Team assembly mechanisms determine collaboration network structure and team performance

Roger Guimerà[1,†], Brian Uzzi[2,†], Jarrett Spiro[3], and Luís A. Nunes Amaral[1,*]

[1]Department of Chemical and Biological Engineering
Northwestern University, Evanston, IL 60208, USA

[2]Kellogg School of Management and Department of Sociology
Northwestern University, Evanston, IL 60208, USA

[3]Graduate School of Business
Stanford University, Stanford, CA 94305, USA

[†]These two authors contributed equally to this work.
[*]To whom correspondence should be addressed; E-mail:amaral@northwestern.edu

Agents in creative enterprises are embedded in networks that inspire, support, and evaluate their work. Here, we investigate how the mechanisms by which creative teams self-assemble determine the structure of these collaboration networks. We propose a model for the self-assembly of creative teams that is based on three parameters: team size, the fraction of newcomers in new productions and the tendency of incumbents to repeat previous collaborations. The model suggests that the emergence of a large connected community of practitioners can be described as a phase transition. We find that team assembly mechanisms determine both the structure of the collaboration network and team performance, for teams derived from both artistic and scientific fields.

Teams are assembled because of the need to incorporate individuals with different ideas, skills, and resources. Creativity is spurred when proved innovations in one domain are introduced into a new domain, solving old problems and inspiring fresh thinking (*1–4*). However, research shows that the right balance of diversity on a team is elusive. While diversity may potentially spur creativity, it typically promotes conflict and miscommunication (*5–7*). It also runs counter to the security most individuals experience in working and sharing ideas with past collaborators (*8*). Successful teams evolve toward a size that is large enough to enable specialization and effective division of labor among teammates, but small enough to avoid overwhelming costs of group coordination (*9*). Here, we investigate empirically and theoretically the mechanisms by which teams of creative agents are assembled. We also investigate how these *microscopic* team assembly mechanisms determine both the *macro* structure of a creative field, and the success of certain teams in using the resources and knowledge available in the field. We develop a model for the assembly of teams of creative agents. In our model , the selection of the members of a team is controlled by three parameters: the number m of team members, the probability p of selecting incumbents, that is, agents already belonging to the network, and the propensity q of incumbents to repeat past collaborations. The model predicts the existence of two phases which are determined mainly by the values of p and q. In one phase there is a large cluster connecting a significant fraction of the agents, while in the other phase the agents form a large number of isolated clusters.

We analyzed data from both artistic and scientific fields where collaboration needs have experienced pressures such as differentiation and specialization, internationalization, and commercialization (*4, 10, 11*): (i) the Broadway musical industry (*BMI*), and (ii) the scientific disciplines of social psychology, ecology, and astronomy (Table 1). For the *BMI*, we considered all $number$ productions in the period 1877–1990 (*12, 13*). For each of the scientific disciplines, we considered all collaborations that resulted in publications in recognized journals within the fields studied (*14*)—seven social psychology journals, ten ecology journals, and six astronomy journals (Table 2). Collaboration networks (*15–19*) were then built for each of the journals independently, and for the whole discipline by merging the data from the journals within a discipline (see Supporting Material).

The evolution of team sizes in the *BMI* bears out the expectation that team size and composition depend on the intricacy of the creative task. In the period 1877–1929, when the form of the Broadway musical show was still being worked out through trial and error (*12*), there was a steady increase in the number of artists per production, from an average of two to an average of seven (Fig. 1A). This increase in size suggests that teams evolved to manage the complexity of the new artistic form. By the late 20's, the Broadway musical reached the form we know today (*4*). For the following 55 years, a period that includes the Great Depression, World War II, and the post-War Boom, the average size of teams remained around seven (*20*).

We find similar scenarios for the evolution of team size in scientific collaborations. The three fields showed an increase in team size with time (Figs. 1B-D). The increase was approximately linear in social psychology and faster than linear in ecology and astronomy. For social psychology, team size growth rate was greater for high impact versus low impact journals, suggesting that team size not only depends on the intricacy of the enterprise but also that successful teams might adapt faster to external pressures.

The analysis of team size cannot capture the fact that teams are embedded in a larger network (*3*). This complex network (*21–23*), which is the result of past collaborations and the *medium* in which future collaborations will develop, acts as a storehouse for the pool of "knowledge" created within the field. The way the members of a team are embedded in the larger network affects the manner in which they access the knowledge in the field. Therefore, teams formed by individuals with large but disparate sets of collaborators are more likely to draw from a more diverse reservoir of knowledge. At the same time, and for the same reasons, the way teams are organized into a larger network affects how likely it is that breakthroughs will occur in a given field.

The agents comprising a team may be classified according to their experience. Some agents are *newcomers*, that is, rookies, with little experience and unseasoned skills. Other agents are *incumbents*. They are established persons with a track record, a reputation, and identifiable talents. The differentiation of agents into newcomers and incumbents results in four possible types of links within a team: (1) newcomer–newcomer; (2) newcomer–incumbent; (3) incumbent–incumbent; and (4) repeat incumbent–incumbent. The distribution of different types of links reflects the team's underlying diversity. For example, if teams have a preponderance of repeat incumbent–incumbent links, it is less likely that they will have innovative ideas because their shared experiences tend to homogenize their pool of knowledge. In contrast, teams with a variety of types of links are likely to have more diverse perspectives to draw from, and therefore to contribute more innovative solutions.

Since quantifying the emergence and effects of team diversity (*2, 9, 24–26*) is more difficult than measuring team size, we consider next a model for the assembly of teams. In our model, we assemble N teams in temporal sequence. The assembly of each team is controlled by three parameters: m, p, and q. The first parameter, m, is the number of agents in a team. In our investigations of the model we considered three situations: keep m constant, draw m from a distribution, or use a sequence of m values obtained from the data. For the theoretical analysis of the model we keep m constant, while comparison with an empirical dataset is done using the sequence of $m(t)$ values in the corresponding dataset.

The second parameter, p, is the average fraction of incumbents that participate in a team. Higher values of p indicate fewer opportunities for newcomers to enter a field. The third parameter, q, represents the inclination for incumbents to collaborate with prior collaborators, rather than initiate a new collaboration with an incumbent they have not worked with in the past.

We start at time zero with an endless pool of newcomers. Newcomers become incumbents the first time step after being selected for a team. Each time step t, we assemble a new team and add it to the network (Fig. 2). We select sequentially $m(t)$ different agents. Each agent in a team has a probability p of being drawn from the pool of incumbents and a probability $1-p$ of being drawn from the pool of newcomers. If the agent is drawn from the incumbents' pool and there is already another incumbent in the team, then: (i) with probability q, the new agent is randomly selected from among the set of collaborators of a randomly selected incumbent already in the team; (ii) otherwise she is selected at random among all incumbents in the network.

Finally, nodes that remain inactive for longer than τ time steps are removed from the network. This rule is motivated by the observation that agents do not remain in the network forever—agents age and retire, change careers, and so on. The removal process enables the network to reach a steady state after a transient time. Our results do not depend in the specific value of τ (see Supporting Material).

Through participation in a team, agents become part of a large network (*27*). We were interested in examining the topology of the network of collaborations among the practitioners of a given field. More specifically, we asked "Is there a large connected cluster comprising most of the agents or is the network comprised of numerous smaller clusters?" A large connected cluster would be supporting evidence for the so-called "invisible college," the web of social and professional contacts linking scientists across universities proposed by de Solla Price (*28*) and Merton (*29*). A large number of small clusters would be indicative of a field made up of isolated "schools of thought." For all four fields considered here, we find that the network contains a large connected cluster.

As is typically done in the study of percolation phase transitions (*30*), we use the fraction S of nodes that belong to the largest cluster of the network to quantify the transition between these two regimes: invisible college or isolated schools. We explore systematically the (p, q) parameter-space of the model. We find that the network goes through a percolation transition at a critical line $p_c(m, q)$. That is, the system experiences a sharp transition from a multitude of small clusters to a situation in which one large cluster, comprising a large fraction S of the individuals, emerges—the so-called giant component (Fig. 3). The transition line $p_c(m, q)$ therefore determines the *tipping point* for the emergence of the invisible college (*31*). Our analysis shows that the existence or not of this transition is independent of the average number of agents $\langle m \rangle$ in a collaboration, although the precise value of $p_c(m, q)$ does depend on m.

The proximity to the transition line—which depends on the distribution of the different types of links—determines the structure of the largest cluster (Fig. 3A). In the vicinity of the transition, the largest cluster has an almost linear or branched structure (Fig. 3A, $p = 0.30$). As one moves toward larger p, the largest cluster starts to have more and more loops (Fig. 3A, $p = 0.35$) and, eventually, it becomes a densely connected network (Fig. 3A, $p = 0.60$).

Networks with the same fraction S of nodes in the largest cluster do not necessarily correspond to networks with identical properties. Each point in the (p, q) parameter-space is characterized by both S and the fraction f_R of repeat incumbent-incumbent links. For example, in Fig. 3C, the line $f_R = 0.32$ corresponds to those values of p and q for which 32% of all links in new teams are between repeat collaborators (*32*). The fraction of repeat incumbent-incumbent links has a significant impact on the dynamics of the network. When f_R is large, collaborations are firmly established and therefore the structure of the network changes very slowly. In contrast, low values of f_R correspond to enterprises with high turnover and very fast dynamics. Intermediate values of f_R are related to situations in which collaboration patterns with peers are "fluid" (see Supporting Material).

For each of the four fields for which we have empirical data, we measure the relative size of the giant component S (see Supporting Material). For all fields considered S is larger than 50% (Table 1). This result provides quantitative evidence for the existence of an "invisible college" in all the fields. Intriguingly, the relative sizes of the giant component is similar for three of the four fields considered: *BMI*, social psychology, and ecology, $S = 0.70$, $S = 0.68$, and $S = 0.75$

respectively. However, for astronomy S was significantly larger (0.92).

To gain further insight in the structure of collaboration networks, we use our model to estimate the values of p and q for each field. Given the temporal sequence of teams giving rise to the network of collaborations, one can calculate the fraction of incumbents and the fraction of repeat incumbent-incumbent links. These fractions and the model enable us to then estimate the values of p and q that are consistent with the data (*33*).

We estimate p and q for each field, and then simulate the model to predict the key properties of the network of collaborations, including the degree distribution of the network and the fraction S of nodes in the largest cluster. By comparing predictions of the model with the empirical results, we are able to validate the model. We first compare the degree distribution of the collaboration networks with the predictions of the model (Fig. 4A-D), and find that the model predicts the empirical degree distributions remarkably well. In Table 1, we compare the predictions of the model for S with the measured values. The model correctly predicts that an "invisible college" containing more than 50% of the nodes exists in all cases. Even more significantly, the values of S predicted by the model are in good agreement with the empirical results.

To investigate how changes of the team assembly mechanism affect the structure of the network, we use the model to generate networks with the same sequence of team sizes as the data, but with different values of p and q. We show in Figs. 4E-H that, notably, at least three out of the four creative networks we consider are very close to the "tipping" line at which an invisible college emerges. The exception is, again, astronomy. We also find that, for astronomy, the fraction of repeat incumbent-incumbent links is significantly larger than for the other fields.

If diversity affects team performance and our model correctly captures how diversity is related to the way teams are assembled, then the parameters p and q must be related to team performance. To investigate this issue, we consider, for the three scientific fields, how teams publishing in different journals are assembled. We used each journal's impact factor as a proxy for the quality of teams' output. We then studied the different journals separately to quantify the relationship between team assembly mechanisms and performance.

In Fig. 5, we show the values of p, q, and S for the journals in each of the fields as a function of the impact of the journal. We found that p was positively correlated with impact factor, whereas q was negatively correlated with impact factor, for both ecology and social psychology. The result for p implies that successful teams have a higher fraction of incumbents, who contribute expertise and know-how to the team, while the result for q implies that teams that are less diverse typically have lower levels of performance.

The relative size S of the giant component in a journal was also associated with performance for ecology and social psychology. Teams publishing in journals with a high impact factor typically give rise to a large giant component, while teams publishing in low-impact journals typically form small isolated clusters. This suggests that teams publishing in high impact journals perform a better sampling of the knowledge within a field, and thus are able to more efficiently use the resources of the "invisible college." Surprisingly, neither p, q, or S were significantly correlated with impact factor in astronomy. This distinguishes astronomy from the other creative enterprises considered.

We have shown that team size evolves with time, probably up to an optimal size, as in the case of the *BMI*. A similar process may be occurring for the parameters quantifying expertise, p, and diversity, q. Three of the four fields considered—all except astronomy—had very similar values of p, q, and S, thus suggesting that a "universal" set of optimal values might exist. The fact that in astronomy there are no correlations between p, q, or S and the impact of journals, also indicates that this field is different from the others. Whether these differences are due to the needs imposed by the creative enterprise itself or to historical or other reasons is a question that we cannot answer conclusively.

Our findings also suggest an explanation for the high impact of research published in top interdisciplinary journals such as *Science*, *Nature*, *PNAS*, *Physical Review Letters*, or *Bioinformatics*. The teams publishing in these journals comprise scientists drawn from across different fields. Hence, these teams draw on the largest pools of creative material—pools that exist not just within a single scientific discipline, but across different disciplines.

References and Notes

1. M. S. Granovetter, *Am. J. Sociol.* **78**, 1360 (1973).
2. R. Reagans, E. W. Zuckerman, *Organ. Sci.* **12**, 502 (2001).
3. R. Burt, *Am. J. Sociol.* **110**, 349 (2004).
4. B. Uzzi, J. Spiro, *Am. J. Sociol.* (forthcoming, 2005).
5. J. R. Larson, C. Christensen, A. S. Abbott, T. M. Franz, *J. Pers. Soc. Psychol.* **71**, 315 (1996).
6. A. Edmondson, *Admin. Sci. Quart.* **44**, 350 (1999).
7. K. A. Jehn, G. B. Northcraft, M. A. Neale, *Admin. Sci. Quart.* **44**, 741 (1999).
8. G. Stasser, D. D. Stewart, G. M. Wittenbaum, *J. Exp. Soc. Psychol.* **31**, 244 (1995).
9. J. R. Katzenback, D. K. Smith, *The Wisdom of Teams* (Harper Bussiness, NY, 1993).
10. J. M. Ziman, *Prometheus Bound* (Cambridge University Press, 1994).
11. J. R. Brown, *Science* **290**, 1701 (2000).
12. S. Green, K. Green, *Broadway Musicals Show by Show* (Hal Leonard Corp., Milwaukee, WI, USA, 1996), fifth edn.
13. R. Simas, *The Musicals No One Came to See: A Guidebook to Four Decades of Musical-Comedy Casualties on Broadway, Off-Broadway and in Out-Of Town Try-Out, 1943-* (Garland Publications, 1988).

14. We impose several requirements on the journals we selected for analysis. First, the main subject category of the journal must be the desired one. For example, we consider only those ecology journals whose subject category is either ”Ecology” or ”Ecology” and ”Biodiversity and conservation”, according to the ISI Journal Citation Reports. We disregard more specialized journals, such as Microbial Biology, whose subject category is more specific. We also require that journals contain a sufficiently large number of papers, typically at least 1500.

15. M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **98**, 404 (2001).

16. A.-L. Barabási, *et al.*, *Physica A* **311**, 590 (2002).

17. M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **101**, 5200 (2004).

18. K. Börner, J. T. Maru, R. L. Goldstone, *Proc. Natl. Acad. Sci. USA* **101**, 5266 (2004).

19. J. J. Ramasco, S. N. Dorogovtsev, R. Pastor-Satorras, *Phys. Rev. E* **70**, art. no. 036106 (2004).

20. This stationary state remains until the mid 1980s, when size drops again precisely at the time when a rash of revivals and revues conceivably simplified production.

21. R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).

22. M. E. J. Newman, *SIAM Review* **45**, 167 (2003).

23. L. A. N. Amaral, J. Ottino, *Eur. Phys. J. B* **38**, 147 (2004).

24. H. Etzkowitz, C. Kemelgor, M. Neuschatz, B. Uzzi, J. Alonzo, *Science* **266**, 51 (1994).

25. D. A. Harrison, K. H. Price, M. P. Bell, *Acad. Manage. J.* **41**, 96 (1998).

26. S. G. Barsade, A. J. Ward, J. D. F. Turner, J. A. Sonnenfeld, *Admin. Sci. Quart.* **46**, 174 (2001).

27. The teams and agents are the nodes in a bipartite network. Technically, agents are connected only to teams and vice-versa. However, this bipartite network can be projected onto a network comprising only agents and in which there is an edge (connection) between two nodes (agents) if the agents have been connected to at least one common team.

28. D. J. de Solla Price, *Little Science, Big Science... and Beyond* (Columbia Univ. Press, 1963).

29. R. K. Merton, *The Sociology of Science* (Univ. of Chicago Press, Chicago, 1973).

30. D. Stauffer, A. Aharony, *Introduction to Percolation Theory* (Taylor & Francis, 1992), second edn.

31. M. Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference* (Little, Brown and Company, Boston, 2000).

32. Figure 3C shows that large f_R occurs when p and q are large, and corresponds to a network in which collaborations among incumbents are firmly established and opportunities for newcomers are few. Conversely, small f_R, which occurs when p and/or q are small, indicates plentiful opportunities for newcomers to join new projects. In this case, newcomers are the norm and collaborations are rarely repeated. Finally, intermediate values of f_R suggest intermediate values of both p and q, that is, a situation for which there is a balance between seasoned incumbents and newcomers with fresh ideas.

33. The value of p is directly given by the fraction of incumbents in new creations, while the value of q must be obtained numerically by simulating the model.

34. We thank A. A. Moreira, V. Hatzimanikatis, J. M. Ottino, M. Sales-Pardo, and D. B. Stouffer for numerous suggestions and discussions. R.G. thanks the Fulbright Program and the Spanish Ministry of Education, Culture & Sports. L.A.N.A. gratefully acknowledges the support of a Searle Leadership Fund Award.

Figure Captions

Figure 1: Time evolution of the typical number of team members in: **(A)** the Broadway musical industry; and scientific collaborations in the disciplines of **(B)** social psychology, **(C)** ecology, and **(D)** astronomy.

Figure 2: Modeling the emergence of collaboration networks in creative enterprises. **(A)** Creation of a team with $m = 3$ agents. Consider, at time zero, a collaboration network comprising five agents, all incumbents (blue circles). Along with the incumbents, there is a large pool of newcomers (green circles) available to participate in new teams. The rules to create the team involve two steps. In step one, one selects an agent for the new team. With probability p this agent is a randomly selected incumbent, while with probability $1 - p$ it is a randomly selected newcomer. For concreteness, let us assume that incumbent "4" is selected as the first agent in the new team. One then repeats step one to determine if the second agent in the team is an incumbent or a newcomer. If the new agent is also an incumbent, then the incumbent is no longer selected at random. Instead, one uses step two: With probability q the second incumbent is selected from the set of incumbents that have *already* collaborated with incumbent "4", while with probability $1 - q$ it is selected at random from the set of all incumbents. In the figure, the new agent is selected from the pool of node "4" previous collaborators. Finally, one repeats step one to determine if the third agent in the team is an incumbent or a newcomer. In the figure, a newcomer is selected. **(B)** Time evolution of the network of collaborations according to the model for $p = 0.5$, $q = 0.5$ and $m = 3$.

Figure 3: Predictions of the model. **(A)** Phase transition in the structure of the collaboration network. We plot only the largest cluster in the network. For small p, the network is formed by numerous small clusters ($p = 0.10$). At the critical point p_c—the "tipping point"—a large cluster emerges, that is, a cluster that contains a significant fraction of the agents. In the vicinity of the transition, the largest cluster has an almost linear or branched structure ($p = 0.30$). As one increases p, the largest cluster starts to have loops ($p = 0.35$) and, eventually, becomes a densely connected cluster containing essentially all nodes in the network ($p = 0.60$). We show results for $q = 0.5$ and $m = 4$, where m is the number of agents in a team. **(B)** The transition described in (A) can be characterized by the fraction S of nodes that belong to the giant component—the order parameter. We display graphically the value of S as a function of p and q. Note that for any value of q, the model displays a percolation transition as the fraction p of incumbents increases from 0 to 1, and that the critical fraction p_c depends on q, defining a percolation line $p_c(m, q)$. The critical line $p_c(m, q)$ is an increasing function of q. Note, also, that the critical line also depends on the size m of the teams. We show results for $m = 4$. Even though the order parameter S is an important parameter to quantify the structure of the network, not all points with the same S, that is, all points represented with the same color, correspond to fields with identical properties. This result is made clear by the lines of equal probability of fraction f_R of repeat incumbent-incumbent collaborations. The upper-right corner of the (p, q) plane is characterized by f_R close to one, while the lower-left corner corresponds to f_R close to zero. As we show in Fig. 4, all fields considered have parameter values abobe the transition line.

Table 1: Global network properties of the fields studied. The sources for the *BMI* are: (i) Broadway Musicals Show by Show (*12*), and (ii) The Musicals No One Came to See (*13*). The data analyzed excludes revivals and focus on the "steady state" period 1940–1985. The data for scientific publications was obtained from Web of Science®. We selected recognized journals in each of the different scientific fields (see Table 2). For each field we show the total number of agents in all the period considered, the values of p and q obtained from the model, the fraction f_R of repeat incumbent-incumbent links, the size N of the network in the last year of the period considered, the value N_{mod} predicted by the model, the fraction S of agents that belong to the largest cluster, and the value S_{mod} predicted by the model. Note that, by definition, S takes values between 0 and 1 and does not depend on the size of the network (*30*).

Field	**Period**	**Agents**	**p**	**q**	$\mathbf{f}_R$	**N**	$\mathbf{N}_{\mathrm{mod}}$	**S**	$\mathbf{S}_{\mathrm{mod}}$
BMI	1877–1995	2303	0.52	0.77	0.16	428	420	0.70	0.80
Social psychology	1955–2004	23029	0.56	0.78	0.22	11412	14408	0.68	0.67
Ecology	1955–2004	38609	0.59	0.76	0.23	23166	26498	0.75	0.84
Astronomy	1955–2004	30192	0.76	0.82	0.39	18021	22794	0.92	0.98

Table 2: Journal-specific network structure. We present the same information as in Table 1 for each of the journals studied. We rank journals within each field according to their impact factor. Note that, for some low-impact journals, the fraction f_R of repeat incumbent-incumbent links is too high to be reproducible with the model. In those cases, which we represent by $q > 1$, simulations of the model are done with $q = 1$. The model still reproduces the empirical results quite well.

Field	**Journal**	**Period**	**Agents**	**p**	**q**	$\mathbf{f}_R$	**S**	$\mathbf{S}_{\mathrm{mod}}$
Social psychology	J. Pers. Soc. Psychol.	1965–2003	9112	0.56	0.74	0.20	0.75	0.79
	J. Exp. Soc. Psychol.	1965–2004	2133	0.40	0.76	0.11	0.44	0.07
	Pers. Soc. Psychol. B.	1976–2004	4339	0.45	0.74	0.14	0.54	0.47
	Eur. J. Soc. Psychol.	1971–2004	1790	0.41	0.93	0.15	0.44	0.08
	J. Appl. Soc. Psychol.	1971–2004	4602	0.33	1.00	0.10	0.06	0.02
	J. Soc. Psychol.	1956–2004	6294	0.32	>1	0.12	0.05	0.01
	Soc. Behav. Personal.	1973–2004	1981	0.26	>1	0.08	0.03	0.01
Ecology	Am. Nat.	1955–2004	4990	0.44	0.70	0.13	0.49	0.19
	Ecology	1965–2003	8885	0.48	0.71	0.15	0.56	0.65
	Oecologia	1969–2004	10545	0.44	0.81	0.15	0.51	0.36
	Ecol. Appl.	1991–2004	3417	0.29	0.99	0.08	0.30	0.06
	J. Ecol.	1955–2004	3639	0.43	0.91	0.15	0.40	0.19
	Funct. Ecol.	1989–2004	2873	0.36	>1	0.13	0.05	0.02
	Oikos	1961–2004	6589	0.43	0.84	0.15	0.48	0.11
	Biol. Conserv.	1977–2004	5821	0.27	>1	0.09	0.08	0.01
	Ecol. Model.	1978–2004	5260	0.35	>1	0.13	0.14	0.02
	J. Nat. Hist	1967–2004	2631	0.36	>1	0.04	0.13	0.01
Astronomy	Astron. J.	1965–2003	10832	0.78	0.86	0.40	0.96	0.99
	Publ. Astron. Soc. Pac.	1955–2004	6769	0.58	0.78	0.22	0.85	0.89
	Icarus	1983–2004	4357	0.72	0.90	0.38	0.89	0.97
	Publ. Astron. Soc. Jpn.	1965–2004	2432	0.77	0.95	0.44	0.95	0.99
	Astrophys. Space Sci.	1968–2004	10823	0.55	1.00	0.29	0.60	0.05
	IAU Symp.	1984–2004	10185	0.60	0.75	0.23	0.80	0.92

Figure 4: Network structure of different creative fields. Degree distributions for: **(A)** the Broadway musical industry, **(B)** the field of social psychology, **(C)** the field of ecology, and **(D)** the field of astronomy. Lines correspond to simulation results and symbols to empirical data. Simulations are carried out using the sequence $\{m(t)\}$ of team sizes found in the empirical data, and with the values of p and q estimated from the measured fractions of the different types of links. We present the predictions of the model with the lines and the empirical degree distributions with the open circles. For all cases considered, the data falls within the 95% confidence intervals of the predictions of the model. The (p, q) parameter-space of the network of collaborators for: **(E)** the Broadway musical industry, **(F)** the field of social psychology, **(G)** the field of ecology, and **(H)** the field of astronomy. The solid lines, separating the red and the blue regions, indicate the values of p and q for which 50% of the nodes belong to the largest cluster, that is, the percolation transition at which a giant component—the invisible college—emerges. The distance from the percolation line predicts the overall structure of the network. For example, the networks in astronomy is well above the "tipping" line and has a very dense structure (see Table 1). In contrast, all other fields are close to the transition and have relatively sparse giant components. Another important characteristic of the network is provided by the value of the fraction f_R of repeat incumbent-incumbent links. To help with the interpretation of the results, we plot with dotted lines the curves $f_R = 0.32$. For three of the creative networks considered, we find $f_R < 0.25$. For astronomy, we find $f_R = 0.39$.

Figure 5: Relation between team assembly mechanisms, network structure, and performance. We calculate the values of p, q, and S for several journals in each of the three scientific fields considered. Note that, in a few cases, q should be larger than one in order to reproduce the empirical values of f_R—in this cases q is considered one and the corresponding points are shaded. We plot the values of p, q, and S as a function of the impact factor of the journal, and then use the Spearman rank-order correlation coefficient r_s to determine significant correlations. A correlation is considered significant if the probability $P(r_s)$ is smaller than 5%. Shaded graphs indicate significantly correlated variables.