



Examining the Emergence of Large-scale Structures in Collaboration Networks: Methods in Sociological Analysis

Jaideep Ghosh¹ and Avinash Kshitij^{2,3}

Abstract

This article introduces a number of methods that can be useful for examining the emergence of large-scale structures in collaboration networks. The study contributes to sociological research by investigating how clusters of research collaborators evolve and sometimes percolate in a collaboration network. Typically, we find that in our networks, one cluster among the leading ones eventually wins the growth race by percolating through the network, spanning it and rapidly filling up a significant volume of it. We show how this process is governed by the dynamics of cluster growth in the network. When operating in a percolating regime, this class of networks possesses many useful functional properties, which have important sociological implications. We first develop the methodological tools to perform a study of the intrinsic clustering process. Then, to understand the actual large-scale structure

¹ School of Management and Entrepreneurship, Shiv Nadar University, Uttar Pradesh, India

² Council of Scientific and Industrial Research–National Institute of Science, Technology and Development Studies (CSIR–NISTADS), K. S. Krishnan Marg, New Delhi, India

³ Centre for Studies in Science Policy, School of Social Sciences, Jawaharlal Nehru University, New Delhi, India

Corresponding Author:

Jaideep Ghosh, School of Management and Entrepreneurship, Shiv Nadar University, Gautam Buddha Nagar, Uttar Pradesh 201314, India.

Email: jghosh20770@gmail.com

formation process in the network, we apply the theoretical methods to simulate a number of realistic scenarios, including one based on actual data on the collaboration behavior of a sample of researchers. From the perspective of social science research, our methods can be adapted to suit the application domains of many other types of real social processes.

Keywords

network analysis methods in sociological research, large-scale structure formation in collaboration networks, cluster evolution and percolation

Introduction

Work on the sociology of knowledge has demonstrated that research creativity often becomes a necessary consequence of a social system of researchers working in collaboration with one another (Uzzi and Spiro 2005). Access to ideas, information, and other resources through this system provides the supporting foundation that enriches a researcher's creative potential. Furthermore, acquaintance, far or near, weak or strong, with individuals through this network helps to build a researcher's social capital (Burt 2000; Coleman 1988). Thus, researchers connected to one another in a networked environment share ideas, thoughts, and research questions, make use of complementary research methods and techniques, and produce interconnected influence on one another (Moody 2004).

In an early stage of the development of this network, budding ideas and thoughts are primarily confined to small, cohesive groups of clusters¹ of researchers, largely disconnected from one another (Durkheim 1984; Moody and White 2003). However, over time, as disjoint clusters begin to be bridged up, the complete network tends to exhibit evidence of broader patterns of interdisciplinary connectivity. This is produced primarily by the mixing of different classes and types of ideas which are exchanged between researchers in small clusters which had largely been isolated from one another before the bridges were formed between them.

The mixing of diverse research ideas, thoughts, and problems by the researchers is highly facilitated when large-scale connectivity is established through the bridging of many of the disjoint clusters that constituted the network at some initial time. This process helps to eventually bring about a cooperative behavior among the researchers in the network. The underlying cluster evolution dynamics governs the specific linkage patterns that result from the joining of the clusters. The process in question works in the following way.

At some initial time (say, $t = t_i$), the network consists merely of a small number of tiny, and largely isolated, groups of clusters.² Over time, new researchers join the network, and some of the existing clusters become connected to one another through internal bridging links. Viewing the network at the micro level at a specific future time ($t = t_j$), it seems reasonable to envisage three possibilities for determining the fate of every new researcher that is added to the network in course of time: (1) the new researcher remains in the network as an isolate for a long time, (2) the researcher, who joined as an isolate, subsequently connects to an existing cluster, and (3) the researcher joins directly as a part of one of the existing clusters in the network.

On the one hand, the dynamics of the underlying social process creating the networked collaboration system governs the formation of its large-scale structures; and, on the other, the structure itself influences the dynamics for sustaining the network over time (Giddens 1979). As large-scale structures start to emerge through the formation of cluster linkages, communication channels open up for the researchers in the individual clusters, serving as means for building their social capital (Burt 2000; Coleman 1988; McFadyen and Cannella 2004). Earlier studies have explored such processes in many different types of networks. Important ones include interorganizational networks (Gulati and Gargiulo 1999; Powell et al. 2005), innovation and knowledge networks (Ahuja 2000), public health networks (Provan, Beagles, and Leischow 2011); biotechnology industry networks (Walker, Kogut, and Shan 1997), social science collaboration networks (Moody 2004) and so on. The problem of the formation of large-scale structures of epidemiological networks driven by underlying social forces has also been investigated (Moore and Newman 2000).

For the present work, it is important to be clear at the outset that we do not intend to investigate specific *mechanisms* responsible for cluster evolution in collaboration networks.³ Rather, we concentrate our attention on the behavioral aspects of the emergence of the large-scale structures in such networks, given whatever underlying mechanism is ultimately responsible for the intrinsic evolutionary processes at play in the networks. Adopting this viewpoint, we address the following issues in this article: (1) In a practical situation, how can one track the evolution of competing clusters in the network? (2) Given the existence of an underlying growth mechanism, how can one parametrize the dynamics of local connectivity between researchers in the network? (3) With this parametrization, how can one characterize large-scale structure formation in the network? This article develops the relevant methodological tools to perform this analysis and demonstrates how the techniques can actually be applied in practice to a class of collaboration networks of researchers in two countries: India and the United States.

The Case of Collaboration Coauthorship Networks

It is frequently the case that only a small number of researchers come to dominate the arena of specialized research work by producing a disproportionately large number of novel ideas and solutions (Crane 1972). A creative mixing of diverse ideas and an emergent consensus among researchers depend on the overall connectivity of their collaboration network (Moody and White 2003). The issue in question has a dual aspect. On the one hand, the highly connected, star researchers in their cohesive clusters render the network highly vulnerable to breakdown, since their removal catastrophically disconnects the clusters (Allison, Long, and Krauze 1982). On the other hand, collaborative brokers, though not necessarily star researchers themselves, generate large-scale connectivity, and ideas are more likely to spread through them over the entire network (Fleming, Mingo, and Chen 2007). For the network, this typically signifies a small world (Watts 1999), which influences the behavior of researchers by modifying the network's large-scale topology. Thus, it enables "creative material in separate clusters to circulate to other clusters as well as to gain the kind of credibility that unfamiliar material needs to be regarded as valuable in new contexts, thereby increasing the prospect that the novel material from one cluster can be productively used by other members of other clusters" (Uzzi and Spiro 2005:449). Local clusters contain specialized knowledge or resources, but their bridging through intermediaries enables the specialized elements to be circulated and mixed, giving the opportunities for generating novelty and innovation in the connected part of the full network. Today, many innovative research works and discoveries have increasingly resulted from the efforts of this kind (Paruchuri 2010; Shrum, Genuth, and Chompalov 2007). It has frequently been the case that collaborating researchers produce more novelty and innovation in research than do single researchers by individual efforts (Kotha, George, and Srikanth 2013). Empirical work also indicates that collaborative research publications have higher citations on the average (Kotha et al. 2013). The social force acting between researchers is now, almost universally, accepted to be the most critical component of research collaboration⁴ (Acedo et al. 2006). Admittedly, the primary objective of research collaboration is the pursuit of new knowledge and the creation of many innovative capabilities (Fleming et al. 2007). Nevertheless, on the practical side, research collaboration also creates new social pathways, enabling researchers to extend their domains of professional contacts and research circles (Ahuja 2000). The social force binding the researchers within exclusive groups of contacts helps to build a personal relationship network, based on the mutual trust and like-mindedness of the

individuals, their sharing of ideas, and their complementary technical skills in research (Bozeman and Corley 2004).

The research collaboration scenario studied in this work is realized through the coauthorship network, in which the collaborative activities of researchers culminate in the publishing of one or more scholarly articles in peer-reviewed journals and conference proceedings (Acedo et al. 2006; Barabási et al. 2002; Ghosh and Kshitij 2014; Ghosh, Kshitij, and Kadyan, 2015; Newman 2001). Although this type of network has been studied for quite some time now, there is rather limited *behavioral* investigation of cluster evolution in it. The present article contributes to the understanding of this specific evolutionary process and investigates the practical question concerning the possibility of whether large-scale connectivity exists in the network through the study of network percolation.

Empirical Setting and Data

As mentioned before, we investigate in this study the network structures of collaborative works of researchers, created by the inclusion of a sample of researchers through their coauthored papers in peer-reviewed journals and conference proceedings. We select the growing field of management and information, including related areas of information technology and economics (designated collectively by MGMT). We built our networks using bibliometric data from Elsevier's SciVerse (Scopus) and Thomson Reuters' Web of Science (WoS) electronic databases. The indexed data were retrieved from the databases using the condition that a researcher is based in a certain country, such as India or the United States in the present case, if they work in an institution in that country. However, it is perfectly possible as well as allowable that one or more coauthors of a researcher may be based in an institution in another country. Research specialization areas included in our data sets are searched within the primary subdisciplines of MGMT including, for instance, management information systems, operations management, operations research, finance, accounting, organizational behavior, strategy, managerial and information economics, and so on. We extensively build separate networks in this way for two countries: India and the United States.

Theoretical Questions, Research Methods, and Computational Details

This section presents the primary theoretical questions and the concerned research methods employed in this work. The related computational details

used to carry out numerical simulations are also provided. Most of the computations are performed by executing fast numerical routines in C and Java. For a few computations, the Pajek software package (version 4.04) is used (Batagelj and Mrvar 2015).

Network Characterization

Our model characterizes a collaboration coauthorship network as a one-mode projection of an *affiliation* network, in which researchers are connected to one another through their common membership in research projects or groups. Of these instances, we consider only those cases where this connection has been *formally* established through the common presence of these researchers in coauthored papers published in peer-reviewed journals and conference proceedings. A one-mode projective network of this type having size n with dichotomous collaborative association between pairs of researchers is represented by an $n \times n$ adjacency matrix A with elements $A_{ij} = 1$ if researchers i and j are the coauthors of a paper and $A_{ij} = 0$ otherwise (Barabási et al. 2002; Ghosh and Kshitiij 2014; Ghosh et. al 2015; Newman 2001). In a more general setting, it is possible to incorporate the strengths of collaborative coupling between pairs of researchers into the network structure (Ghosh et al. 2014). However, for the primary purpose of investigating the problem of large-scale structure formation in the network, in this work we leave out this additional complication and consider the simpler dichotomous network instead.

Cluster-tracking Procedure

In a social network, cluster growth is intrinsically a competitive process. The primary theoretical question in this regard pertains to the issue of global dominance of one or more clusters in the competition for growth. For example, in an interorganizational network, a group of organizations may form a large cluster through their local supply-chain connections. Initially, this process may start through collusion of a few firms by forming only a small clique. Eventually, this small initial group may emerge as a fully functional cartel that comes to control and dominate the entire consumer market in a few years' time.

It has been found empirically that, when one of the clusters in the network starts growing, it tends to swamp all others smaller than it in size and begins to fill a significant volume of the entire network (Callaway et al. 2000; Newman and Watts 1999). This cluster is a large subset of researchers that

are all connected to one another through intermediate collaborative ties. When such a large cluster persists in the network, the complete network operates in the percolating regime. The ratio $\pi = \frac{\text{Largest cluster size}}{\text{Full network size}}$ serves as the *order parameter* for the concerned percolation problem. Empirically, a large value of π signals a *percolation transition*, being the onset of the formation of the giant cluster in the network. Numerically, π goes from a small value to a large one at the *transition threshold* at which the giant cluster forms (Newman and Watts 1999).

In empirical studies, the threshold values have been determined for many different types of collaboration coauthorship networks (Newman 2001; Newman and Watts 1999). For example, in biomedical research, π is close to 93 percent for the MEDLINE (MEDLINE is the U.S. National Library of Medicine bibliographic database for journal articles in life sciences and biomedicine.” For reference, please see <http://www.nlm.nih.gov/pubs/factsheets/medline.html>) database; in astrophysics, it is about 89 percent for the Los Alamos e-Print database; in condensed matter physics, it is about 85 percent for the Los Alamos database; in theoretical high-energy physics, it is approximately 71 percent for the Los Alamos database; in computer science, it is roughly 57 percent for the Networked Computer Science Technical Reference Library database (Newman 2001). MGMT collaboration coauthorship networks in India and the United States have π values lying typically in the range of about 60–70 percent (Ghosh and Kshitij 2014). Collaboration coauthorship networks in cancer research in India have π values lying in the range of 83–96 percent (Kshitij, Ghosh, and Gupta 2015).

To witness cluster formation and growth patterns in the networks, we employ, as our first strategy, an empirical method of individually tracking a selected sample of the leading clusters over the entire window of study from 2000 to 2011 (inclusive) in time increments of one year⁵ and then plotting the values of π against time. Thus, at time $t = 0$, we start our observational recording by putting labels on all of the individual clusters in a sample of the leading clusters in the network (the top five clusters in increasing orders of size, for example). Subsequently, at each increment of one year ($t = 1, 2, \dots$), we record the sizes of the previous clusters. As the process unfolds in time, some of the previous clusters merge together to form new clusters of increasingly larger sizes. A few of the initial clusters may remain essentially isolated from the rest of the clusters in the network and show only minimal growth, if at all. Depending on their initial size, these clusters may be in or out of the growth race at subsequent times and eventually by the time our clock stops ticking at a predetermined time (in our case, $t = 11$).

Additionally, entirely new clusters⁶ may join the competition and oust some of the previous ones from the race. The method is of great practical utility: first, it allows us to visualize how some of the clusters exhibit a tendency to percolate; and second, it helps us to specifically locate at what points of time the cluster size begins to grow significantly in the network. As another matter, if two large clusters merge to give rise to a very large cluster, it is also possible to locate the individual researchers that serve to bridge the previously disconnected clusters in the network.

Collaboration Willingness

Given an underlying growth mechanism at play in the network, now comes the important theoretical question as to how one can parametrize the instances of local connectivity among the different actors in the network. In the context of our collaboration network constructed using cross-sectional data over a specific window of time, this is, in effect, a snapshot of a dynamical process in which many of the existing isolated clusters of researchers in the network are continuously bridged by means of new connections formed in the network. Also, concurrently, many new researchers are entering the network by collaborating and coauthoring papers with researchers who are already established in collaborative research. The growth of clusters in this way is an evolutionary process.

Since the practice of collaboration has a major social underpinning, an important social indicator in this regard is the collaboration willingness of a researcher in the network. This measure can be operationalized in terms of an average probability that a certain researcher selected at random will be willing to collaborate in order to form *new* connections with other researchers in the network. If the researcher is averse to starting any further collaboration beyond what is involved in their present collaboration circle, then the researcher will not serve as a source to form new links with which to generate cluster growth in the network. In other words, there is no further flow of resources through the researcher into directions previously unexplored in collaboration. On the other hand, a researcher's willingness to form new collaborative ties cannot increase indefinitely. As collaborative research work demands a definite commitment in time, resources, and creative energy on the part of a researcher, it may become increasingly burdensome for the researcher to collaborate beyond a certain number of collaborative research connections.⁷ Besides, when research collaboration is of interdisciplinary nature, there is a mixing of diverse types of knowledge, ideas, and expertise (Cummings and Kiesler 2007). Coordination in the form of management of

interdependencies among research activities becomes essential in this practice (Malone and Crowston 1994). A very high engagement in collaborative activities is the likely cause of complex coordination problems in research. This indicates, therefore, that there must exist a threshold, θ , to cut off increasing collaborative activities of researchers.

The actual number of collaborators of a researcher also governs the collaboration willingness. In our model, we represent this parameter by $\omega(k, \theta)$ and investigate its behavior to see how it governs cluster growth in the network in a number of practical scenarios. Finally, we perform Monte Carlo simulations of a situation based on centered values of *perceived* collaboration willingness that a sample of researchers have revealed to us in a recent survey of research collaboration practices and strategies in India.

Cluster Growth and Distribution

A very important theoretical question bears on the issue of large-scale structure formation in a social network. Considering this question in the context of our collaboration network, at some initial time in the history of growth of a network of this type, there are many isolated clusters of research collaboration with only very few connections among the clusters of comparatively large size. The individual clusters are unevenly distributed in size, but it frequently transpires that among the various clusters present at a particular time, there is a leading one having the largest size.⁸ As the large-scale collaboration structure unfolds, the clusters in the network continue to grow in size by forming intercluster bridges as well as by the addition of new researchers to the existing networked system.

As collaborative ties are thus being formed, it commonly becomes overwhelmingly probable that a new link will emerge in the direction of one of the larger clusters in the network. The reason is that, probabilistically, the larger clusters already enjoy the benefits of higher connectivity, resulting in an *attractive force* drawing new collaborative links toward themselves. Among the existing larger clusters, there is already a competition to grow in size and win. In most situations involving a social process, only one or two large clusters show excessive growth, quickly swamping all others in the network. This phenomenon has a strong socioculturally directed force as its driver, and it is not the result of a purely stochastic growth mechanism (Capocci et al. 2006).

A researcher can be connected to an existing large cluster through one or more of their first-order neighbors. This is obvious if such a neighbor is already a member of this large cluster. Contrariwise, it becomes impossible

for a researcher to connect to a large cluster if the neighbors are not themselves members of the set of one of the large clusters in the network or if they themselves exhibit unwillingness to form any further collaborative ties. In any case, at a certain stage of growth, if a researcher already has k collaborating connections, then the probability that the person is willing to collaborate at this degree value is given by $P(k, \theta) = p(k)\omega(k, \theta)$. The generating function for this distribution can be written as $\varphi(z, \theta) = \sum_{k=0}^{\infty} P(k, \theta)z^k$ (Callaway et al. 2000; Newman, Strogatz, and Watts 2001). Additionally, $\langle \omega \rangle = \varphi(1, \theta) = \sum_{k=0}^{\infty} P(k, \theta)$ is the average level of collaboration willingness in the network. Sociologically, $\langle \omega \rangle$ accounts for the mean proportion of researchers in the network that are willing to start *new* collaborative activities. The *largest* cluster in this case consists of a certain fraction π of the researchers in the entire network given by $\pi(\theta) = \varphi(1, \theta) - \varphi(\xi, \theta)$, where ξ is the largest cluster exclusion probability of a randomly selected researcher in the network.

In a situation where a researcher has a prospective neighbor in a small cluster, the focal researcher remains outside the largest cluster in the network, because that neighbor is still not a member of the largest cluster. Additionally, if the concerned neighbor of the focal researcher is unwilling to collaborate, then the focal researcher has no way of connecting to the largest cluster through the neighbors of that neighbor.⁹ At the present level of k collaborators of the neighbor, the probability that the focal researcher fails to make an entry into the largest cluster can be calculated as $1 - \omega(k+1, \theta) + \omega(k+1, \theta)\xi^k$, where the ξ^k -term arises from the consideration of the second-order neighbors of the focal researcher or the first-order neighbors of the selected neighbor. The selected neighbor of the focal researcher has a collaborator distribution given by $kp(k)$ (Newman 2003). The probability distribution that this neighbor is willing to collaborate at this degree value is given by the generating function $\psi(z, \theta) = \frac{1}{\mu} \sum_{k=1}^{\infty} kp(k)\omega(k, \theta)z^{k-1}$, where μ is network's average degree. A comparison with the previous equation yields a polynomial equation for ξ thus: $\xi = \frac{1}{\mu} \sum_{k=1}^{\infty} kp(k)[1 - \omega(k, \theta) + \omega(k, \theta)\xi^{k-1}]$, where $\mu = \sum_k kp(k)$ is the average number of collaborators in the network. Finally, expanding the individual terms on the right-hand side of the above ξ equation, we obtain the following result¹⁰: $\xi = 1 - \psi(1, \theta) + \psi(\xi, \theta)$. The solution for ξ obtained from this equation can be directly inserted into the expression for $\pi(\theta)$ given above to obtain the size of the largest cluster in the network. Note that the ξ equation is actually polynomial equation of large order, whose real roots,

lying in the range $(0, 1)$ ¹¹ are obtained numerically in the simulations described below.

Collaboration Willingness Scenarios

It is clear that different functional characteristics of cluster growth will result in the network for different functional forms representative of $\omega(k, \theta)$. To this end, we examine several scenarios to get a behavioral picture of cluster growth in the network. Starting with a simple scenario, we incrementally improve it by incorporating more realistic complexities into it, and finally, we perform Monte Carlo simulations, bootstrapped on actual values of perceived collaboration willingness of researchers obtained from a recent survey of researchers in MGMT in India.

Scenario 1. In this scenario, we model $\omega(k, \theta)$ as a 0 – 1 step function as follows: $\omega(k, \theta) = 1$ for $k < \theta$ and $\omega(k, \theta) = 0$ for $k \geq \theta$. Thus, before reaching the threshold, a researcher is perfectly willing to collaborate, but this willingness cuts abruptly off to zero as soon as the threshold is crossed. In real situations, however, this type of willingness perception profile is a bit unnatural, since it is not usually the case that a researcher becomes entirely unwilling to collaborate immediately as their current number of collaborative ties crosses a definite threshold. We investigate the behavior of cluster growth and, in particular, the formation of the largest cluster as the parameter θ is varied over a range of degree values. Admittedly, a researcher’s collaboration willingness behavior is hardly ever so abrupt in reality. Nevertheless, the scenario itself serves as a valid check to see if our model, at least *qualitatively*, shows the correct behavior of growth and evolution in this case.

Scenario 2. In this scenario, researchers’ willingness profile is modeled as $\omega(k, \theta) = 1$. For $k < \theta$ and $\omega(k, \theta) = \exp(-\frac{k}{\theta} + 1)$. For $k \geq \theta$. Thus, researchers are perfectly willing to build new collaborative connections with other willing researchers with their present degrees rising up to θ , but after this threshold is crossed, their willingness does not fall abruptly to 0, as in the previous scenario, but declines exponentially by an amount given by the fraction by which k exceeds θ . This scenario is much more realistic than the previous one, since it accounts for a rapid, but not abrupt, drop of willingness after the threshold is crossed. Most researchers in real life are likely to behave in this way. For example, a research advisory faculty can take only a limited number of doctoral students to work on research problems over a

fixed period of time. This is true not just with student collaborators but with the researcher’s other coworkers as well. Basically, when a researcher’s threshold capacity to do collaborative work is exceeded, it becomes increasingly difficult for the person to spend the necessary energy and time in any new collaborative endeavor, and coordination as well as management problems invariably arise. However, since the threshold is actually a researcher’s *perception* of the physical and intellectual capacity to collaborate, given the present number of collaborative ties, the willingness function falls exponentially fast as this threshold is crossed, and it does not abruptly drop to zero.

Scenario 3. Here, the willingness profile has the form $\omega(k, \theta) = \frac{k}{\theta}$ for $k < \theta$ and $\omega(k, \theta) = \exp(-\frac{k}{\theta} + 1)$ for $k \geq \theta$. This scenario is even more realistic than the previous one. A researcher’s willingness behavior beyond the threshold is the same as in scenario 2, but for the range of k values less than the threshold, it is perhaps unrealistic to simply assume (as in the previous two scenarios) that the researcher is *perfectly* willing to collaborate uniformly at all levels of k . Perhaps, it is more natural to suppose that the focal researcher’s willingness function increases linearly¹² as he or she succeeds in building new collaborative ties with other researchers.

In an analysis of the large-scale structural topology of the MGMT networks, we empirically found the degree distributions to be truncated power laws (Ghosh and Kshitij 2014). To model the present scenario for numerical computations, we implement such a distribution as possessing a general analytical form $p(k) = \frac{k^{-\alpha} e^{-\frac{k}{\tau}}}{Li_s(e^{-\frac{1}{\tau}})}$, where α is the exponent and τ is the cutoff of the power law. The function $Li_s(z)$ in the denominator of the above expression is the *polylogarithm* function¹³ of order s and argument z defined by $Li_s(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^s}$ (Abramowitz and Stegun 1965; Cvijović 2007). Incorporating the willingness function for this scenario into the ξ equation obtained above yields, after some nontrivial algebraic manipulations, the following exact equation: $\xi = 1 - \frac{1}{\mu \theta Li_s(e^{-\frac{1}{\tau}})} \sum_{k=1}^{\theta-1} k^{2-\alpha} (1 - \xi^{k-1}) e^{-\frac{k}{\tau}} - \frac{1}{\mu Li_s(e^{-\frac{1}{\tau}})} \sum_{k=\theta}^{\infty} k^{1-\alpha} (1 - \xi^{k-1}) e^{1 - (\frac{1}{\theta} + \frac{1}{\tau})k}$ where μ is the average degree of the network. Although the equation looks rather formidable at first sight, it is straightforward to implement computationally and solve numerically as a polynomial equation in ξ . Similarly, with the above distribution for $p(k)$ and the functional form for $\omega(k)$, some amount of algebraic manipulations

yield the following π equation:
$$\pi = \frac{1}{\theta Li_x(e^{-\frac{1}{\theta}})} \sum_{k=1}^{\theta-1} k^{1-\alpha} (1 - \xi^k) e^{-\frac{k}{\tau}} - \frac{1}{\mu Li_x(e^{-\frac{1}{\theta}})} \sum_{k=\theta}^{\infty} k^{-\alpha} (1 - \xi^k) e^{1 - (\frac{1}{\theta} + \frac{1}{\tau})k}$$
. With regression-fitted values of the parameters α and τ in the empirical degree distributions, it is now straightforward to obtain the functional form of the order parameter π as a function of the collaboration threshold θ .

Scenario 4. This scenario is a generalization of scenario 1. In this case, the willingness profile has a two-parameter (θ, γ) distribution as follows: $\omega(k, \theta, \gamma) = \frac{1}{e^{\frac{k-\theta}{\gamma} + 1}}$, where $\gamma \geq 0$. When $\gamma = 0$, the distribution is a step function: For $k > \theta$, ω vanishes and for $k < \theta$, ω is unity. The function cuts abruptly off at $k = \theta$. At a finite γ , the distribution becomes fuzzy over a length of the order several γ . At very high values of the degree, $k - \theta \gg \gamma$, and the distribution has the asymptotic form given by $\omega \sim e^{-\frac{\theta-k}{\gamma}}$. In spirit as well as in implementation, this scenario is different from all of the previous ones: rather than directly examining the large-scale growth characteristics with the threshold θ , we investigate here how the order parameter is impacted by the values of γ for a *fixed* value of θ . Physically, γ is an *adversity* determinant, which measures the lack of idealness of the environment for collaboration. The ideal, albeit nonpractical, situation occurs at $\gamma = 0$, when each researcher in the network is willing to build new collaborative ties all the way up to the maximum capacity determined by the threshold level but stops completely once this threshold is crossed. As γ starts to increase from 0, the situation becomes increasingly nonideal, in which researchers exhibit unwillingness to forge any new collaborative ties much before the threshold capacity is reached. This behavior is possibly due to a number of exogenous factors not inclusively accounted for in the perceived collaboration willingness profile. For example, a researcher's current work pressure, in the form of administrative duties, teaching load, and so on, that is not directly related to research activities may contribute to a lack of eagerness to collaborate any further. Also, in some cases, researchers' geographical isolation or their lack of conference and workshop participation might constrain researchers to grossly limit collaborative activities.¹⁴ The onset of this tendency to prematurely reduce collaboration willingness is controlled in an aggregative manner by the adversity parameter γ in our simulations.

Using the same analytical form of the general degree distribution pertaining to our model as used in the previous scenario, the ξ equation in the

present case can be shown to assume the form $\xi = 1 - \frac{1}{\mu Li_x(e^{-\frac{k}{\gamma}})} \sum_{k=1}^{\tilde{\theta}} k^{1-\alpha}$
 $\frac{e^{-\frac{k}{\gamma}}}{e^{\frac{k-\tilde{\theta}}{\gamma}} + 1} (1 - \xi^{k-1})$, where $\tilde{\theta}$ is a predetermined fixed value of θ (in this work,

$\tilde{\theta} = 0.5 \theta_{max}$). Now, with regression-fitted values of τ and α , the above polynomial equation can be numerically solved for ξ , where $0 \leq \xi \leq 1$. This solution, say $\bar{\xi}$, can then be substituted into the corresponding π equation, which, after some mathematical manipulations, can be shown in this case

$$\text{to reduce to the following form: } \pi(\gamma) = \frac{1}{Li_x(e^{-\frac{k}{\gamma}})} \sum_{k=1}^{\tilde{\theta}} k^{-\alpha} \frac{e^{-\frac{k}{\gamma}}}{e^{\frac{k-\tilde{\theta}}{\gamma}} + 1} (1 - \bar{\xi}^k).$$

A plot of $\pi(\gamma)$ for different values of γ then exhibits the functional behavior of the order parameter against the adversity parameter of the model. Since rising adversity in research environment discourages researchers to forge new collaborative ties in the network, the order parameter should exhibit a gradual, nonlinear decline with increasing adversity.

Scenario 5. In this final scenario, we perform Monte Carlo simulations to obtain a distribution of ω centered on the *actual* values of perceived collaboration willingness reported by a sample of researchers from the India network. The computational procedure employed to perform these simulations is described as follows.

Based on the last three years' (2010–2012 inclusive) collaboration engagements of a sample of 23 researchers in India working in the MGMT field, data for their perceived collaboration willingness are collected through survey questionnaires and semistructured interviews. Empirical degree distributions of their coauthorship network are computed based on the above window of time.¹⁵ Employing the bootstrap procedure (Efron 1993; Mooney and Duval 1993), we assume that this data set is the real population giving the true values of the parameters of the model. Using the full range of unique degree values ($k_i, i = 1, \dots, s$) in this data set and the corresponding values of the perceived collaboration willingness,¹⁶ we fit a cubic spline interpolated polynomial to get a functional profile of ω using the s data points in the sample (Press et al. 1992). Employing this willingness profile and the corresponding $p(k)$ from the network, we compute π_{true} , the *true* value of the order parameter.

Next, we select s degree values $k_i, i = 1, \dots, s$ from the sample uniformly at random with replacement. In this selection, because of the replacement procedure, a fraction of the degrees from the actual set will be repeated. For each such k_i , we do not take the corresponding $\omega(k)$ directly from the actual

Table 1. Summary of Metrics and Parameters.

Quantity	Symbol	Explanation
Degree	k	Number of direct collaborating coauthors of a researcher
Order parameter	π	A measure of the proportion of total network volume occupied by the leading cluster in network
Collaboration threshold	θ	Maximum number of collaborators a researcher is willing to collaborate with (may depend on time window in a general situation). When this threshold is crossed, the collaborative activities of a researcher are grossly reduced
Collaboration willingness profile	$\omega(k, \theta)$	A distribution function for the perceived collaboration willingness of a researcher that depends on the current number of collaborators and the collaboration threshold
Mean collaboration willingness	$\langle \omega \rangle$	Mean proportion of researchers in network that are willing to start new collaborative activities
Giant cluster exclusion probability	ξ	Leading cluster exclusion probability of a randomly selected researcher in network
Adversity parameter	γ	A measure of the lack of nonidealness of the environment for collaboration. In a general situation, it represents environmental adversities limiting the growth of a social movement
Tolerance parameter	η	An acceptance bound for bootstrapped Monte Carlo simulated values for selecting the profile distribution of $\omega(k)$

set. Instead, each value is selected such that it is accurate to within a preassigned measure of tolerance. The value of this tolerance parameter is generated by drawing a real random number from the range $[0, 1]$. If this number lies within a small range of $\pm \eta$, where η is the tolerance parameter, centered on the actual sample value, then we select the point and use it as the corresponding $\omega_\eta(k)$, otherwise we reject it and draw again.¹⁷ We repeat this procedure for all the s degree values from the data set to generate an entire manifold of simulated ω values. Using this set, we then fit a cubic spline interpolated polynomial for ω_η and compute π_1 as above. We then repeat the above steps r times¹⁸ to generate the sample of π as follows: $\pi_1, \pi_2, \dots, \pi_r$. This *synthetic* data set is used to obtain the sampling distribution of π . The entire procedure is repeated for different values of the tolerance parameter η .

A summary of the metrics and parameters used in the present work appears in Table 1.

Name-resolution Algorithms

Our data samples collected from Scopus and WoS are of unequal total volume because of coverage variations in the databases. Furthermore, the journals and the published conference proceedings indexed in these databases have different listing styles and arrangements of author names that appear in the papers. The exact value of the total number of authors cannot, therefore, be precisely estimated. Following Newman (2001), we employ two separate algorithms to address the issue: (1) computing a first initial (FI) limit: an author in a paper is identified only by their last name and the FI and (2) computing the all initials (AI) limit: an author is identified by their last name and AI. The FI method underestimates the total count of author names when it identifies two authors as one individual. By contrast, the AI method overestimates when it identifies one author as two if their initials are listed differently in different papers. However, the complete interval (FI, AI) can *statistically* capture the actual number of authors in the data sets. It is important to realize that, in this work, we regard an author as simply an *actor* in the concerned network and not as a *person* having a specific identity to be revealed by network analysis.

We still have to cope with the problem of duplicate author names in the data. For example, there might be a situation in which the U.S.-based collaborator of an India-based researcher was included in the India network; however, this collaborator's primary network would truly be the U.S. network. These problems notwithstanding, the algorithms do include all the relevant cases of author names in the designated interval. The possibility of a bias exists in this type of estimation procedure, but a redeeming feature is that, the bias affects both the India and the U.S. networks to the same degree as well as in the same direction. A number of other methods for author name disambiguation also appear in the research literature (Kang et al. 2009; Milojević 2013).

Results

Cluster Tracking

In Figures 1–4, we display the growth of the five leading clusters over the entire period of study from 2000 to 2011 (inclusive) in time increments of one year for the Scopus and WoS networks of India and the United States. The specific internal details of the actual cluster evolution patterns exhibited in the figures are shown in Table 2. In the Scopus India network (Figure 1), at $t = 0$ (year 2000), there are five clusters, rather small in size and labeled a_1, \dots, a_5 . Figure 1 shows that, in this network, the largest cluster

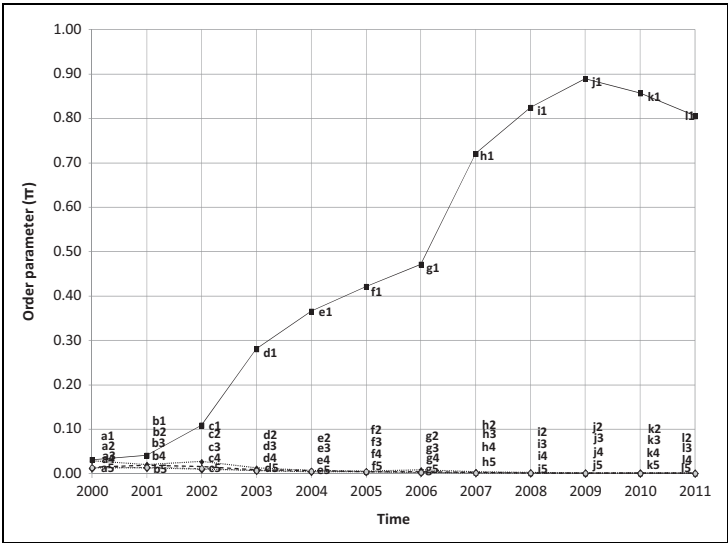


Figure 1. Five leading clusters (Scopus, India).

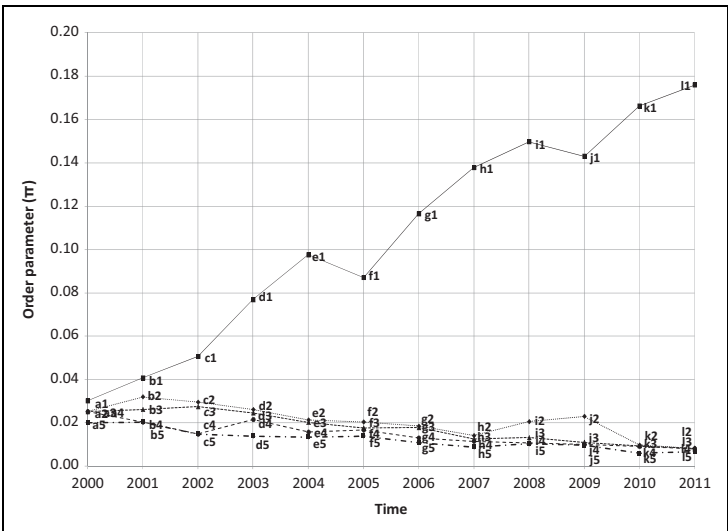


Figure 2. Five leading clusters (Web of Science, India).

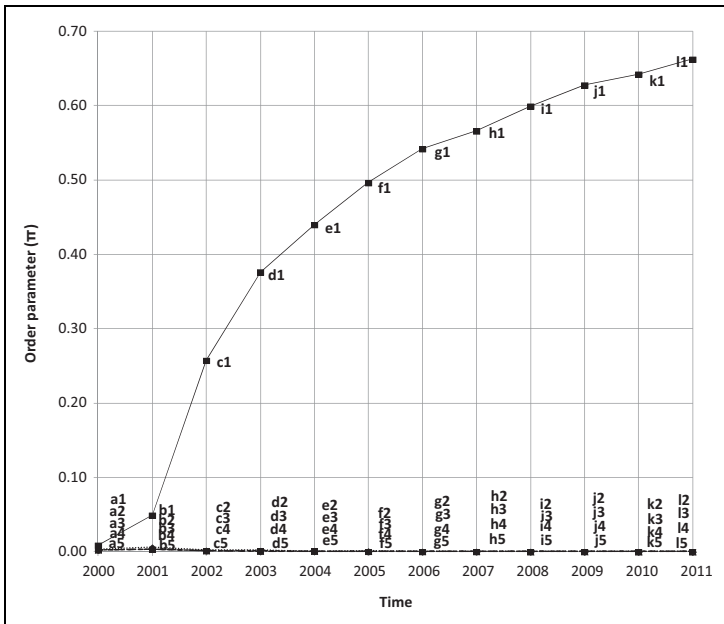


Figure 3. Five leading clusters (Scopus, U.S.).

(#1) has propagated over time as $(a_1, b_1, c_1, d_1, e_1, f_1, g_1, h_1, i_1, j_1, k_1)$. The letters, a, b, \dots, l indicate the names given to the cluster for the years 2000, 2001, \dots , 2011. The b_1 cluster in 2001, for example, has formed out of the a_1 cluster of year 2000, plus some other clusters (which could even be a single node, considered as a cluster) that has joined a_1 in this period. The symbol x stands for a smaller cluster or a group of clusters in the network, which are not one of those that constitute the set of the top five considered in the evolution process. Similarly, in 2003, the d_1 cluster has formed out of the c_1, c_2, c_4, c_5 clusters of 2002 plus some other clusters not from the top five set. Running the picture backward in time, c_1 of 2002 is the b_1 of 2001 plus x , c_2 is of 2001 plus x , c_4 is b_4 of 2001 plus x , and finally, c_5 is b_5 of 2001 plus x . The other clusters (2, 3, 4, and 5) have not performed well in the competition for growth; they have been swamped by cluster 1 in almost all the periods from 2000 to 2011.¹⁹

In the Scopus India network as well as in both of the Scopus and WoS U.S. networks, there is a predominantly leading cluster. It is clearly the winner in 2011, swamping in size all the remaining four in the growth race. In the WoS India network, it is almost a similar situation, although in this case the other

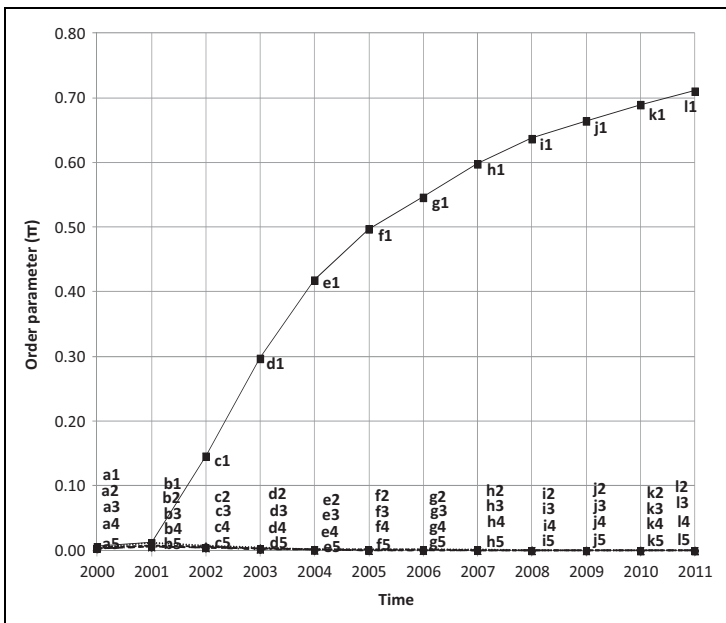


Figure 4. Five leading clusters (Web of Science, U.S.).

four clusters (2, 3, 4, and 5) were observed to have put up a more promising competition than that exhibited by the corresponding clusters in other three networks. It is important to note that, in the figures, the order parameter π is plotted against time (in years), and it is possible for π to decrease from one year to another in a situation in which the largest cluster’s size has not increased in proportion to the overall size of the full network. This can happen, for example, if many small, isolated clusters have formed in the network over the concerned period, without the largest cluster scaling proportionately in size so as to show a clear increase in π . The benefits and implications of tracking the growth and evolution of the large clusters in the network for formulating research policies will be discussed in the next section.

Cluster Growth Dynamics

A researcher’s perceived willingness to collaborate is what gives rise to the possibility of the leading cluster in the network to percolate. The first four scenarios described in the previous section combine ways to mathematically specify this distribution by starting with a simple case (scenario 1) and

Table 2. Details of Formation and Time Evolution of Five Leading Clusters in India and U.S. Networks.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Country												
Cluster												
India:												
Scopus	1 $a_1 = a_1$	$b_1 = a_1, x$	$c_1 = b_1, x$	$d_1 = c_1, c_2,$ c_4, c_5, x	$e_1 = d_1, d_2,$ d_3, d_4, x	$f_1 = e_1, e_2,$ e_4, e_5, x	$g_1 = f_1,$ f_2, f_4, x	$h_1 = g_1,$ g_2, g_3, x	$i_1 = h_1, h_2, x$	$j_1 = i_1,$ i_2, i_3, x	$k_1 = j_1,$ j_2, x	$l_1 = k_1, k_2,$ x
	2 $a_2 = a_2$	$b_2 = a_2, x$	$c_2 = b_2, x$	$d_2 = c_2, x$	$e_2 = e_2$	$f_2 = e_2, x$	$g_2 = g_2$	$h_2 = h_2$	$i_2 = h_2$	$j_2 = i_2$	$k_2 = k_2$	$l_2 = l_2$
	3 $a_3 = a_3$	$b_3 = a_2, x$	$c_3 = b_3, x$	$d_3 = d_3$	$e_3 = e_3$	$f_3 = f_3$	$g_3 = g_3$	$h_3 = g_4,$ x	$i_3 = i_3$	$j_3 = j_3$	$k_3 = i_3$	$l_3 = l_3$
	4 $a_4 = a_4$	$b_4 = b_4$	$c_4 = b_4, x$	$d_4 = d_4$	$e_4 = e_4$	$f_4 = f_4$	$g_4 = f_5$	$h_4 = h_4$	$i_4 = i_4$	$j_4 = i_5$	$k_4 = i_4$	$l_4 = k_5, x$
	5 $a_5 = a_5$	$b_5 = b_5$	$c_5 = b_5, x$	$d_5 = d_5$	$e_5 = e_5$	$f_5 = f_5$	$g_5 = g_5$	$h_5 = h_5$	$i_5 = i_5$	$j_5 = j_5$	$k_5 = i_5$	$l_5 = k_4$
India:												
WoS	1 $a_1 = a_1$	$b_1 = b_1$	$c_1 = b_3, x$	$d_1 = c_1, x$	$e_1 = d_1, x$	$f_1 = e_1, x$	$g_1 = f_1,$ f_3, x	$h_1 = g_1,$ g_2, x	$i_1 = h_1, x$	$j_1 = i_1, x$	$k_1 = j_1,$ j_2, x	$l_1 = k_1$
	2 $a_2 = a_2$	$b_2 = a_4, x$	$c_2 = b_1, x$	$d_2 = c_3, x$	$e_2 = d_3, x$	$f_2 = e_3, x$	$g_2 = f_2, x$	$h_2 = g_3,$ x	$i_2 = h_5, x$	$j_2 = i_2, x$	$k_2 = j_5, x$	$l_2 = k_2, x$
	3 $a_3 = a_3$	$b_3 = b_3$	$c_3 = b_2, x$	$d_3 = d_3$	$e_3 = d_2, x$	$f_3 = e_2, x$	$g_3 = f_4, x$	$h_3 = g_5,$ x	$i_3 = h_2, x$	$j_3 = i_3, x$	$k_3 = j_3, x$	$l_3 = k_3$
	4 $a_4 = a_4$	$b_4 = b_4$	$c_4 = b_4, x$	$d_4 = c_2, x$	$e_4 = d_4, x$	$f_4 = f_4$	$g_4 = g_4$	$h_4 = g_4,$ x	$i_4 = h_3, x$	$j_4 = i_5, x$	$k_4 = j_4, x$	$l_4 = k_4, x$
	5 $a_5 = a_5$	$b_5 = b_5$	$c_5 = c_5$	$d_5 = d_5$	$e_5 = e_5$	$f_5 = f_5$	$g_5 = g_5$	$h_5 = h_5$	$i_5 = h_4, x$	$j_5 = i_4$	$k_5 = k_5$	$l_5 = l_5$

(continued)

Table 2. (continued)

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Country Cluster												
U.S.: Scopos	1 $a_1 = a_1$ $a_2, a_3,$ a_4, x	$b_1 = b_1$ $b_2 = b_2$ $b_3 = b_3$ $b_4 = b_4$	$c_1 = b_1, b_2,$ $b_3, b_4, b_5,$ x	$d_1 = c_1, x$ $d_2 = d_2$ $d_3 = d_3$ $d_4 = d_4$	$e_1 = d_1, d_2,$ d_4, x $e_2 = d_3$ $e_3 = e_3$ $e_4 = e_4$	$f_1 = e_1, e_2,$ $e_3, e_4,$ e_5, x $f_2 = f_2$ $f_3 = f_3$ $f_4 = f_4$	$g_1 = f_1,$ f_2, x $g_2 = g_2$ $g_3 = g_3$ $g_4 = f_4$	$h_1 = g_1,$ g_4, x g_5, x $h_2 = h_2$ $h_3 = g_3,$ x $h_4 = g_2,$ x $h_5 = h_5$	$i_1 = h_1, h_2$ $i_2 = i_2$ $i_3 = i_3$ $i_4 = i_4$ $i_5 = i_5$	$j_1 = i_1, i_2,$ x $j_2 = j_2$ $j_3 = j_3$ $j_4 = j_4$ $j_5 = j_5$	$k_1 = j_1,$ $j_3, j_4,$ j_5, x $k_2 = j_2$ $k_3 = k_3$ $k_4 = k_4$ $k_5 = k_5$	$l_1 = k_1, k_2,$ $k_3, k_4,$ k_5, x $l_2 = l_2$ $l_3 = l_3$ $l_4 = l_4$ $l_5 = l_5$
U.S.: WoS	1 $a_1 = a_1$ b_3, x	$b_1 = b_1, b_2,$ b_3, x	$c_1 = b_1, b_2,$ $c_3, c_4, c_5,$ x	$d_1 = c_1, c_2,$ $c_3, c_4, c_5,$ x $d_2 = d_2$	$e_1 = d_1, d_2,$ $d_3, d_4,$ d_5, x $e_2 = e_2$	$f_1 = e_1, e_2,$ $e_3, e_4,$ e_5, x $f_2 = f_2$	$g_1 = f_1,$ f_5, x $g_2 = g_2$	$h_1 = g_1,$ g_2, x $h_2 = g_2,$ x $h_3 = h_3$ $h_4 = h_4$ $h_5 = h_5$	$i_1 = h_1, h_2,$ h_4, h_5, x $i_2 = i_2$ $i_3 = i_3$	$j_1 = i_1, i_2,$ x $j_2 = j_2$ $j_3 = j_3$ $j_4 = j_4$ $j_5 = j_5$	$k_1 = j_1, x$ $k_2 = k_2$ $k_3 = k_3$ $k_4 = k_4$ $k_5 = k_5$	$l_1 = k_1, k_2,$ $k_3 = k_3$ $l_2 = l_2$ $l_3 = l_3$ $l_4 = l_4$ $l_5 = l_5$

Note: The letter x symbolizes isolates or smaller sized clusters which are not part of the set of the leading five clusters in the network. WoS = Web of Science

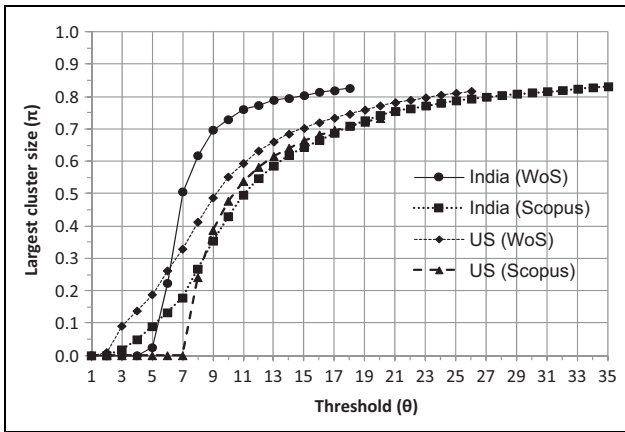


Figure 5. Order parameter against willingness threshold in scenario 1.

incrementally incorporating more realistic complexities into the model. The idea is to see what kinds of cluster evolution patterns emerge from their execution. As we see in this section, the results of most of these simulations have realistic features, in one form or another, that can be used to explain cluster evolution behavior in simple situations. This knowledge is of great value in model building. However, in major practical applications, not one but rather a combination of these features are expected to contribute, most frequently, in a nontrivial way to the cluster-building process. Scenario 5 is a realistic, albeit limited, case based on real willingness data collected from a small sample of researchers.

Scenario 1. Scenario 1 depicts a situation in which a researcher's collaboration willingness drops abruptly to zero right after their current number of collaborative ties crosses a threshold. Although rather unrealistic in practice, the scenario is theoretically intuitive and serves as a toy example to validate the behavior of cluster growth and evolution in our model. In Figure 5, we plot the order parameter π against different values of the collaboration threshold θ for all four networks considered in this study. Qualitatively, all the graphs exhibit similar behavior. Initially, small values of θ yield small values of π , as a network is largely fragmented at this stage. However, with θ increasing subsequently, π rises steeply, as more and more clusters are joined to form a large cluster in the network. This phenomenon is caused by a "probabilistic attraction" mentioned earlier, where an increasing number of clusters (including isolates) exhibits a

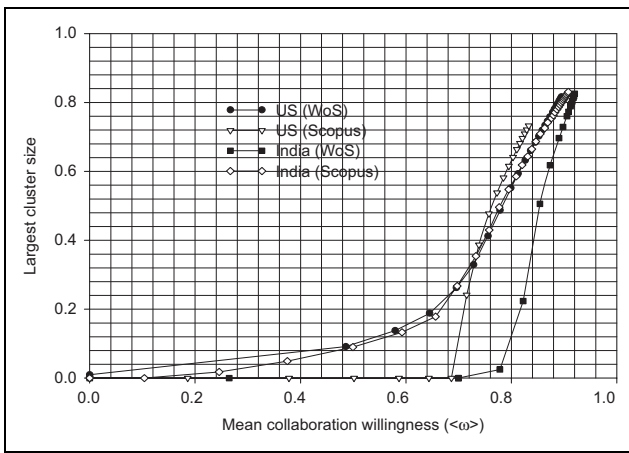


Figure 6. Order parameter against average willingness in scenario 1.

tendency, on the average, to connect to the leading cluster in the network. When the threshold becomes moderately high, the clusters are already well formed in the network, and the π curve reveals a saturation plateau, its rate of increase becoming smaller with further increase of θ .

Figure 6 shows the behavior of π with the mean collaboration willingness $\langle\omega\rangle$. As mentioned earlier, $\langle\omega\rangle$ gives the mean proportion of willing candidates available for new collaboration. When it is small, there are few researchers in the network willing to start new collaboration. Therefore, the probability of formation of new connections among the currently existing isolated clusters is small, and the corresponding π value is small on the average. With more researchers willing to collaborate, the order parameter rises sharply. Alternatively, it is possible to look at the scenario in terms of a cluster *breakdown* process, leading to a loss of connectedness in the collaboration network. Thus, as higher and higher fractions of researchers become unwilling to form collaborative ties, the giant cluster begins to fall apart. In Figure 6, the decline in the value of π is seen to be the sharpest for average willingness lying in the range of (0.7, 0.9). Among the four networks, the fall is most rapid for the India WoS and the U.S. Scopus networks.

Scenario 2. In this scenario, every researcher in the network is perfectly willing to build new collaborative ties before they reach their threshold willingness. However, as this threshold is crossed, researchers may still be somewhat willing to collaborate, but this willingness falls very sharply (but

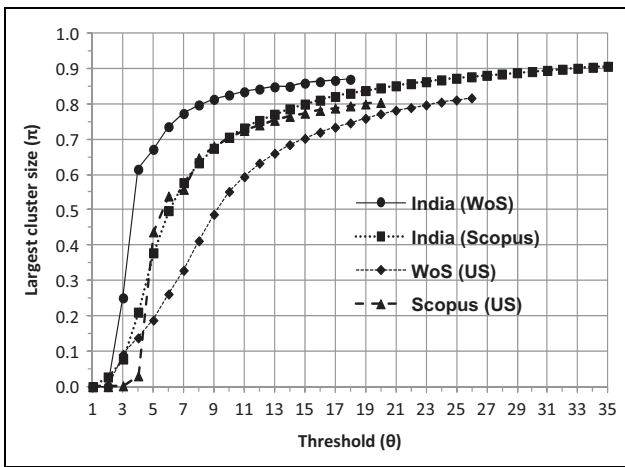


Figure 7. Order parameter against willingness threshold in scenario 2.

continuously) with their current degrees in the network. From an initially fragmented network, the order parameter rises very sharply from a very small value as the threshold is made higher and higher. With every researcher in the network perfectly willing to start new collaborations, the isolated clusters are bridged very quickly. Subsequently, as large values of θ are reached, the increase of π slows down. However, even at a large threshold, π does not decrease. This is due to the fact that a significant contribution ($\omega = 1$) to π comes from the low end ($k < \theta$) of the willingness function, over and above the large contribution to it that comes from the $p(k)$ distribution for small k . This behavior, seen in Figure 7, is very similar to what was found in scenario 1. The qualitative behavior for a sudden drop in willingness in scenario 1 and an exponential drop in the present case is not significantly different.

In Figure 8, we show π against the mean collaboration willingness $\langle \omega \rangle$. This behavior is also similar to that in scenario 1. As $\langle \omega \rangle$ increases, isolated clusters are bridged, and a large-sized cluster starts to form in the network quite rapidly. Alternatively, with a functional large cluster operating in the network, if $\langle \omega \rangle$ starts to decrease as previously willing researchers become unwilling to form new collaboration and are effectively removed from the network, the large cluster begins to fall apart, and π drops very sharply.

Scenario 3. Figures 9 and 10 exhibit the simulation results of this scenario. Initially, the network is largely fragmented, populated primarily by isolates

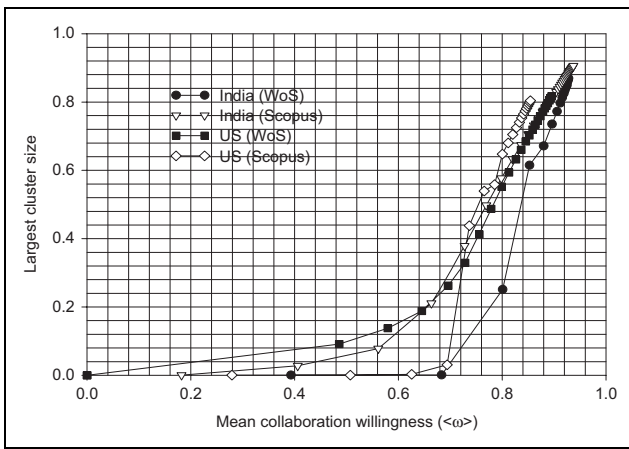


Figure 8. Order parameter against average willingness in scenario 2.

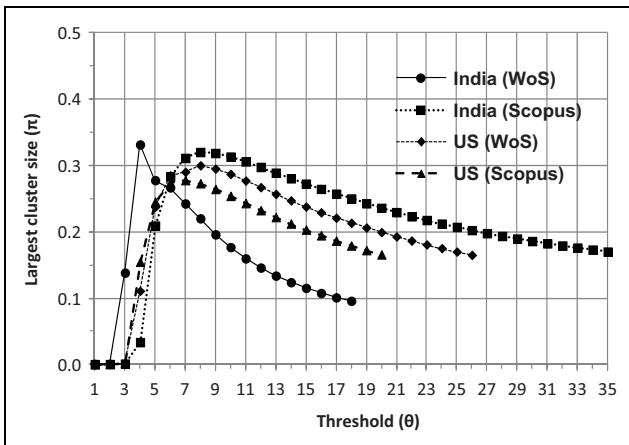


Figure 9. Order parameter against willingness threshold in scenario 3.

and, perhaps, some small-sized clusters. For a specific value of the threshold θ , as the willingness to form new collaborative ties increases linearly with the current degree of a researcher, there is an increasing probability that previously disconnected clusters will be bridged through one or more of these new ties formed in the network. On an average, a particular researcher willing to form new collaborative ties with others in the network is highly likely to be

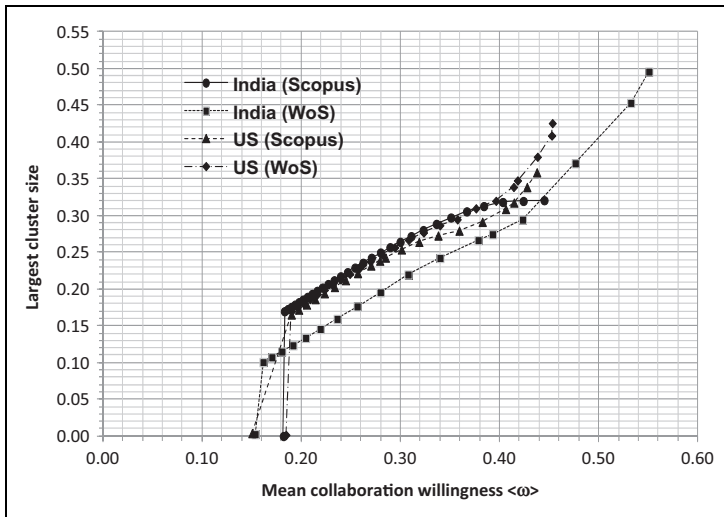


Figure 10. Order parameter against average willingness in scenario 3.

connected to one or more of the network's large-sized clusters. This bridging process continues until the threshold is crossed, after which collaboration willingness falls exponentially fast, and the size of the order parameter increases very slowly, if at all.

When θ is set to high values, researchers may remain willing to forge new collaborative ties even at large values of their current degrees. However, the actual degree distributions of the MGMT networks are not perfect power laws but have finite cutoffs. Therefore, most of the degrees in the network are confined to intervals around the mean degree (which is rather small), and there are only very few researchers with really large degrees in all our MGMT networks for both countries. When the threshold is set to a high value, there is then a negligible contribution to the order parameter from the distribution's tail. With increasing threshold, π therefore declines. The particular value of θ that signals the onset of this decline is unique to the topological structure of the network. These values are 4, 8, 7, and 7 for the WoS India, Scopus India, WoS U.S., and Scopus U.S. networks, respectively. This behavior of π is clearly visible in all four of our MGMT networks in Figure 9. This is in sharp contrast to the behavior exhibited by scenarios 1 and 2 above, in which there is always a uniformly large contribution to π from the low end of the distribution ($k < \theta$), where $p(k)$ is relatively high and $\omega = 1$ consistently throughout this range. In this case, π does not decrease even at high values of

the threshold, although its rate of increase becomes increasingly smaller, which characterizes the saturation plateau seen in the figure.

The behavior of the average collaboration willingness (i.e., the average fraction of researchers who are still willing to collaborate), given by $\langle \omega \rangle = \sum_{k=0}^{\infty} p(k) \omega(k, \theta)$, is characterized in this case by a slow fall of $\langle \omega \rangle$ with increasing θ . The fall pattern is unique to the specific topology of a network. The θ values are 2, 5, 4, and 3 for the WoS India, Scopus India, WoS U.S., and Scopus U.S. networks, respectively. For small values of willingness, the network is largely fragmented, consisting primarily of isolates and small clusters. As $\langle \omega \rangle$ increases, the probability of bridging links among the disjoint clusters also increases, and a large cluster begins to take shape. The dynamics of this process determines whether or not the leading cluster percolates in the long run.

On the other hand, in a cluster dissolution scenario, as the willingness threshold decreases from a high value with previously willing researchers now becoming unwilling to collaborate on the average, they and their associated ties are effectively removed from the network, and π immediately starts to decline. Figure 10 exhibits the simulation results pertaining to this behavior. In all our networks in this scenario, the order parameter initially declines almost linearly as mean willingness falls in the network, but eventually a catastrophic breakdown of large-scale connectivity is observed. This is a unique characteristic of this scenario, in that the initial rate of decline of π with decreasing $\langle \omega \rangle$ is much slower than what was encountered in the previous two scenarios. For example, the average decline rate in the linear portion of the curve is about 80 percent steeper in scenario 2 than in scenario 3.

Scenario 4. In this scenario, Figures 11–14 show the behavior of growth shrinkage of the giant cluster with rising adversity values for fixed collaboration willingness thresholds. As we hypothesized before, in all of the four cases displayed, the trend in the fall of the order parameter with increasing adversity is quite clear. Overall, the decline is nonlinear: for initially low values of adversity, the order parameter declines slowly, but thereafter, its fall is somewhat close to linear. The average fractional percentage drop in π per unit increase in adversity is small: about 4 percent for WoS India, 3 percent for WoS U.S., 1 percent for Scopus India, and 2 percent for Scopus U.S. Note that the maximum values of the threshold are all different in the four cases. We also keep well away from the asymptotic regime realized at very high values of the degree by limiting our simulations to a range below 50 percent of θ_{max} .

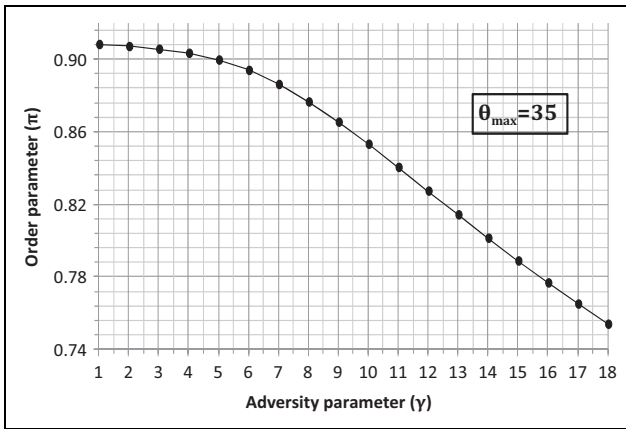


Figure 11. Order parameter against adversity for fixed willingness threshold in scenario 4 (Scopus India).

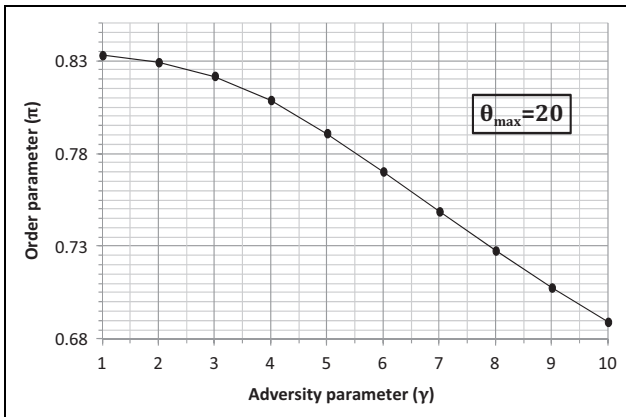


Figure 12. Order parameter against adversity for fixed willingness threshold in scenario 4 (Scopus U.S.).

Scenario 5. The results of the bootstrapped Monte Carlo simulations pertaining to scenario 5 are exhibited in Figure 15 for the Scopus and WoS India networks. The figure shows the order parameter π plotted against the willingness tolerance η centered on the actual data points, which are a set of values of actual researchers' perception of collaboration willingness, conditioned on

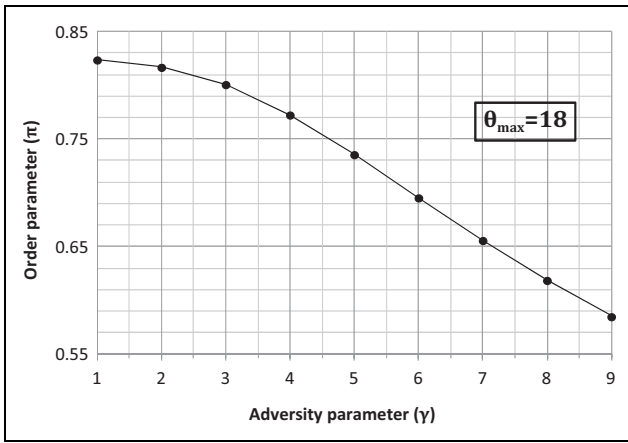


Figure 13. Order parameter against adversity for fixed willingness threshold in scenario 4 (Web of Science India).

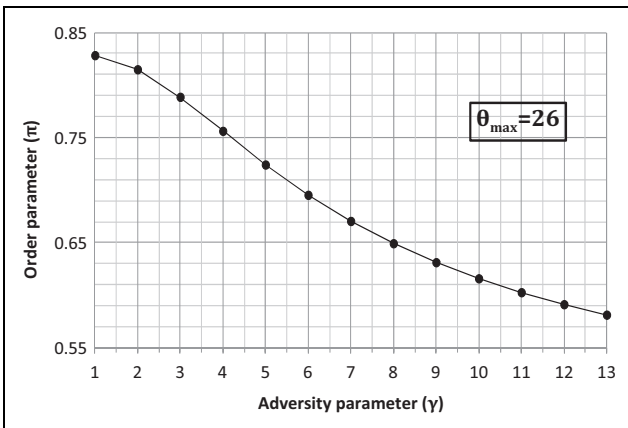


Figure 14. Order parameter against adversity for fixed willingness threshold in scenario 4 (WoS U.S.).

their last three years’ engagements in collaborative research activities. For both Scopus and WoS, the average fractional change in π per unit tolerance starting from $\eta = 0.1$ is close to about 1 percent. Thus, in both cases, π does not appear to be overly sensitive to increasing tolerance limits, which signifies that the structure of the giant cluster is robust against limited variations

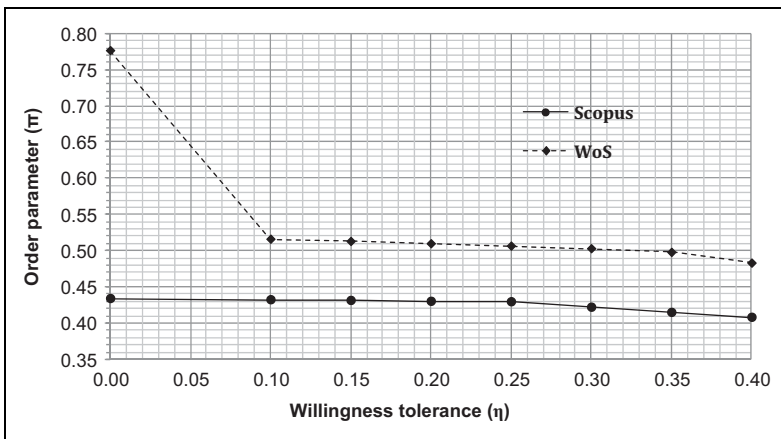


Figure 15. Order parameter against willingness tolerance in scenario 5 (Scopus and Web of Science India).

in assessing the willingness profile of actual researchers. In consequence, if the existing giant cluster fills a significant volume of the entire network, then this result indicates that the network possesses a highly desirable resilient structure, in which the giant cluster remains in place even when there are some deviations in the practical fulfillment of researchers' actual collaborative engagements.

Discussions of Results

Utility of Cluster Tracking

In this work, we examined the problem of large-scale structure formation and growth in collaboration coauthorship networks, which are driven by an underlying process of socioprofessional interactions among research collaborators. In a sociological study of the structure formation process in this type of network, the cluster-tracking method is an excellent utility for individually recording the temporal growth of some of the leading clusters in the network starting from some initial instant of time. This knowledge is useful for ascertaining how the network eventually approaches a stage in which it operates in the percolating regime containing a fully operational giant cluster. The specific points in the developmental phases of this cluster at which important links form between this cluster and previously disconnected cluster sets are also crucial elements contributing to the study of the ongoing

growth process in the network. The particular researchers that constitute these critical connectivity points in the giant cluster possess high values of betweenness (irrespective of whether they possess high degrees).²⁰

In almost all of the networks we studied over a 12-year window of time, the final giant cluster emerged as a consistently clear winner in the race for growth, dwarfing the next four clusters by a wide margin. This points to the predominance of a largely connected community of researchers in the networks that grew considerably bigger and bigger over time through the addition of new entrants to the research fields as well as existing researchers from other smaller clusters in the networks. Interestingly, in the WoS India network, although the final giant cluster was once again the clear winner, the next four clusters also exhibited some competitive growth. This network is much smaller in size than the other three and is also of more specialized categorization of fields in the WoS electronic database.

Willingness Profile and Large-scale Structure Formation

Rather than focusing on the underlying mechanism of structure formation (the *why*), we concentrated our attention in this study on the sociological aspects of the problem (the *how*) by looking at the dynamics of how clusters connect to one another through internal bridging links that form between them in course of time. The primary observable in this dynamics is what we have modeled by means of the willingness profile distribution, which is a measure of a researcher's perception of their eagerness and capacity for forging new collaborative ties with other willing researchers in the network. Unlike a causal mechanism that begets large-scale cluster growth, the willingness function is not a mechanistic device per se. Instead, it is a manifestation of an underlying sociological characteristic of the actors in the network. The entire unfolding process, at an aggregate level, is stochastic.

After ascertaining the existence of a definite giant cluster in each of the networks under study, we performed numerical simulations in a number of different scenarios to try to understand the characterization of the process of large-scale structure formation in the networks. In this regard, cleverly devising a willingness profile distribution function is the key to successfully approximating the actual pattern of cluster evolution in the networks. The scenarios we formulated to do this job capture different aspects of this behavior. Starting from a simple scenario, we incrementally built more realistic, albeit complicated, scenarios that approximate the actual growth of the giant cluster in the real networks.

In the settings of scenarios 1 and 2, each researcher in the network has a willingness profile in which there is perfect willingness to collaborate up to the level of the threshold θ , independent of the current local collaboration ties of researchers. This makes the order parameter π grow with progressively higher values of θ . The scenarios, however, differ in their individual patterns of decline of the willingness after the threshold is crossed. As π reaches saturation, the decrease in willingness, independent of whether it is abrupt or smooth, does not cause the order parameter to drop significantly, although its growth rate visibly slows down, and it may ultimately become negative at very large values of θ . This type of behavior is vastly different from the one that arises from a willingness profile that depends on the current degree values of collaborators below the threshold level. The simplest case pertaining to this characteristic behavior is modeled in scenario 3 by a profile that is linearly increasing with degree below the threshold. The order parameter cannot saturate in this case but rapidly falls off immediately as the threshold is crossed.

It is important, in particular, to compare these three scenarios with scenario 5, which is a bootstrap Monte Carlo realization of a sample of willingness profiles based on actual data obtained from a sample of researchers from the India network. The bootstrap results show that the order parameter is not much affected by limited deviations in the actual willingness profiles of researchers. Thus, when the giant cluster occupies a considerable fraction of the network's volume, its large-scale structure is robust against errors made in assessing the probabilities of researcher's willingness to undertake new collaborative activities. This result is reassuring, since the actual assessment of these probabilities depends on subjective elements arising out of researchers' current perception of collaboration willingness that projects into a future state of the network.

In this simulation, one can conceive of two independent cases: (1) resampling of actual data points (used in this work), and (2) resampling of actual data points as well as inclusion of intermediate points from the profile. There is a conceptual difference, however, between the two cases. If we consider only the first case, then we are not supposed to use any willingness values between the sample points. Given that we use only these points when computing π , the profile function, strictly, contains no information about the values in between. The function might take any form whatsoever between the sample points and we would never know these values, since these intermediate points did not exist in the sample in the first place. Nevertheless, if a willingness profile is reasonably smooth, with no major irregularities between sample points, then knowing the values at the sample points only is sufficient to get a picture of the general

shape of the profile. For some of the intermediate points not present in the sample, the $p(k)$ values as well as the corresponding $\omega(k)$ values may be found to be high. The ξ value obtained as a root of the polynomial equation for such a point makes the π value larger in this case. For example, we may write $\pi^{(0)} = \sum_k p(k)\omega^{(0)}(k) - \sum_k p(k)\omega^{(0)}(k)\xi_{(0)}^k$ and $\pi^{(\varepsilon)} = \sum_k p(k)\omega^{(\varepsilon)}(k) - \sum_k p(k)\omega^{(\varepsilon)}(k)\xi_{(\varepsilon)}^k$ for two-order parameters corresponding to $\eta = 0$ and $\eta = \varepsilon$, respectively. Then, $\pi^{(\varepsilon)} - \pi^{(0)} = \sum_k p(k)[\{\omega^{(\varepsilon)}(k) - \omega^{(0)}(k)\} - \{\omega^{(\varepsilon)}(k)\xi_{(\varepsilon)}^k - \omega^{(0)}(k)\xi_{(0)}^k\}]$. For the intermediate points not in the sample, $\omega^{(0)}(k) = 0$. In this case, $\pi^{(\varepsilon)} > \pi^{(0)}$ for all k values, because $\xi_{(\varepsilon)}^k < 1$. For the original resampled points, however, $\pi^{(\varepsilon)} > \pi^{(0)}$ holds only for those k values for which $\omega^{(\varepsilon)}(k) - \omega^{(0)}(k) > \omega^{(\varepsilon)}(k)\xi_{(\varepsilon)}^k - \omega^{(0)}(k)\xi_{(0)}^k$. This depends on the specific values of ξ , where the corresponding polynomial equations in the two cases make a zero crossing in the interval $[0, 1]$. In practical situations, all we can get are a finite number of sample points and not really a nice continuous function; for this kind of data, our method of using the spline interpolation works comfortably well.

Scenario 4, by contrast, examines an altogether different sociological aspect of research collaboration. Rather than focusing on the variations in threshold level, it examines the behavior of π for different values of adversity in the research environment for a fixed value of the threshold level. A researcher works in an environment that is frequently constrained by a number of external factors, which include, for instance, teaching responsibilities, administrative duties, student advising as well as various personal obligations. The effects of these factors on the freedom to do collaborative research are aggregated into the adversity parameter in our model. With no adversity present (an ideal situation), the willingness profile is a step function with a uniform maximum value all the way up to the threshold level. With increasing influences of external factors, the profile is moderated by the adversity parameter, causing the order parameter to exhibit a growing shrinkage. However, on account of the nonuniform decrease in the profile around θ , the decline of the order parameter becomes generally nonlinear.

This type of behavior is exhibited, almost uniformly, by all four of the networks studied in this work; it can therefore be regarded as a general feature of these networks. Although we have modeled only the average adversity in this work, it is possible as well as useful to extend the model by disaggregating the entire effect into separate, individual influences of all external constraints, resulting in different adversity levels at which researchers perform

collaborative research activities in the prevailing research environment of an institution or a country. This model, albeit more realistic, is a little harder to conceptualize and formulate; however, its numerical implementation should be straightforward.

Sociological Implications

As mentioned earlier, the evolution of clusters in a research collaboration network is a result of an intrinsically competitive dynamics, although the underlying process, per se, may not necessarily be generated by a purposive social force. Two centrality measures commonly contribute to growth in this regard—the degree and the betweenness of the researchers. In the context of collaboration, influential researchers who are collaboratively associated with many others in the network play the role of attractors or star performers, drawing existing as well as new researchers toward them. The underlying association is one of cumulative advantage (Allison et al. 1982; Merton 1957; Simon 1955), which helps to enlarge the ego-centered clusters of these researchers over time. On the other hand, collaborative brokers, surrounded by structural holes, become influential as well by dint of their high betweenness values (Fleming et al. 2007). Even possessing low degree centrality in some cases, these researchers serve to bridge previously disconnected clusters, thereby enlarging the average cluster size in the network (Moody and White 2003; Reagans and McEvily 2003).²¹ These two processes are simultaneously at play in most common cases of evolution. Using our cluster-tracking algorithm, we found that both these processes contribute heavily to the evolution of the leading cluster, causing it to eventually cross the percolation threshold. Thus, even in a network with a highly fragmented topology at some initial time (which is the case with all our networks in this study), a small number of clusters is seen to grow quickly in time. Finally, if the network enters the percolating regime, there is one clear winner.

Although there are a number of models of cluster evolution in a social network, it is frequently unclear as to what constitutes the *fundamental* mechanism of growth in a particular social context. Commonly, not one but multiple, highly interlinked processes contribute. Instead of adopting this mechanistic approach, we have focused our attention in this work on a behavioral aspect of evolution through a researcher's perceived willingness to collaborate, which is the manifestation of a researcher's socioprofessional attitude toward collaborative research activities (Powell et al. 2005). This not only engages a researcher in the excitement of doing research through the mixing of diverse domains of knowledge and technical expertise available from collaborators,

it also opens up new channels for them to build their social capital (McFadyen and Cannella 2004; Walker et al. 1997). The existence of a giant cluster is a precondition for ensuring the necessary large-scale connectivity needed to build and expand this capital. In our present approach to cluster evolution, the clusters, as units of analysis, do not evolve through the imposition of a predefined set of rules. Rather, they evolve *adaptively* as researchers connect to one another based on their willingness to collaborate. We are, therefore, not concerned about whether specific ideas or schools of thoughts will get passed on over time through cohesive embedding in local groups through star performers or otherwise but about the search, flow, access, and the mixing of ideas in a collaboration context. Importantly, only when a network operates in the percolating regime, these benefits can be reaped. For example, the accessibility and the flow of information and resources benefiting a researcher is restricted in practice to investigating whether the network's percolating cluster, if it exists, possesses the small-world property (Watts 1999). Although cohesion and group embeddedness are important structural characteristics, their actual impact on a collaboration network's large-scale connectivity is neither transparent nor particularly relevant to the present context.²²

Our first three scenarios are based on the assumption that the willingness profile, under ordinary circumstances, depends on the number of present collaborators of a researcher as well as on their maximum capacity to collaborate. This assumption does not take into account the presence of the external environment in which a researcher performs collaboration research. In scenario 4, we made the additional assumption that the work environment can also influence a researcher's willingness profile distribution. Thus, subject to the constraint imposed by the collaboration threshold, a congenial or supportive work environment may go a long way to enhance a researcher's collaborative activities above what is normally possible in a neutral environment. Conversely, a nonsupportive environment, often realized through institutional pressures for productivity, growth, and functionalism; rapid technological changes; and constraints in the funding situations, may reduce collaborative activities to a level below what is ordinarily achievable in a neutral environment (Lieberson and Lynn 2002; Shrum et al. 2007). The adversity parameter models this situation in an aggregate way. Our simulations based on the assumed profile supports this hypothesis. However, in the current context, the effect is incorporated into the profile only in an aggregate way. This approach, therefore, must be regarded only as an approximation, which does not allow the collaboration environment to be subdivided into components for which identifiable sources of adversity can be modeled separately.

Conclusions, Limitations, and Future Directions

In this work, we introduced techniques in sociological research methods to perform a behavioral analysis of cluster formation, growth, and percolation in social networks. The methodology and the techniques can be applied to a wide variety of research problems in which specific domains are sociological research. The accompanying numerical metrics are utilized to explore as well as to confirm the validity and applicability of many observational variables of practical significance. We specialized in a cross-country, comparative study of the emergence of large-scale structures in collaboration, coauthorship networks of researchers by investigating the growth and evolution patterns of the clusters in these networks. Assuming the existence of some underlying social force that is responsible for growth in these networks, we adopted a behavioral perspective to examine how some small-sized, isolated clusters at an early instant of time grow in size through the merging of multiple such clusters, as crucial intercluster bridging links are formed subsequently between them. Typically, one cluster among the leading ones may eventually win the growth race by percolating through the network, spanning it and filling up a significant volume of it. Our findings in this direction of investigation contribute to a better understanding of the dynamics of cluster growth and evolution in social networks and should stimulate future research to integrate the actual causal mechanisms for network growth and the sociological manifestation of large-scale structural patterns in these networks.

This study, at its present stage, has a few limitations that offer fruitful opportunities for future research in this area. First, it is important to realize that the methods introduced in this work apply to a specific research problem at an aggregate level. In other words, the final outcome pertains to the structure of an entire network (hence, a large community, organization, or society). It is not the study, per se, of an ego-centered group. Of course, it is the individual dynamics that underlies the fundamental mechanism responsible for the social attraction among individual actors mentioned earlier in the article. In this work, we have assumed its existence as providing the background of the cluster formation process and coarse-grained over the individual interactions. In the spirit of an agency-structure integration problem, we will show, in a future work, how to make this connection between the micro and the macro structures of social interactions.

Next, we note that the willingness parameter, which contributes to the formation of bridging links between disconnected clusters, has a subjective element built into its actual phenomenological assessment. It actually is the perception of a researcher of their willingness to collaborate at the present

time or in the near future, given their current involvement in collaborative research activities. Thus, two researchers with comparable present involvement in collaboration research may possess widely dissimilar perceptions of future research interests. Researchers also vary widely in their individual attitudes toward research. For example, there are researchers who intend to build large collaboration groups, although they may not personally undertake a significant amount of research activities themselves. Rather, they tend to play the role of collaborative brokers, bringing other researchers together to collaborate through their critical intermediation. The collaboration willingness of these individuals is usually high and does not depend significantly on the current level of their research involvement. By contrast, some researchers who already have high current involvement may attempt to grossly reduce, or at least limit, their future collaborative activities and consequently exhibit low levels of willingness. To obtain accurate results useful in simulation works, large data sets containing a wide distribution of the willingness measure must be used. In the absence of a known distribution of population values of willingness, the Monte Carlo bootstrap procedure used in the present work is a useful parametric estimation technique. However, since our current data set is rather small, the statistical assessment of the order parameter is not highly accurate at this stage. Also, we could only employ a small set of primary data collected from researchers in India to witness the behavior of the order parameter. Besides enlarging this set in the future, it is also necessary to obtain primary willingness data from researchers in the United States. In the absence of this component, our cross-country comparative analysis must therefore be acknowledged to have remained only partially complete. At this time, our work is under way to make this assessment more robust.

To simplify our analysis in this work, we made use of a constant collaboration threshold. In reality, it is conceivable that the threshold depends on the time. The resulting simulations will have further complications in this case. Even more importantly, we have assumed in scenarios 1–4 that all researchers behave in the same way with respect to the individual capacities to shape their collaborative activities. In practice, there certainly are deviations from this behavior. Besides, we have not considered past history of collaboration, the availability of collaboration resources, and the homophily on various demographic and geographic dimensions.^{2,3} However, it is also important to note that the network operating in a percolating regime has characteristics that are not dependent on specific initial conditions pertaining to group embeddedness, local cohesion, and homophily.

Finally, we have considered the adversity parameter only at an aggregate level. In this sense, it is only a *single* parameter that represents all manners of

external difficulties encountered by researchers in their collaboration research engagements. In reality, there could be many sources to limit their collaborative activities. The individual impacts of the separate sources on the order parameter are expected to contribute at different levels. The aggregation philosophy is, therefore, only an approximation. Computationally, it is straightforward to disaggregate the general effect and incorporate the separate sources into an empirical willingness profile analysis. However, it is quite difficult to perform an appropriate conceptual assessment of the individual effects of all the sources that contribute to the difficulties encountered in collaboration research.

Ongoing work at this time involves incorporating these enhancements into the present model. Primary data collection from researchers based in the United States is also in active contemplation.

Acknowledgments

The authors wish to acknowledge and thank three anonymous reviewers of SMR for suggesting a number of improvements in the paper.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Jaideep Ghosh would like to thank the Department of Science and Technology, Government of India, for financial support of the Ramanujan Fellowship to carry out this work.

Notes

1. A cluster, also called a network *component*, is a connected set of actors in the network. The term *cluster* is usually used in the context of percolation problems in networks (Albert and Barabási 2000; Callaway et al. 2000).
2. A single actor as an *isolate* in the network is also a cluster in our view.
3. There are a number of important earlier studies of the growth mechanisms in social networks (Albert and Barabási 2000).
4. This is true of most works of collaboration. There are a few exceptions, however. For example, a paper published in Elsevier's journal *Physics Letters B* (Compact Muon Solenoid [CMS] 2012) on the CMS collaboration experiment at the Large Hadron Collider to detect the Higgs boson has listed about 2,893 authors from

- 63 different countries of the world. The social relevance of these types of extremely large collaboration projects is unclear (Newman 2001).
5. Still shorter periods using data over four or six months are theoretically possible, but, unfortunately, the listings of the journals in the actual databases are incompletely indexed over such short periods.
 6. These clusters were not one of the initial clusters in our fixed sample at previous instances of time.
 7. In recent survey interviews of researchers in India, we have found that many researchers, especially those who hold permanent faculty positions in various institutions in India, also have to perform a number of administrative functions. Most of these researchers have expressed concerns to us to the effect that, it frequently becomes difficult for them to balance administrative duties with collaborative research (and teaching) activities.
 8. There may be more than one in rare cases.
 9. These, then, are the second-order neighbors of the focal researcher.
 10. Callaway et al. (2000) have found similar results by the use of secondary generating functions.
 11. The root of this equation is actually a probability, which is restricted to this range.
 12. The linear function is the simplest behavior used as a first approximation to the functional form. Other empirical functions or parameter-dependent functional forms may also be used.
 13. For nonintegral s , the polylogarithm function is referred to as the Jonquière function (Maximon 2003).
 14. As we have noted in a number of interviews of a subset of researchers in India, the unavailability of sufficient research funds in recent years prevents many researchers from attending national and, especially, international conferences, which provide worthwhile opportunities for researchers to come into contact with other researchers to start new research collaboration projects.
 15. The coauthorship bibliometric data were first collected from the Scopus and WoS databases for this period. Subsequently, the electronic data were confirmed and validated directly through interviews of the selected sample of researchers.
 16. For all researchers having the same degree value in the network, we take the average of the values of their perceived collaboration willingness as the representative willingness value for that degree.
 17. If the ω value obtained in this way falls beyond the upper and lower bounds of the $[0, 1]$ probability range, we reject the point and draw again.
 18. In the actual program implementation, we used $r = 10,000$.
 19. It is important to recognize that our networks, and hence the clusters in them, have been built by fixing a specific window of time and using cross-sectional data which are sectionally segmented over short periods of one year and

subsequently pieced together. To a certain extent, therefore, cluster formation in a network of this type depends on the initial clock time (here, the year 2000), and a particular configuration of clusters in a network at a certain period of time may be the result of a particular historical accident. Be that as it may, the tracking technique we employ here does not depend on the initial time, but different clusters may have different temporal predominance if clocks are started off at different initial instants of time.

20. In this article, although we have desisted from identifying by naming individual researchers who serve as the endpoints of the bridging links, it is easy to do so in any specific practical application of the cluster-tracking procedure.
21. Using search criteria based on fields or specialization areas within MGMT, we have found through cluster tracking that many of these bridging connections are due to interdisciplinary research work within MGMT (particularly in the U.S. networks); for example, researchers working in finance have collaborated and published with researchers in organizational behavior or strategy.
22. We investigated these structural effects in earlier papers (Ghosh and Kshitij 2014; Ghosh, Kshitij, and Kadyan 2015).
23. We thank a reviewer for pointing out this limitation to us.

References

- Abramowitz, M. and I. A. Stegun. 1965. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications
- Acedo, F. J., C. Barroso, C. Casanueva, and J. L. Galán. 2006. "Co-authorship in Management and Organizational Studies: An Empirical and Network Analysis." *Journal of Management Studies* 43:957-83.
- Ahuja, G. 2000. "Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study." *Administrative Science Quarterly* 45:425-55.
- Albert, R. and A.-L. Barabási. 2000. "Topology of Evolving Networks: Local Events and Universality." *Physical Review Letters* 85:5234-37.
- Allison, P. D., S. J. Long, and T. K. Krauze. 1982. "Cumulative Advantage and Inequality in Science." *American Sociological Review* 47:615-25.
- Barabási, A.-L., H. Jeong, R. Ravasz, Z. Néda, T. Vicsek, and A. Schubert. 2002. "On the Topology of the Scientific Collaboration Networks." *Physica* 311:590-614.
- Batagelj, V. and A. Mrvar. 2015. Pajek—Program for Large Network Analysis. Retrieved January 2, 2014 (<http://vlado.fmf.uni-lj.si/pub/networks/pajek>).
- Bozeman, B. and E. Corley. 2004. "Scientists' Collaboration Strategies: Implications for Scientific and Technical Human Capital." *Research Policy* 33:599-616.
- Burt, R. S. 2000. "The Network Structure of Social Capital." *Research in Organizational Behavior* 22:345-423.

- Callaway, D. S., M. E. J. Newman, S. H. Strogatz, and D. J. Watts. 2000. "Network Robustness and Fragility: Percolation on Random Graphs." *Physical Review Letters* 85:5468-71.
- Capocci, A., V. D. P. Servedio, F. Colaiori, L. S. Buriol, D. Donato, S. Leonardi, and G. Caldarelli. 2006. "Preferential Attachment in the Growth of Social Networks: The Internet Encyclopedia Wikipedia." *Physical Review E* 74:1-6
- CMS (Compact Muon Solenoid). 2012. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC." *Physics Letters B* 716:30-61.
- Coleman, J. S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology* 94:S95-120.
- Crane, D. 1972. *Invisible Colleges; Diffusion of Knowledge in Scientific Communities*. Chicago: University of Chicago Press.
- Cummings, J. N. and S. Kiesler. 2007. "Coordination Costs and Project Outcomes in Multi-university Collaborations." *Research Policy* 36:1620-34.
- Cvijović, D. 2007. "New Integral Representation of the Polylogarithm Function." *Proceedings of the Royal Society A* 463:897-905.
- Durkheim, E. 1984. *The Division of Labor in Society*. Translated by W. D. Hall. New York: The Free Press.
- Efron, B. and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Fleming, L., S. Mingo, and D. Chen. 2007. "Collaborative Brokerage, Generative Creativity, and Creative Success." *Administrative Science Quarterly* 52: 443-75.
- Ghosh, J. and A. Kshitij. 2014. "An Integrated Examination of Collaboration Coauthorship Networks through Structural Cohesion, Holes, Hierarchy, and Percolating Clusters." *Journal of the Association for Information Science and Technology* 65: 1639-61.
- Ghosh, J., A. Kshitij, and S. Kadyan. 2015. "Functional Information Characteristics of Large-scale Research Collaboration: Network Measures and Implications." *Scientometrics* 102:1207-39.
- Giddens, A. 1979. *Central Problems in Social Theory: Action, Structure, and Contradiction in Social Analysis*. Berkeley: University of California Press.
- Gulati, R. and M. Gargiulo. 1999. "Where Do Interorganizational Networks Come from?" *American Journal of Sociology* 104:1439-93.
- Kang, I. -S., S. -H. Na, S. Lee, H. Jung, P. Kim, W. -K. Sung, and J. -H. Lee. 2009. "On Co-authorship for Author Disambiguation." *Information Processing and Management* 45:84-97.
- Kotha, R., G. George, and K. Srikanth. 2013. "Bridging the Mutual Knowledge Gap: Coordination and the Commercialization of University Science." *Academy of Management Journal* 56:498-524.

- Kshitij, A., J. Ghosh, and B. M. Gupta. 2015. "Embedded Information Structures and Functions of Coauthorship Networks: Evidence from Cancer Research Collaboration in India." *Scientometrics* 102:285-306.
- Liebersohn, S. and F. B. Lynn. 2002. "Barking up the Wrong Branch: Scientific Alternatives to the Current Model of Sociological Science." *Annual Review of Sociology* 28:1-19.
- Malone, T. W. and K. Crowston. 1994. "The Interdisciplinary Study of Coordination." *ACM Computing Surveys* 26:87-119.
- Maximon, L. C. 2003. "The Dilogarithm Function for Complex Argument." *Proceedings of the Royal Society A* 459:2807-19.
- McFadyen, M. A. and A. A. Cannella. 2004. "Social Capital and Knowledge Creation: Diminishing Returns of the Number and Strength of Exchange." *The Academy of Management Journal* 47:735-46.
- Merton, R. K. 1957. "Priorities in Scientific Discovery: A Chapter in the Sociology of Science." *American Sociological Review* 22:635-59.
- Milojević, S. 2013. "Accuracy of Simple, Initials-based Methods for Author Name Disambiguation." *Journal of Informetrics* 7:767-73.
- Moody, J. 2004. "The Structure of Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999." *American Sociological Review* 69: 213-38.
- Moody, J. and D. R. White. 2003. "Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups." *American Sociological Review* 68:103-27.
- Mooney, C. Z. and R. D. Duval. 1993. *Bootstrapping: A Nonparametric Approach to Statistical Inference*. Newbury Park, CA: Sage.
- Moore, C. and M. E. J. Newman. 2000. "Epidemics and Percolation in Small-world Networks." *Physical Review E* 61:5678-82.
- Newman, M. E. J. 2001. "The Structure of Scientific Collaboration Networks." *Proceedings of the National Academy of Sciences* 98:404-9.
- Newman, M. E. J. 2003. "The Structure and Function of Complex Networks." *SIAM Review* 45:167-256.
- Newman, M. E. J., S. H. Strogatz, and D. J. Watts. 2001. "Random Graphs with Arbitrary Degree Distributions and Their Applications." *Physical Review E* 64:1-7.
- Newman, M. E. J. and D. J. Watts. 1999. "Scaling and Percolation in the Small-world Network Model." *Physical Review E* 60:7332-42.
- Paruchuri, S. 2010. "Intraorganizational Networks, Interorganizational Networks, and the Impact of Central Inventors: A Longitudinal Study of Pharmaceutical Firms." *Organization Science* 21:63-80.
- Powell, W. W., D. R. White, K. W. Koput, and J. Owen-Smith. 2005. "Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences." *American Journal of Sociology* 110:1132-205.

- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. *Numerical Recipes in C. The Art of Scientific Computing*. New York: Cambridge University Press.
- Provan, K. G., J. E. Beagles, and S. J. Leischow. 2011. "Network Formation, Governance, and Evolution in Public Health: The North American Quitline Consortium Case." *Health Care Management Review* 36:315-26.
- Reagans, R. and B. McEvily. 2003. "Network Structure and Knowledge Transfer: The Effects of Cohesion and Range." *Administrative Science Quarterly* 48: 240-67.
- Shrum, W., J. Genuth, and I. Chompalov. 2007. *Structures of Scientific Collaboration*. Cambridge, MA: MIT Press.
- Simon, H. A. 1955. "On a Class of Skew Distribution Functions." *Biometrika* 42: 425-40.
- Uzzi, B. and J. Spiro. 2005. "Collaboration and Creativity: The Small World Problem." *American Journal of Sociology* 111:447-504.
- Walker, G., B. Kogut, and W. Shan. 1997. "Social Capital, Structural Holes and the Formation of an Industry Network." *Organization Science* 8:109-25.
- Watts, D. J. 1999. "Networks, Dynamics, and the Small-world Phenomenon." *American Journal of Sociology* 105:493-527.

Author Biographies

Jaideep Ghosh is an associate professor of Decision Sciences, Operations Management, and Management Information Systems in the School of Management and Entrepreneurship at Shiv Nadar University, India. He is also a Ramanujan Fellow of the Department of Science and Technology, Government of India.

Avinash Kshitij is a scientist at CSIR–National Institute of Science, Technology and Development Studies working on research policy in R&D analysis and growth. He is currently pursuing doctoral studies in the Centre for Studies in Science Policy, School of Social Sciences at Jawaharlal Nehru University, India.