolve with the project and that the creative process underlying the final scientific product is not constrained within the team boundary but intertwined with external processes. Therefore, these external processes should be accounted for when studying the organization of science.

To empirically test the interdependency between collaborative teams linked through membership of a single scientist, we test the knowledge spillover from one type of coauthor team to another. Note that while the preceding theoretical discussions mainly adopted the broad definition of *collaborative teams*, the empirical part will mainly adopt the narrow definition of *coauthor teams*, and discuss the interdependence that comes directly from the fact that collaborative teams have fuzzy boundaries and that coauthor teams do not capture all the actions responsible for the final product. Specifically, we argue that when a scientist has more new coauthors, his/her other coauthor teams not involving these new coauthors will have better performance in terms of citation impact. The organizational learning literature has long emphasized newcomers as important sources of innovation in organizations: newcomers are more likely to explore new knowledge, while old-timers are more likely to exploit exiting knowledge (Gupta et al., 2006; March, 1991; Perretti and Negro, 2006). Consequently, there is a negative association between repeated collaboration and creativity in science (Guimera et al., 2005; Porac et al., 2004).

This paper does not intend to detangle the difference between new and repeated collaborations, but takes advantage of this difference to test the knowledge spillovers between them. Newcomers are important sources of innovation for an organization because they are more likely to bring in different knowledge and perspectives that are not yet shared in the organization (Gupta et al., 2006; March, 1991; Perretti and Negro, 2006). We further argue that a scientist can transfer the new knowledge and perspectives that he learned from his/her new coauthor to his/her other coauthor teams not involving these new coauthors. Another reason that newcomers are important sources of innovation is that they are not constrained by inert and shared mental models emerging from repeated collaborating experience. The mental model shapes not only the way that individuals explain, predict, and describe events but also the way that the team differentiates roles among members (Skilton and Dooley, 2010). However, newcomers do not share such mental models, and the new team composition may call for an updated or a completely new mental model. Here we further argue that exposure to different perspectives may also help a scientist to break the mental model governing his/her other teams not directly “disturbed” by newcomers. In summary, there will be knowledge spillovers from new coauthor teams to repeated coauthor teams. Furthermore, a larger number of new coauthors will provide more diverse knowledge and perspectives and therefore will have higher positive effect on the performance of repeated coauthor teams. Using the same logic, more repeated coauthors would also suffocate creativity in other teams not direct involving these old-timers. Therefore, we hypothesize that

Hypothesis 1: when a scientist has more new coauthors, his/her other coauthor teams not involving these new coauthors will produce papers with higher impact.

Hypothesis 2: when a scientist has more repeated coauthors, his/her other coauthor teams not involving these repeated coauthors will produce papers with lower impact.

4 Data and Methods

4.1 Data

To test our hypotheses, we used fixed effects models and unbalanced panel data of American scientists' life-time publication histories. The sample of American scientists came from a project funded by the United States National Science Foundation (NSF). It surveyed 3,677 stratified randomly sampled American scientists (i.e., egos) in 2007 in six disciplines: biology, chemistry, computer science, earth and atmospheric sciences, electrical engineering, and physics. The random sample was stratified by gender, 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 1,774 completed surveys, 176 were removed because of ineligible rank or discipline, resulting in a final total sample size of 1,598. 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 gender, rank, and discipline are very similar to the survey population.

Life-time publication data for these 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 name disambiguation algorithm documented in Wang et al. (2012). Coauthor names were also cleaned and disambiguated to identify unique coauthors. The publication data were 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 1,323 scientists in the remaining five disciplines for analysis. Out of these 1,323

scientists, 1,310 scientists published 43,996 WoS journal articles in total.

For each of the 43,996 papers, its number of forward citations in a five-year time window was retrieved from WoS, that is, for a paper published in year t , citations received by this paper from year t to $t+4$ were counted. Citation ageing pattern differs across papers: many highly cited papers take a long time to establish themselves as elite papers, while many other papers have very early citation peaks (Glänzel et al., 2003; Rogers, 2010; Van Raan, 2004), so a sufficient citation 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. We did not exclude author self-citations because that they are an organic part of the citation process (Glänzel et al., 2004), and also because of the difficulty and potential errors in matching citing-author and cited-author names.

For this analysis, we further imposed some sample restrictions. First, observations before 1983 were excluded because our complete WoS data start in 1980, and we classified coauthors into new- or repeated-coauthors based on coauthoring history in the preceding three years. Second, if the ego’s first paper was published in year t , then this ego’s observations before (not including) year $t+3$ were excluded for the same reason. Third, observations after 2007 were excluded because these papers do not have a full five-year period to accumulate citations. We are left with 11,850 observations of 1,310 egos.

4.2 Measures

We first classified, in each year and for each ego, coauthors into two types: new (not coauthored with the ego in the preceding three years) and repeated (coauthored at least once with the ego in the preceding three years). Correspondingly, papers of each ego in each year were classified into four types: *solo* (single-authored paper), *new* (coauthored

with only new coauthors), *repeated* (coauthored with only repeated coauthors), and *mixed* (coauthored with both new and repeated coauthors).

Dependent variables. To test the knowledge spillover from new to repeated coauthor teams, we use the average number of citations per paper for repeated coauthor team papers (*Citations.AVG.REP*). To test the constraining effect from repeated to new coauthor teams, we use the average number of citations per paper for new coauthor team papers (*Citations.AVG.NEW*).

Independent variables. The number of new coauthors or the number of projects involving new coauthors were counted: (1) the number of new coauthors (*Coauthors.NEW*) and (2) the number of projects/papers involving new coauthors, that is, the number of new- and mixed- coauthor team papers (*Pubs.NEW&MIX*). Similarly, *Coauthors.REP* and *Pubs.REP&MIX* were constructed for the number of repeated coauthors and the number of projects involving repeated coauthors. The natural logarithms of these independent variables are roughly normally distributed and were used in regressions.

Control variables. 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. First, following Lee and Bozeman (2005), we used ego’s career age, defined as the number of years after receiving Ph.D., to control for both age and experience effects. Both career age and career age squared were incorporated to account for the nonlinear trajectory of research performance over the life cycle (Cole, 1979; Lee and Bozeman, 2005; Stephan and Levin, 1991). We also tried with physical age instead of career age and got consistent results. Furthermore, both the number of citations (the dependent variables) and the number of coauthors (the focal independent variables) are likely to be correlated with the ego’s previous performance, in other words, successful history breeds further success and also attracts more coauthors. To address this endogeneity issue, ego’s citation performance in the preceding year was also controlled for, that is, the average number of citations received by all papers of the

ego published in the preceding year (*Citations.AVG.LAG*). Morgan and Winship (2007) warned of potential issues when incorporating a lagged dependent variable. In our case, strictly speaking, *Citations.AVG.LAG* is not a lagged dependent variable because it measures overall citation impact of the ego (not new/repeated coauthor team papers separately corresponding to the dependent variables). Our theoretical argument is also very clear: previous success is the mutual cause for both the current citation impact and the current number of coauthors. Therefore, we directly control for such an observable confounder, instead of attempting to decompose such confounder (when unobservable) into a lagged dependent variable and two separate unobserved variables: one affects both the current number of coauthors and the lagged citation performance, and the other affects both the lagged and the current citations. Therefore, we avoided the problem explained by Morgan and Winship (2007). The log of this control variable was used in regressions. Descriptive statistics and correlations are reported in Table 1.

Table 1: Descriptive statistics and correlations

	Variables	N	Mean	S.D.	Min	Max	1	2	3	4	5	6	7
1	Citations.AVG.REP	4556	12.28	17.15	0	504							
2	Citations.AVG.NEW	6815	12.81	32.07	0	1741	.29						
3	Coauthhors.NEW	17202	3.27	7.03	0	328	.14	.22					
4	Pubs.NEW&MIX	17202	1.35	1.80	0	22	.12	.11	.94				
5	Coauthors.REP	17202	1.61	4.04	0	173	.12	.11	.54	.59			
6	Pubs.REP&MIX	17202	1.14	1.94	0	25	.09	.11	.55	.61	.97		
7	Citations.AVG.LAG	17202	7.68	20.53	0	1741	.36	.23	.28	.30	.40	.40	
8	Career age	17151	14.07	9.42	-10	54	-.06	-.09	.11	.12	.14	.13	.09

5 Results

5.1 Fluidness of Coauthor Teams

Before reporting the regression results for the interdependence between coauthor teams, we first use some descriptive statistics to illustrate the fluidness of coauthor teams in science. Based on the sample of papers published between 2005 and 2007, that is, 7,678

papers of 1,148 egos, we retrieve coauthor teams associated with these egos, and classify these coauthor teams based on size: doublet (two members), triplet (three members), quartet (four members), and so on. Subsequently we evaluate the stability of these coauthor teams. For each coauthor team, the number of papers that were published by exactly the same group of authors was counted. Table 2 reports how many coauthor teams were repeated, that is, having more than one paper. In addition, we also count how many teams also existed in some slightly different forms: (1) teams that have papers authored by all the team members plus some other scientists and (2) teams that have papers authored by a subset of the team (not all team members) and possibly (but not necessarily) plus some other scientists. However, this second count is actually underestimated in this paper, because our data are egocentric data, that is, all papers in our sample have to include the ego. Therefore, for a coauthor team, it is possible that a subset of the team has their own collaborations without the ego, but these papers are not observed in our data. The implication is that the actual fluidness indicated by the second count is actually even higher than reported in Table 2. Note that Table 2 reports teams with up to six members, while singlet and teams with more than six members are left out.

Table 2: Fluidness of coauthor teams

	Doublet	Triplet	Quartet	Quintet	Sextet
Total number of teams	1169	1460	1404	891	601
Number of teams which have more than one paper (repeated teams)	228 (20%)	133 (9%)	57 (4%)	31 (3%)	13 (2%)
Number of teams which have paper(s) with someone outside of the team	517 (44%)	256 (18%)	139 (10%)	56 (6%)	30 (5%)
Number of teams which have paper(s) by the ego, a subset of the team, and maybe also someone outside of the team		917 (63%)	991 (71%)	684 (77%)	462 (77%)
Maximum number of papers of the team	16	6	4	3	3

Percentages in parentheses.

As shown in Table 2, coauthor teams are mostly one-time phenomena and do not repeat. About 20% of the doublets have repeated, that is, have more than one paper, and only about 9% of the triplets and 4% of the quartets have repeated. As the team

size increases, the percentage of repeated teams decreases. Although coauthor teams do not repeat in the exact form, they do frequently “repeat” in a slightly different form. Take triplets as an example, 18% of them have paper(s) authored by the whole team and some other scientist(s), and more than 63% of them have paper(s) authored by a subset of the team and some other scientist(s).

5.2 Interdependence between Coauthor Teams

Because our dependent variables are over-dispersed count variables, we use Poisson regression with robust standard errors, following previous literature (Hall and Ziedonis, 2001; Hottenrott and Lopes-Bento, 2014; Somaya et al., 2007). 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 adopt the Poisson model with robust standard errors in our empirical analysis. Furthermore, we incorporate individual fixed effects to account for time-invariant individual heterogeneities, so that within-ego effects are estimated. Such fixed effects Poisson models can be fitted by conditioning out the individual fixed effects (Hausman et al., 1984). Specifically, we use the *xtpoisson* command in *STATA* (StataCorp, 2013b), which implements the formula as presented in Wooldridge (1999). In addition, because of the exponential specification in the Poisson model (i.e., describing the log of the dependent variable by a linear equation of independent variables), we take natural logarithm to transform all the independent count variables, so that both dependent and independent variables are scaled in the same way. As a result, the coefficients can be interpreted as elasticity (except for career age and squared, which are not log transformed), that is, a percentage change in the dependent variable corresponding to a percentage change in the independent variable, *ceteris*

paribus.

Table 3: Fixed effects Poisson models: interdependence between coauthor teams

	Citations.AVG.REP		Citations.AVG.NEW	
	(1)	(2)	(3)	(4)
Coauthors.NEW (ln)	0.0426*			
	(0.0252)			
Pubs.NEW&MIX (ln)		0.0753*		
		(0.0419)		
Coauthors.REP (ln)			-0.0762*	
			(0.0413)	
Pubs.REP&MIX (ln)				-0.0747
				(0.0582)
Citations.AVG.LAG (ln)	0.0451**	0.0444**	-0.0343*	-0.0357**
	(0.0219)	(0.0219)	(0.0179)	(0.0179)
Career age	-0.0068	-0.0071	-0.0179	-0.0182
	(0.0121)	(0.0119)	(0.0119)	(0.0125)
Career age squared	0.0002	0.0003	0.0009***	0.0009***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Ego fixed effects	YES	YES	YES	YES
Log pseudolikelihood	-21632	-21622	-44698	-44718
χ^2	10**	10**	26***	24***
N obs	4307	4307	6674	6674
N egos	738	738	1053	1053

Cluster-robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Two-tailed tests

Regression results are reported in Table 3. Panel data are used for regressions, so one observation is an ego-year. Take Table 3 column 1 as an example, the dependent variable would be the average number of citations received by repeated-coauthored papers of ego i in year t , and the focal independent variable is the log number of new coauthors of ego i in year t . In addition, egos with only one observation or no within-ego variations are automatically dropped out from the regression, so the number of groups and observations are smaller than the total number of observation without missing values.

As fixed effects models are implemented, coefficients report within-individual effects, but not between-ego effects. We observe significant positive effect from the number of new coauthors to repeated coauthor team papers. For the same scientist, as the number

of new coauthors increases by 1%, the average number of citations received by repeated coauthor team papers increases by 0.04%. As the number of projects/papers involving new coauthors increases by 1%, the average number of citations received by repeated coauthor team papers increases by 0.08%. On the other hand, as the number of repeated coauthors increases by 1%, the average number of citations received by new coauthor team papers decreases by 0.08%, while the number of projects/papers involving repeated coauthors has no significant effect on the citation impact of new coauthor team papers. Overall, empirical results confirm our first hypothesis about the positive knowledge spillover effects from new to repeated coauthor teams and partially support our second hypothesis about the negative constraining effect from repeated to new coauthor teams.

5.3 Alternative Explanations

An alternative explanation for our findings is that successful scientists are more likely to attract new potential coauthors, and these new coauthors would crowd out repeated ones. Therefore, we would observe that the citation performance in general (regardless of new or repeated coauthor team papers) is positively correlated with the number of new coauthors but negatively correlated with the number of repeated coauthors. We believe that we have already addressed this issue by incorporating the average citations in the preceding period as a control variable. To further investigate this alternative explanation, we run another set of regressions, using the same model specification but replacing the dependent variables by the number of new/repeated coauthors. Regression results are reported in Table 4. First, we observe no evidence of crowding-out effects. Table 4 column 1 shows that, as the number of new coauthors increases by 1%, the number of repeated coauthors increases by 0.24%. Second, we observe no evidence that, as scientists become more successful, they have more new coauthors and fewer repeated coauthors. On the contrary, more successful scientists initiate fewer new coauthors and retain more repeated coauthors. This makes sense, previous success may motivate scientists to

keep and reinforce the existing collaborative ties instead of seeking new collaborators. However, under such circumstances, although not motivated to, when successful scientists actually do establish more new collaborative ties, their repeated coauthor teams would achieve higher citation impact. Combining Table 3 and 4 provides stronger support for our hypotheses.

Table 4: Fixed effects Poisson models: crowding-out effects

	Coauthors.REP		Coauthors.NEW	
	(1)	(2)	(3)	(4)
Coauthors.NEW (ln)	0.3632*** (0.0297)			
Pubs.NEW&MIX (ln)		0.7459*** (0.0298)		
Coauthors.REP (ln)			0.4698*** (0.0312)	
Pubs.REP&MIX (ln)				0.7328*** (0.0319)
Citations.AVG.LAG (ln)	0.1785*** (0.0143)	0.1654*** (0.0146)	-0.0570*** (0.0167)	-0.0627*** (0.0155)
Career age	0.0626*** (0.0068)	0.0577*** (0.0062)	0.0728*** (0.0067)	0.0668*** (0.0069)
Career age squared	-0.0006*** (0.0002)	-0.0005*** (0.0001)	-0.0010*** (0.0002)	-0.0008*** (0.0002)
Ego fixed effects	YES	YES	YES	YES
Log pseudolikelihood	-20796	-20069	-43806	-42401
χ^2	1020***	1257***	816***	1047***
N obs	15950	15950	17003	17003
N egos	1096	1096	1202	1202

Cluster-robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Two-tailed tests

In addition, following much prior work in this area (Guimera et al., 2005; Simonton, 2004; Uzzi et al., 2013), our empirical design uses citations to evaluate the performance of science production. However, citations depend on not only the intrinsic quality of the paper but also the knowledge diffusion process which is not under the control of the authors (Bornmann and Daniel, 2008; Gilbert, 1977; Whitley, 2000). Therefore, there is an alternative explanation for the positive association between the number of new coauthors

and the number of citations received by repeated coauthor team papers: having more new coauthors helps disseminate the work of repeated coauthor team papers. However, Wang (2014) demonstrated that there is no retroactive networking effect (i.e., the number of coauthors of the ego) on citations to solo-authored papers. In addition, if this alternative explanation is true, then we should also observe a significant positive association between the number of repeated coauthors and the number of citations received by new coauthor team papers, because repeated coauthors should also contribute to the diffusion of new coauthor team papers. However, this is not supported by our data (Table 3).

5.4 Robustness and Limitations

We used the conditional fixed effects Poisson models and incorporated robust standard errors to deal with the over-dispersion. Given that the negative binomial model is also commonly used in the literature, we also tried with negative binomial models. Considering the nature of our research question, we incorporated ego fixed effects to account for unobserved heterogeneity among individual, so that we can estimate within-individual effects, e.g., how does citation rate of repeated coauthor papers change as the number of new coauthors changes, for the same individual rather than between different individuals. In a least squares framework, consistent estimates can be obtained by incorporating a set of individual dummies, but such strategy is not available for nonlinear models such as Poisson or negative binomial. For estimating within-individual effects, the conditional maximum likelihood is developed for Poisson models which conditions out the individual fixed effects (Cameron and Trivedi, 2013; Hausman et al., 1984; Wooldridge, 1999), and this method is implemented in the *xtpoisson* function in *STATA* (StataCorp, 2013b). 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; Guimaraes, 2008). To the best of our knowledge, we are not aware of any statistics software providing a true fixed effect negative binomial solution. Therefore, we implement the following two imperfect strategies as robustness checks: (1) standard negative binomial models with the complete set of individual dummies, and (2) the *xtnbreg* model grouped by individuals. Both yielded consistent results.

In addition, as our study is based on bibliometric data, we cannot avoid some fundamental limitations of the bibliometric data. For example, our empirical analysis is based on coauthorship data, but not all collaborations result in coauthored papers and not all coauthorships embody collaborations (Haeussler and Sauermann, 2013; Katz and Martin, 1997; Laudel, 2002; Melin and Persson, 1996). In addition, we do not observe collaborative teams that failed to publish papers, and such selection biases may cause errors in our retrieval of coauthor teams and coding of new/repeated coauthors. These are important limitations of our study.

6 Discussion

Science is increasingly produced by self-assembled and fluid collaborative teams, which calls for an open approach to studying them. Scientists assemble teams based on the needs of the project, and project idea generation precedes or co-evolves with the process of team assembly, in other words, the creative process underlying a scientific production is initiated before the intra-team process starts. Therefore, sources of creativity outside the team should be accounted for. Collaborative teams in science also have unstable memberships and fuzzy boundaries. Collaborative teams have intense exchanges with other scientists and teams, and a considerable amount of the activities responsible for the final team product take place in the scientific community outside of the team. Such external activities should be taken into account in explaining team performance, in addition to

intra-team activities. This paper empirically tests the dependence of collaborative teams on people outside the team. Specifically, we demonstrate positive knowledge spillover effects from new to repeated collaborations.

Our findings have important implications for the study of science production following a team approach. Given that both team assembly and operation are embedded in a wider community, and teams sharing members are interdependent, scholars should account for such external activities responsible for the team performance, in order to better understand team processes and sources of team creativity.

This study suggests the organization of science must be studied at multiple levels simultaneously. The creative process in science production is not bounded within closed teams but takes place in a fluid, evolving environment. Idea generation (finding and defining the problem) precedes team assembly. Even after a team is assembled, the team keeps interacting with other scientists and teams to tap into external sources of knowledge. Furthermore, because of the uncertainty in scientific research, the initial idea behind the team is constantly subject to change, so is the team, adding new expertise and resources or removing exiting ones. This dynamic process leads to the observed instability of teams as described in this paper. Furthermore, in this dynamic process, teams have intensive exchange with other scientists and teams. Therefore, the boundaries of teams are fuzzy, and teams are interdependent. Scientific teams must be studied not in isolation, but as contributors to and beneficiaries of a larger environment, the scientific community.

References

Aime, F., Humphrey, S., Derue, D. S., and Paul, J. B. (2014). The riddle of heterarchy: Power transitions in cross-functional teams. *Academy of Management Journal*, 57(2):327–352.

- Allison, P. D. and Waterman, R. P. (2002). Fixed-effects negative binomial regression models. *Sociological Methodology*, 32(1):247–265.
- Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, 45(2):357–376.
- Ancona, D. G. (1990). Outward bound: Strategies for team survival in an organization. *Academy of Management Journal*, 33(2):334–365.
- Ancona, D. G. and Caldwell, D. F. (1988). Beyond task and maintenance: Defining external functions in groups. *Group & Organization Studies*, 13(4):468–494.
- Ancona, D. G. and Caldwell, D. F. (1992). Bridging the boundary: External activity and performance in organizational teams. *Administrative Science Quarterly*, 37(4):634–665.
- Bagozzi, R. P. and Dholakia, U. M. (2006). Open source software user communities: A study of participation in linux user groups. *Management Science*, 52(7):1099–1115.
- Bornmann, L. and Daniel, H.-D. (2008). What do citation counts measure? a review of studies on citing behavior. *Journal of Documentation*, 64(1):45–80.
- Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*. Cambridge University Press, Cambridge ; New York, NY, second edition. edition.
- Chompalov, I., Genuth, J., and Shrum, W. (2002). The organization of scientific collaborations. *Research Policy*, 31(5):749–767.
- Cohen, S. G. and Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, 23(3):239–290.
- Cole, S. (1979). Age and scientific performance. *American Journal of Sociology*, 84(4):958–977.

- Contractor, N. S. (2013). Some assembly required: leveraging web science to understand and enable team assembly. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):1–14.
- Cummings, J. N. and Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science*, 35(5):703–722.
- Cummings, J. N., Kiesler, S., Zadeh, R. B., and Balakrishnan, A. D. (2013). Group heterogeneity increases the risks of large group size: A longitudinal study of productivity in research groups. *Psychological Science*, 24(6):880–890.
- de Solla Price, D. (1986). *Little Science, Big Science—and Beyond*. Columbia University Press, New York, NY.
- DeChurch, L. A. and Marks, M. A. (2006). Leadership in multiteam systems. *Journal of Applied Psychology*, 91(2):311–329.
- Denison, D. R., Hart, S. L., and Kahn, J. A. (1996). From chimneys to cross-functional teams: Developing and validating a diagnostic model. *Academy of Management Journal*, 39(4):1005–1023.
- Falk-Krzesinski, H. J., Borner, K., Contractor, N., Fiore, S. M., Hall, K. L., Keyton, J., Spring, B., Stokols, D., Trochim, W., and Uzzi, B. (2010). Advancing the science of team science. *Clinical and Translational Science*, 3(5):263–266.
- Fiore, S. M. (2008). Interdisciplinarity as teamwork: How the science of teams can inform team science. *Small Group Research*, 39(3):251–277.
- Ford, C. M. (1996). Theory of individual creative action in multiple social domains. *Academy of Management Review*, 21(4):1112–1142.
- Gilbert, G. N. (1977). Referencing as persuasion. *Social Studies of Science*, 7(1):113–122.

- Gino, F., Argote, L., Miron-Spektor, E., and Todorova, G. (2010). First, get your feet wet: The effects of learning from direct and indirect experience on team creativity. *Organizational Behavior and Human Decision Processes*, 111(2):102–115.
- Glänzel, W., Schlemmer, B., and 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.
- Glänzel, W., Thijs, B., and Schlemmer, B. (2004). A bibliometric approach to the role of author self-citations in scientific communication. *Scientometrics*, 59(1):63–77.
- Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Applications to poisson models. *Econometrica*, 52(3):701–720.
- Greene, W. (2005). Functional form and heterogeneity in models for count data. *Foundations and Trends in Econometrics*, 1(2):113–218.
- Guimaraes, P. (2008). The fixed effects negative binomial model revisited. *Economics Letters*, 99(1):63–66.
- Guimera, R., Uzzi, B., Spiro, J., and Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308(5722):697–702.
- Gupta, A. K., Smith, K. G., and Shalley, C. E. (2006). The interplay between exploration and exploitation. *Academy of Management Journal*, 49(4):693–706.
- Guzzo, R. A. and Dickson, M. W. (1996). Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology*, 47:307–338.
- Hackman, I. R. and Morris, C. G. (1975). *Group tasks, group interaction processes, and group performance effectiveness: A review and proposed integration*, pages 47–99. Academic Press, New York, NY.

- Haeussler, C. and Sauermann, H. (2013). Credit where credit is due? the impact of project contributions and social factors on authorship and inventorship. *Research Policy*, 42(3):688–703.
- Hall, B. H. and Ziedonis, R. H. (2001). The patent paradox revisited: An empirical study of patenting in the us semiconductor industry, 1979-1995. *Rand Journal of Economics*, 32(1):101–128.
- Hausman, J., Hall, B. H., and Griliches, Z. (1984). Econometric models for count data with an application to the patents r&d relationship. *Econometrica*, 52(4):909–938.
- Hemlin, S., Allwood, C. M., Martin, B. R., and Mumford, M. D. (2013). *Creativity and Leadership in Science, Technology, and Innovation*. Routledge studies in innovation, organization and technology. Routledge, New York, NY.
- Hertel, G., Geister, S., and Konradt, U. (2005). Managing virtual teams: A review of current empirical research. *Human Resource Management Review*, 15(1):69–95.
- Hertel, G., Niedner, S., and Herrmann, S. (2003). Motivation of software developers in open source projects: An internet-based survey of contributors to the linux kernel. *Research Policy*, 32(7):1159–1177.
- Hicks, D. and Katz, J. S. (1996). Where is science going? *Science, Technology, & Human Values*, 21(4):379–406.
- Hottenrott, H. and Lopes-Bento, C. (2014). Quantity or quality? knowledge alliances and their effects on patenting. *Industrial and Corporate Change*, forthcoming.
- Hulsheger, U. R., Anderson, N., and Salgado, J. F. (2009). Team-level predictors of innovation at work: A comprehensive meta-analysis spanning three decades of research. *Journal of Applied Psychology*, 94(5):1128–1145.

- Katz, J. S. and Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26(1):1–18.
- Keller, R. T. (2001). Cross-functional project groups in research and new product development: Diversity, communications, job stress, and outcomes. *Academy of Management Journal*, 44(3):547–555.
- Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, 11(1):3–15.
- Lee, S. and Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, 35(5):673–702.
- Lee, Y.-N., Walsh, J. P., and Wang, J. (2014). Creativity in scientific teams: Unpacking novelty and impact. *Research Policy*, forthcoming.
- Levine, J. M. and Moreland, R. L. (1990). Progress in small group research. *Annual Review of Psychology*, 41:585–634.
- Levine, J. M. and Moreland, R. L. (2004). Collaboration: The social context of theory development. *Personality and Social Psychology Review*, 8(2):164–172.
- Lipnack, J. and Stamps, J. (1997). *Virtual teams: Reaching across space, time, and organizations with technology*. Wiley, New York.
- Lungeanu, A., Huang, Y., and Contractor, N. S. (2014). Understanding the assembly of interdisciplinary teams and its impact on performance. *Journal of Informetrics*, 8(1):59–70.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1):71–87.
- Marks, M. A., DeChurch, L. A., Mathieu, J. E., Panzer, F. J., and Alonso, A. (2005). Teamwork in multiteam systems. *Journal of Applied Psychology*, 90(5):964–971.

- Mathieu, J., Marks, M., and Zaccaro, S. (2001). *Multi-team systems*, volume 2, pages 289–313. Sage Publications, London, UK.
- McGrath, J. E. and Kravitz, D. A. (1982). Group research. *Annual Review of Psychology*, 33:195–230.
- Melin, G. and Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, 36(3):363–377.
- Morgan, S. L. and Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge University Press, New York, NY.
- Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, 5(1):14–37.
- Pearce, C. L. and Ensley, M. D. (2004). A reciprocal and longitudinal investigation of the innovation process: The central role of shared vision in product and process innovation teams (ppits). *Journal of Organizational Behavior*, 25(2):259–278.
- Perretti, F. and Negro, G. (2006). Filling empty seats: How status and organizational hierarchies affect exploration versus exploitation in team design. *Academy of Management Journal*, 49(4):759–777.
- Pfeffer, J. and Salancik, G. R. (1978). *The External Control of Organizations: A Resource Dependence Perspective*. Harper and Row, New York, NY.
- Porac, J. F., Wade, J. B., Fischer, H. M., Brown, J., Kanfer, A., and Bowker, G. (2004). Human capital heterogeneity, collaborative relationships, and publication patterns in a multidisciplinary scientific alliance: A comparative case study of two scientific teams. *Research Policy*, 33(4):661–678.
- Reagans, R., Argote, L., and Brooks, D. (2005). Individual experience and experience

- working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Science*, 51(6):869–881.
- Rogers, J. (2010). Citation analysis of nanotechnology at the field level: Implications of r&d evaluation. *Research Evaluation*, 19(4):281–290.
- Shrum, W., Chompalov, I., and Genuth, J. (2001). Trust, conflict and performance in scientific collaborations. *Social Studies of Science*, 31(5):681–730.
- Simonton, D. K. (2004). *Creativity in Science: Chance, Logic, Genius, and Zeitgeist*. Cambridge University Press, Cambridge, UK; New York, NY.
- Skilton, P. F. and Dooley, K. J. (2010). The effects of repeated collaboration on creative abrasion. *Academy of Management Review*, 35(1):118–134.
- Somaya, D., Williamson, I. O., and Zhang, X. M. (2007). Combining patent law expertise with r&d for patenting performance. *Organization Science*, 18(6):922–937.
- StataCorp (2013a). xtnbreg.
- StataCorp (2013b). xtpoisson.
- Stephan, P. E. (2012). *How Economics Shapes Science*. Harvard University Press, Cambridge, MA.
- Stephan, P. E. and Levin, S. G. (1991). Inequality in scientific performance: Adjustment for attribution and journal impact. *Social Studies of Science*, 21(2):351–368.
- Stewart, G. L. (2006). A meta-analytic review of relationships between team design features and team performance. *Journal of Management*, 32(1):29–55.
- Stokols, D., Hall, K. L., Taylor, B. K., and Moser, R. P. (2008). The science of team science: Overview of the field and introduction to the supplement. *American Journal of Preventive Medicine*, 35(2, Supplement):S77–S89.

- Tang, L. and Hu, G. (2013). Tracing the footprint of knowledge spillover: Evidence from u.s.china collaboration in nanotechnology. *Journal of the American Society for Information Science and Technology*, 64(9):1791–1801.
- Townsend, A. M., DeMarie, S. M., and Hendrickson, A. R. (1998). Virtual teams: Technology and the workplace of the future. *The Academy of Management Executive*, 12(3):17–29.
- Uzzi, B., Mukherjee, S., Stringer, M., and Jones, B. (2013). Atypical combinations and scientific impact. *Science*, 342(6157):468–472.
- Van Raan, A. (2004). Sleeping beauties in science. *Scientometrics*, 59(3):467–472.
- Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, 94(3):851–872.
- Wang, J. (2014). Unpacking the matthew effect in citations. *Journal of Informetrics*, 8(2):329–339.
- Wang, J., Berzins, K., Hicks, D., Melkers, J., Xiao, F., and Pinheiro, D. (2012). A boosted-trees method for name disambiguation. *Scientometrics*, 93(2):391–411.
- Wellman, B. (1997). *An electronic group is virtually a social network*, pages 179–208. Lawrence Erlbaum Associates, Publishers, Erlbaum, NJ.
- Whitley, R. (2000). *The Intellectual and Social Organization of the Sciences*. Oxford University Press, Oxford, UK; New York, NY, 2nd edition.
- Woodman, R. W., Sawyer, J. E., and Griffin, R. W. (1993). Toward a theory of organizational creativity. *Academy of Management Review*, 18(2):293–321.
- Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, 90(1):77–97.

Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–1039.

Zhu, M., Huang, Y., and Contractor, N. S. (2013). Motivations for self-assembling into project teams. *Social Networks*, 35(2):251–264.