Social Networks 33 (2011) 20–30

Contents lists available at ScienceDirect

Social Networks

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

The local and global structure of knowledge production in an emergent research field: An exponential random graph analysis

Neha Gondal

Department of Sociology, Rutgers University, 26 Nichol Avenue, New Brunswick, NJ 08901, United States

ARTICLE INFO

Keywords:
Knowledge networks
Emergent literature
Exponential random graph models
Bipartite networks

ABSTRACT

Previous research has characterized knowledge networks by diffuse connectivity and/or clusters and the *absence* of centrality. In contrast, *exponential random graph* models used in this article demonstrate that the uncertainty and centralized influence typical of an emerging area of research leads to the creation of a densely interconnecting *core* that acts to cohere the network. Moreover, eclecticism and innovativeness, also characteristic of a developing area, lead to a diffusely connected structure. The data, comprising 2200 authors and 76 papers have been manually coded from articles on the feminization of the labor force in Asia.

1. Introduction

Previous scholarship on knowledge networks has mostly been concerned with the discovery and interpretation of patterns at the level of the complete network. This aerial view has demonstrated that the structure of knowledge networks can be described in terms of diverse patterns, including clustering (Kaplan, 1965; Small and Griffith, 1974; Hill and Carley, 1999; Daipha, 2001), fragmentation (Carolan, 2008), and structural cohesiveness (Moody, 2004). Researchers have interpreted these patterns to signify the specialty, group, or organizational structure of disciplines, changing scientific paradigms, academic stratification, and the diffusion of ideas. While systematic analyses have allowed us to ascertain these macro-level properties, the microprocesses responsible for their emergence, such as preferential attachment, closure, brokerage, and own-group preference, have largely been the subject of informed speculation (see, for example, Daipha, 2001, p. 86–87; Moody and Light, 2006, p. 83). This has created a gap in the literature such that inferences about the state(s) of research disciplines are mostly based upon the large-scale structural features of knowledge networks, while little in is known about the micromechanisms generating those patterns and their implications. Using recent advances in *exponential random graph modeling* (ERGM), I address this gap by linking the macro-level properties of a network of citations to the microprocesses that concatenate to produce those structural features, and I offer a substantive interpretation of the findings.

The data for this study are derived from manually coding the bibliographies of 76 articles in an emergent field of literature (published between 1997 and 2007) pertaining to the rise in the employment of women in export production zones (EPZs) in South and East Asia. The data are recorded in two-mode format from citing papers to cited authors resulting in a network comprising 2200 authors and 76 papers.[1] Until recently, research based on two-mode data has relied upon analyzing their one-mode projections (see, Breiger, 1974). While this technique is frequently valuable, the two- to one-mode conversion results in the loss of considerable information (Latapy et al., 2008). Based on recent advances in the ERGM analysis of bipartite networks (Wang et al., 2009), I analyze the citation data in their original two-mode format.

Following current research, I examine three macro-level properties of the citation network. That is, I investigate the extent to which it manifests a star or core–periphery structure, structural cohesiveness, and/or the properties of a small-world. I next investigate the microprocesses that are most likely to have generated the macro-structure. Briefly, at the macro-level, I find that the data reveal a core of highly cited authors that contributes significantly towards the connectivity of the network. In addition, I also find evidence in support of structural cohesion and small-world patterns. ERGM reveals that microprocesses jointly interpretable as *preferential attachment*, *homophily*, and *differentiation* are responsible for generating the network. In emphasizing preferential attachment and corresponding tendencies towards centralization of the network, the findings from this study diverge

[1] It is important to note that while I cannot guarantee that every paper written in this area is included in the dataset, I have made every effort to meet that target. Thus, the relative smallness of the dataset in comparison to some other studies is largely attributable to the emergent nature of the field.

E-mail address: ngondal@sociology.rutgers.edu.

0378-8733/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.socnet.2010.09.001

from those of previous research that has found fragmentation or cohesion to exist alongside the *absence* of centrality (see, Daipha, 2001; Moody, 2004; Carolan, 2008). I argue that the social forces structuring a literature in its early years are somewhat different from those affecting an established field. Citing a small stock of influential authors, a practice according the network with a heavy center of gravity, is more crucial for authors hoping to publish in an emergent field. Simultaneously, consistent with previous arguments (Gilbert, 1977; Paisley, 1990; Hill and Carley, 1999), the findings on differentiation suggest that scholars publishing in an emergent area, the boundaries of which are as yet uncertain, are likely to be both innovative and influenced by eclectic sources.

2. Global, dyadic, and local properties of knowledge networks

2.1. Data construction

Research focused upon the structure of knowledge production frequently relies on network data. The nodes in the network may be researchers, documents, concepts, or organizations. The edges connecting these nodes correspondingly are collaborative authorship (Babchuk et al., 1999; Moody, 2004; Goyal et al., 2006), social and intellectual contacts between scientists (Lievrouw et al., 1987), co-occurrence of references in the bibliographies of other documents or co-citation (Small and Griffith, 1974; Moody and Light, 2006), shared citations of the same other documents or authors also known as bibliographic coupling (Kessler, 1963), shared membership in organizations (Barnett and Danowski, 1992; Cappell and Guterbock, 1992; Daipha, 2001), or conceptual similarity between documents (Small, 1978; Lievrouw et al., 1987; Hill and Carley, 1999). The analysis of such networks constructed from citation indices, organizational memberships, and authorships is largely conducted at two levels. At the dyadic level, researchers have been concerned with the meaning attributed to the edges interlinking the nodes. At the 'global' or 'macro' level, researchers analyze the topological properties of the network as a whole providing a bird's-eye description of the research field. There is yet another level – the 'local' or 'micro' level – involving more than one tie but significantly less than the complete network which remains relatively underanalyzed in the literature. I briefly describe these three levels next.

2.2. Dyadic properties

Considerable debate surrounds the meaning of the ties linking any two nodes. That is, what exactly is it that flows through the pipes linking individuals, organizations, or documents? Citations, for instance, have variously been interpreted as indicative of social influence, reputation, political motivations, self-aggrandizement, property rights, symbol-making activity, intellectual heritage, 'window-dressing', and intellectual debt relations between authors (Kaplan, 1965; de Solla Price, 1965; Small and Griffith, 1974; Gilbert, 1977; Small, 1978; Cozzens, 1989; Leydesdorff and Amsterdamska, 1990; Kostoff, 1998; Makino, 1998; Fujigaki, 1998; Kostoff, 1998; Collins, 2001; Jarneving, 2005). Similarly, co-specialization has been argued to be represented by organizational memberships (Cappell and Guterbock, 1992; Daipha, 2001) on the one hand and declared areas of interest (Ennis, 1992) on the other.

Meanings of ties are highly contested areas of research because they have implications for the meanings attributed to the network under study. Assuming citations symbolize ideational connections between documents implies that the network represents the intellectual structure of a research field. On the other hand, a network made up of citations viewed as links between authors is thought to reflect the social organization of scientific communities (Leydesdorff and Amsterdamska, 1990).

2.3. Global properties

A significant proportion of the research investigating knowledge networks is dedicated to the detection and analysis of macro-level structure. Using clustering algorithms such as the level of the strength of co-citation between documents, Small and Griffith (1974), for instance, investigate the specialty structure of scientific disciplines. Others have likewise used alternative techniques such as hierarchical clustering and multi-dimensional scaling to decipher the intellectual and social structure of a discipline (Daipha, 2001), similarity indices to map the position of one discipline relative the position of others (Moody and Light, 2006), and factor analysis to discover sub-specialties within a discipline (Van Den Besselaar, 2001). Another variant of global-level analysis is concerned with analyzing the importance of particular documents, concepts, or authors on the basis of their position in the complete network (Hoffman and Holbrook, 1993; Otte and Rousseau, 2002; Evans, 2005).

Substantively, this research delineates a number of hypotheses linking global network properties to the research structure of disciplines. Moody (2004) discusses three distinct models to describe a co-authorship network in sociology. A *star* network structure exists when a small number of authors are on the receiving end of a large number of credits. Such 'stars' are expected to exert disproportionate influence in the research community as their ideas diffuse more rapidly through the network. A *small-world* network structure is made up of locally dense clusters (such that one's connections are also tied to one another) that are nevertheless linked by few interconnecting paths of relatively short length. This structure is likely to be associated with a theoretically fractured underlying knowledge space. A *structurally cohesive* network, with evenly distributed ties and numerous paths connecting any two nodes, would be characteristic of a space with cross-topic collaborations and permeable boundaries. Carolan (2008) discusses three comparable models applicable to research in the field of education: the 'plural-world' model made up of sub-disciplinary fragments that impede consensus formation; Moody's structurally cohesive model with substantial interconnectivity between sub-disciplines; and the small-world model. Moody finds that while the network contains influential people, as connectivity is not contingent upon them, the star model is an inappropriate description of the data. Likewise, he rejects the fit of the small-world model, arguing that "high levels of intergroup contact, weak internal structure, and strong overall connectivity point towards generalized cohesion" (p. 231). Daipha (2001) similarly argues for the lack of centrality in sociology. But instead of cohesion, she argues in favor of an environment marked by pluralism and segregation. In between these two sets of findings, Carolan finds the education research field to be a mix of small-world and structural cohesion features.

These findings, however, largely pertain to mature areas of research. An investigation of emergent fields paints a rather different structural story. According to Crane (1988[1972], p. 40, 54), during the early stages of an area of research, a few productive scholars become crucial for the development and overall connectivity of different parts of the field. Substantiating this argument, Moody and Light (2006) find that the discipline of sociology to be comparatively more centralized in 1970 than in 1980. Hill and Carley (1999) similarly find that the early years of a discipline are associated with greater (conceptually measured) structural centrality. Evans (2005) argues that this higher tendency toward centralization in the early years can also be explained on the basis of a need for legitimation. In the early years, Evans argues, when

gatekeepers of publication are not sure of how to evaluate a text, they rely on the prestige of a small number of key predecessors. Analogously, 'task uncertainty' or the "extent to which scientific production is (not) routinized and predictable," can be expected to be higher when problems and concepts are not clearly defined (Fuchs, 1992:82, 190) leading scholars to be reliant on exemplars. As distinct from previous arguments about the lack of centrality, these studies suggest that in the early years, an area of research characterized by low consensus, high uncertainty, and centralized influence may lead to the creation of a star-like structure of knowledge space.

2.4. The missing link: local properties

While the existing literature makes important contributions to the global structure of knowledge networks as well as to dyadic-level meanings, analyses that account for the frequency with which network structures involving more than one tie but significantly fewer ties than the complete network occur are largely missing. Such micro-structures are network configurations composed of a few ties, such as, triads or four-cycles. For example, in a network composed of authors citing documents, a four-cycle could be composed as follows: Author A cites documents T1 and T2, both of which are also cited by author B. Larger configurations of this type could involve nested four-cycles involving the same two documents being cited by multiple other authors or the same authors citing many of the same other documents.

These configurations, however, are not merely structures, but indicators of *social processes* that influence actors' involvement in knowledge networks. That is, similar to interpretations of structural properties at the global (for example, fragmentation implies sub-specialties) and dyadic (meanings of citations) levels, micro-network configurations have associated (context-specific) substantive meanings. Consequently, they are also known as *localized attachment logics* (LALs) in ERGM literature. For instance, a star structure at the micro-level may plausibly be thought to reflect preferential attachment (a social process), whereby nodes entering a new field attach themselves to popular nodes either because of the latter's inescapable influence (Crane, 1988) or for purposes of legitimation (Evans, 2005).

The statistical modeling of networks within the ERGM framework helps to link such local configuration with global structural outcomes. This is achieved by examining the degree to which tendencies towards or against the formation of micro-configurations contribute towards the emergence of the overall network structure. A model obtained by applying ERGM may reveal, for instance, that a general propensity in a literature towards the formation of four-cycles of the type described above leads to structural cohesion with many paths linking authors at the global level. Similarly, a core–periphery structure may be obtained from the concatenation of many local star structures (Snijders et al., 2006). While several authors have speculated upon such local processes as the formation of 'exemplars' (Gilbert, 1977; Evans, 2005) to explain a core–periphery structure; competition, own-group preference, and ideological schisms (Small and Griffith, 1974; Egghe and Rousseau, 2004; Daipha, 2001; Moody, 2004; Moody and Light, 2006) to explain structural pluralism; and cross-topic collaboration and brokerage (Moody, 2004) to explain structural cohesion at the global level, little has been done in the way of examining this micro–macro-link in a statistically rigorous way. Using recent advances in ERGM, I explicitly investigate the LALs responsible for the creation of the citation network of an emergent field of literature.

Stated simply, the application of ERGM proceeds by examining the global properties of a network; speculating upon the micro-structures generative of that structure; and finally specifying a model composed of the configurations thought to be suitable. Interpretation of the models so obtained is the final step. The following analysis is organized in accordance with this schema. First, based on current literature, I test if the global properties of the network fit a star, small-world, or structurally cohesive pattern. Next, on the basis of that evidence, I use ERGM to investigate the underlying microprocesses leading to the emergence of those global properties.

3. The data: the emergent literature on EPZs

Data for this paper are derived from the bibliographies of 76 articles. The coding was done in two-mode format from text to author. This was achieved by tabulating the reference lists of the 76 papers in digital forms recognizable by *Pajek* (Batagelj and Mrvar, 1996), a network visualization software package and *BPNET* (Wang et al., 2004), a network statistical modeling software package. An edge in this network connects a paper to an author if that author's name is found in the bibliography of that paper. There is thus a duality between authors and texts. Authors are indirectly connected to each other by virtue of being cited by the same papers, while the papers are similarly interconnected by the authors they cite in common. The data in their two-mode format can be converted to unipartite form by using either one of Breiger's (1974) algorithms to derive one-mode projections from two-mode matrices. However, rather than converting the data to one-mode as has been done by others (Daipha, 2001; Carolan, 2008), consistent with recent advances in the analysis of bipartite network data (Latapy et al., 2008; Wang et al., 2009), they were analyzed in their original two-mode form, preserving the duality between authors and documents.

The articles, written in the English language, pertain to the employment of women in export processing zones (EPZs) in Asia. The four nations of Bangladesh, Sri Lanka, China, and India began implementing reforms aimed at the privatization and liberalization of the economy in 1975, 1977, 1978, and 1991, respectively. As a part of the liberalization process, international companies producing mostly consumer goods were given incentives by domestic governments to set up factories in demarcated regions called EPZs. It has been demonstrated that factories in EPZs frequently employ a larger proportion of women, a phenomenon termed as the 'feminization' of the labor market (Hale, 1996; Standing, 1999; Kabeer, 2004; Datta, 2005). Feminist literature examining the consequences of structural adjustment policies has tried to assess their gender-differentiated implications, explicitly questioning the alleged gender neutrality of market-oriented policies. This literature stands in stark contrast to traditional economic models that explain sex-based discrimination as 'residual variance' rather than viewing markets as structurally gendered (Elson, 1999). Feminist writers have offered a number of critiques to this traditional rational choice model of agency, arguing that this type of employment is severely exploitative (Safa, 1981), places an unequal burden on women (Berik et al., 2007), and does not automatically translate into empowerment in the long run (Tzannatos, 1999; Ghosh, 2002). Despite the criticism registered against economistic measures of empowerment, it is nevertheless argued that women in this space should be viewed as agents of change renegotiating gender and work relations in their everyday actions.

These papers, whose publication dates range from 1997 to 2007, were identified on the basis of keyword searches including, but not limited to, globalization, EPZ, garment, Asia, feminization, India, China, Sri Lanka, and Bangladesh, on search engines including Google Scholar, EBSCO Host, University Libraries, Searchlight, and JSTOR. The search was conducted until no new papers related to the subject matter of the feminization of the labor force in Asia could be found. The papers so identified were then read to ensure that

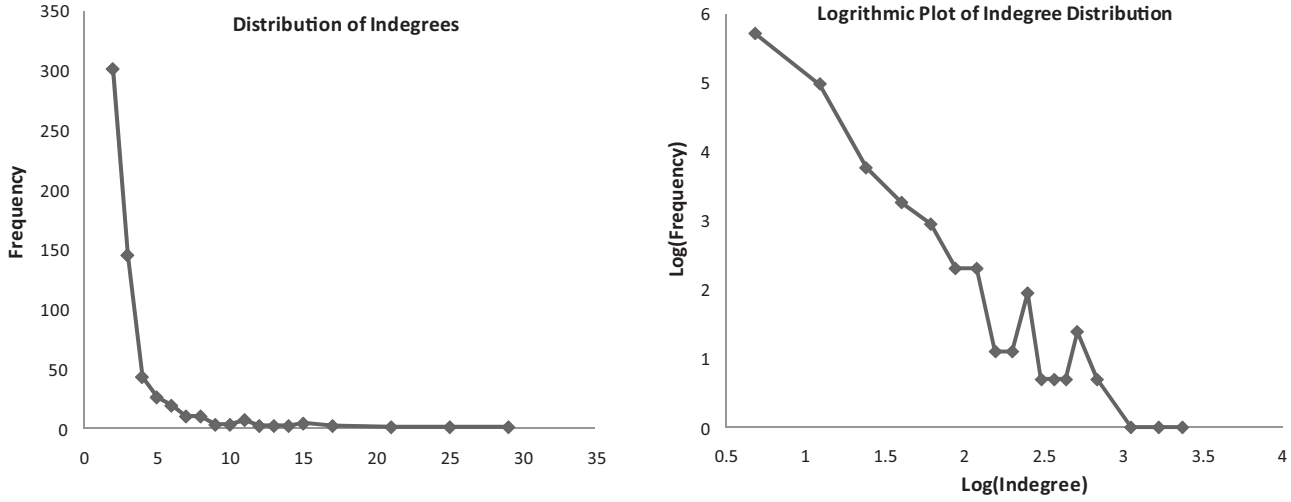


Fig. 1. Distribution of indegrees.

they were indeed related to the subject matter of EPZs and feminization. The papers were subsequently classified on the basis of the specific region within Asia upon which they focused. This process resulted in the delineation of six regions (number of papers): Asia (8), Bangladesh (21), China (12), India (16), South Asia (4), and Sri Lanka (15).[2]

The resulting dataset comprises 2,220 authors and 76 papers. As the data were coded manually, rather than from citation indices, the effect of errors from spelling or the inclusion of middle names is likely to be lower. Moreover, after the coding, the data were reorganized by cited author and checked for the occurrence of such errors. Of the total authors, 1658 are cited only once. They fail to link any papers together and hence may be regarded as unimportant for the purposes of general pattern detection. The exclusion of all authors cited once reduces the original set of 2220 authors to 562, resulting in the 'multi-cited' network.

4. Analysis

4.1. Global properties of the network

4.1.1. Star structure

Fig. 1 depicts the indegree distribution of the network. As can be seen from the left panel, the distribution has a long tail, indicating that whereas most authors are cited a few times, some authors or 'stars' are cited repeatedly in the network. The same information is reflected on a logarithmic scale in the right panel. The negative slope of the natural log[3] of frequency to the natural log of indegree is a property of 'scale-free' or 'power-law' networks (Barabási and Albert, 1999).[4] This star structure of the data is also evident from the fact that the network of papers is robust to the removal of lower indegree authors from the network. To begin with, the network is a single component (Wasserman and Faust, 1994), such that every paper is reachable from every other, via their mutual connections to cited authors and vice versa. Not only is the network a single component, but it continues to remain so when authors receiving fewer citations are eliminated from the network. In fact, the network remains completely connected until the sixth iteration of the sequential exclusion of the lowest indegree authors. Once all authors receiving *fewer* than seven citations are removed, two papers break-off, becoming isolates in the network. Further iterations reveal a very slow rate of paper isolation. The ninth iteration, eliminating all authors receiving fewer than 10 citations, results in the loss of 10 papers from the entire network leaving 25 authors, each on the receiving end of 10 or more citations that interconnect 66 of the 46 papers. A mere 1.12% of the original set of 2220 cited authors keeps 87% of the papers interconnected.[5]

4.1.2. Cohesiveness

In addition to the robustness of the network to the removal of less popular authors, the multi-cited network is also a bicomponent. Moody and White (2003) argue that bicomponents (components in which there are at least two node-independent paths connecting every pair of actors) are more robust than components because the network is less sensitive to becoming disconnected even if some nodes are eliminated. Moody argues this to be a feature of the structural cohesion model. In fact, reversing the node removal procedure described above, the exclusion from the network of the set of 25 authors receiving 10 or *more* citations, leaves the entire set of 76 papers interconnected by virtue of citing the less frequently cited authors. Continued iterations reveal that all 76 papers remain connected to the network until the final iteration of removing all authors receiving more than two citations. At this point, three papers drop out of the network, leaving 73 of the original 76 papers indirectly connected to one another. This is quite a remarkable level of connectedness. The continued removal of highly cited authors could ultimately have resulted in a fragmentation of the network into many components. That this did not occur implies that papers are not attached to the network solely by virtue of citing the more popular authors. Less prominent authors in the network also serve to interconnect the papers, lending support to the structural cohesion model. However, as the core of merely 25 authors in the center of the network maintains the interconnectiv-

[2] To the extent possible, the papers were classified into the four nation specifications. However, some papers did not fit neatly into this classification, yet they addressed issues relevant to feminization in Asia. Those papers were classified as pertaining to Asia in general, or more specifically to South Asia.

[3] An OLS regression fitting the log of frequency on the log of the outdegree results in a negative slope of −0.45 with an R-square of 0.92.

[4] In social network literature, distributions with long tails that do not exhibit a representative observation are called scale-free. It should be noted however, that as the data do not display a strict power law structure (the slope of the curve is quite low), this model may not be sufficient to capture the structural properties of the data.

[5] This star structure of the data is also confirmed by the Gini coefficient (0.408) of the continuous coreness scores of the author-by-author valued matrix (see Borgatti and Everett, 2000).

Table 1
Descriptive statistics of the network and its constituent regional components.

Network	Number of citing papers	Number of cited authors	Density	Average distance
Complete network	76	2220	2.13%	4.40
Multi-cited network	76	562	4.54%	3.67
Asia	8	119	16.49%	3.39
Bangladesh	21	200	13.04%	3.24
China	12	178	17.79%	3.09
India	16	219	10.27%	3.60
South Asia	4	80	30.31%	3.12
Sri Lanka	15	170	15.68%	3.29

Table 2
The correlations between the regional profiles of citations by authors ($N = 588$).

	Bangladesh	China	India	Sri Lanka
Bangladesh	–	−0.080	−0.089*	−0.130**
China			−0.117**	−0.196**
India				−0.172**

* Significant at the 0.05 level (two-tailed).
** Significant at the 0.01 level (two-tailed).

ity of most of the papers, it correspondingly has a lower diameter and average distance than the periphery. The diameter of the connected portion of the core is 6 and the average distance is 2.71. The corresponding figures are 7 and 3.95 for the non-core (authors with an indegree of 9 or less and the papers connected to them) and rise to 14 and 5.85 for authors receiving two citations each.

4.1.3. Regions as small-worlds

Table 1 shows some descriptive features of the whole network as well its constituent regional parts. In this network, the density is lowest when the entire network is considered but doubles for the multi-cited network. The density of all the regional components is considerably higher than that of the complete and the multi-cited networks. The average distance is similarly smaller for all the regional components. The higher density and shorter distances within the regional components clearly suggest internal clustering.

Regional patterns in citations become all the more evident on an examination of each author's citation profile cross-tabulated by region. Faulkner (1983) discusses the 'typecasting' of music composers by genre of Hollywood films. In a similar manner, certain authors are more likely to be cited within literature originating in a specific region. In order to test this at the global level, each author's citations were tabulated by the region of the sending paper. The overall patterns in the author by region citations indicate significant, but low, negative correlations between the four countries (with the exception of Bangladesh and China) (Table 2). This pattern demonstrates that, in general, there is a tendency for low overlap between regional citation profiles, corroborating earlier evidence in support of region-based cohesion.

At a more local level, the typecasting of authors is also evident from variations in their region-based citation profiles. Authors with high variance are more likely to be cited in only one region to the exclusion of other regions whereas authors with low variance are more likely to be on the receiving end of citations from literature on all regions. Kabeer, the most cited author in this network, receives most of her citations (63%) from Bangladeshi papers, considerably skewing her citation profile. Similarly, Zohir, Kibria, Paul-Majumdar and Jahan receive most of their citations from Bangladeshi papers, Jacka, Lee and Pun from Chinese papers, Banerjee and Jhabvala from Indian papers, and Jayaweera, De Alwis, Lynch and Hewamanne from Sri Lankan papers.

Despite this evidence in favor of regional concentration, the clusters do not amount exactly to small-worlds. In contrast to authors with high region-based variance, others, such as Pearson, Elson, Beneria, Ong, Wolf, and Joekes, mostly writing about feminization in other geographical contexts, have much more even citation profiles, making them more central to the genre as a whole. These and other authors create multiple passageways deeply interlinking regional clusters, much like Crane's (1988) community of core researchers – the invisible college – that binds disparate groups within a research area. Table 3 shows the absolute number of linkages connecting the four country clusters at distinct levels of author indegree. For instance, accounting for authors who are cited only twice (indegree of 2), Bangladesh papers are mutually connected to China papers via two citations, to India via eighteen, and to Sri Lanka via five. It is apparent from the table that the clusters are strongly interlinked. But it is also clear that periphery authors contribute significantly less towards fusing regional clusters than do more popular ones. In the aggregate, even though there are 537 authors with an indegree of 9 or less (excluding singly cited authors), they are a source of far *fewer* interconnections between regions than the 25 authors who have an indegree of 10 or more.

These properties of the network can be confirmed from Fig. 2. Fig. 2, drawn using the Fruchterman–Reingold[6] algorithm depicts the complete network as well as the multi-cited component. The left panel shows that the skewed degree distribution manifests in a core–periphery structure. The author nodes towards the center of the figure receive most of the citations, while the nodes towards the periphery and semi-periphery are the source of connections between fewer papers. The figure also shows that the entire network is single cohesive component. The comparatively higher region-specific interactions can be seen from the right panel of this figure. It shows the Bangladesh, China and Sri Lanka papers to be clustered together in the northwest, southeast and southwest of the figure, respectively. Asia, India and South Asia, although not as tightly visually clustered as the other three, are nonetheless still found towards the middle and northeast of the picture. The spatial centrality of the Asia and South Asia components is quite remarkable because it indicates that their relationally derived position is supported by their lack of nation-specific, attribute-based categorization.

To summarize, in line with the findings on emergent areas of research, the data reveal a core of highly cited authors that contributes significantly towards overall as well as regional paper interconnectivity and considerably lowers the average distance and diameter of the network. While the stars may not hold the network together in a strictly structural sense, they are nonetheless very important for *cohering* it, and in that latter sense, research on EPZs in Asia appears to be highly dependent upon them. In addition, consistent with previous research on mature areas of research, the deep interconnectivity of the overall network well into the periphery and higher intra-regional connections point towards structural cohesion and pluralistic small-worlds, respectively. In the next section, I use exponential random graph modeling to investigate the micro-configurations and associated localized attachment logics that are most likely to have generated these global properties.

4.2. Exponential random graph modeling

4.2.1. Modeling framework

The ERGM framework assumes a stochastic environment in which ties/edges serve as random variables. Modeling begins with dependence assumptions such as Bernoulli, Markov or realization dependence (Pattison and Robins, 2002; Snijders et al., 2006;

[6] The Fruchterman–Reingold algorithm is a force-directed algorithm. Essentially, it draws a layout by pulling adjacent vertices together and pushing non-adjacent ones apart until a local minimum force is reached.

Table 3
Regional group interlinkages by author indegree.

	China				India				Sri Lanka			
Author Indegree	2	<6	<9	>10	2	<6	<9	>10	2	<6	<9	>10
Bangladesh	3	49	89	170	18	96	128	329	5	41	89	110
China					12	57	66	96	4	11	28	65
India									7	41	55	76
Cumulative authors	283	514	537	25	283	514	537	25	283	514	537	25

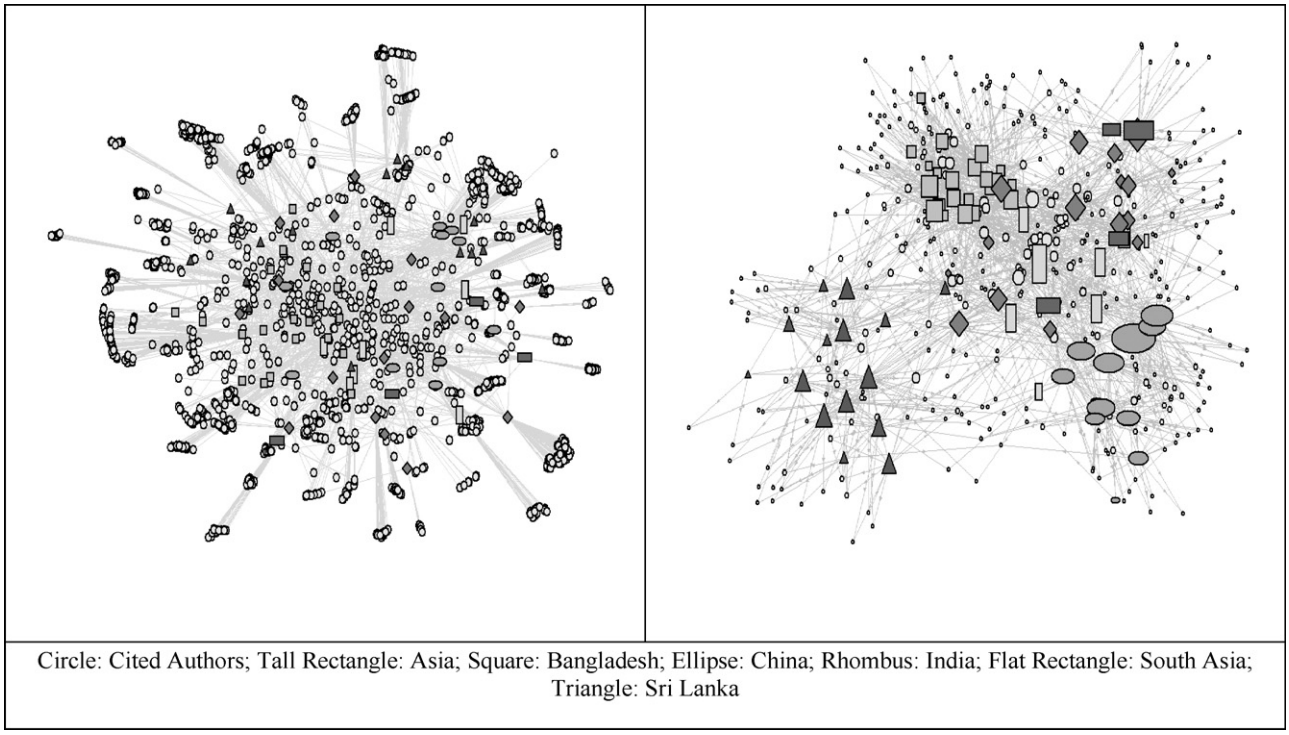


Fig. 2. Network visualizations.

Robins et al., 2007a,b) based on the Hammersley–Clifford theorem (Besag, 1974). Dependence assumptions are hypotheses about how ties in the network are expected to be interconnected. For example, a Bernoulli model postulates that all ties are independent of one another and the occurrence of higher order configurations in the data, such as triads, are attributable to chance. In such a case, it would be safe to assume that the aggregate pattern is simply an agglomeration of individual ties, and larger micro-configurations are unimportant. In addition to endogenous network configurations, a model may also include exogenous attribute variables to capture homophily/heterophily effects. The Hammersley–Clifford theorem helps to impose homogeneity across subsets of ties that are isomorphic to each other, making the model estimable. Until recently, most estimation was done using either Maximum Likelihood Estimation (MLE) (Holland and Leinhardt, 1981) or Maximum Pseudolikelihood Estimation (MPLE) (Strauss and Ikeda, 1990). More recently, Monte Carlo Markov Chain Maximum Likelihood Estimation (MCMCMLE) (Snijders, 2002; Handcock, 2003; Hunter, 2007) has become the preferred fitting algorithm. While most literature on ERGM pertains to one-mode data (Pattison and Robins, 2002; Snijders, 2002; Robins et al., 2009), foundational work by Skvoretz and Faust (1999), and Agneessens et al. (2004) has recently been extended and implemented by Wang et al. (2009) to include two-mode data. The latter modeling strategy has been applied in this article.[7]

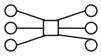
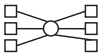
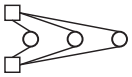
4.2.2. Model

The global properties of the network enumerated above are modeled here using the MCMCMLE algorithm. Several models with various combinations of network configurations were tested on the data. The best results of fitting a model to the multi-cited network are shown in Table 4, while a few other model fits and goodness of fit statistics can be found in the appendix. Ties represented in the table connect circles (authors) and squares (papers), with regions represented in distinct shapes. The model shown here fits endogenous *edge*, *alternating-paper-k-star*, *alternating-author-k-star*, *alternating-paper-k-two-path* and exogenous *country-wise homophily* locally specified parameters to the data.[8] This model performs much better than the Bernoulli model which assumes ties to be independent of each other, demonstrating that the other net-

[7] A comprehensive review of the literature in this area is not possible here. Interested readers are encouraged to consult Robins et al. (2005) for a good introduction to the material.

[8] In the model fit, Asia, South Asia and India have not been presented because they did not exhibit converging estimates of homophily.

Table 4
ERGM fit of the multi-cited network.

Configuration	Statistic	Estimate	Standard error	*T*-ratio
	Edge	−9.442*	0.835	−0.025
	Alternating-paper-k-star	3.158*	0.420	−0.030
	Alternating-author-k-star	0.3617*	0.122	−0.025
	Alternating-paper-k-two-paths	−0.375*	0.043	0.001
	Homophily Bangladesh	0.215*	0.024	−0.045
	Homophily China	0.362*	0.031	−0.018
	Homophily Sri Lanka	0.289*	0.029	0.019

* Significant estimate.

work configurations listed here are important determinants of the aggregate structure.

The first column of Table 4 graphically depicts the network configurations employed in the model. The first edge parameter simply accounts for the number of ties in the model. The alternating-author-k-star parameter captures the effect of multiple papers citing the same author, while the alternating-paper-k-star conversely captures the effect of multiple citations by the same paper. Kabeer, for instance, the most frequently cited author, receives citations from 29 of the 76 papers making her an 'author 29-star'. On the other extreme, all authors on the receiving end of two citations are author 2-stars. The alternating-paper-k-two-path explains the tendency for two papers to cite the same multiple other authors, or the strength of bibliographic coupling (from the viewpoint of the papers). For example, the citation of five of the same authors by two papers would create a paper 5-two-path between them. In fact, the data show that 95 pairs of papers share exactly five author citations. Finally, the homophily parameters capture the tendency for papers within the same regions to cite the same authors. A negative parameter estimate indicates low probability, while a positive estimate indicates high probability. Parameter estimates are considered significant if the reported standard error is less than half the corresponding parameter estimate.

The interpretation of the negative edge parameter is similar to the intercept in standard regression analysis. The high positive paper k-star estimate captures the tendency of papers to cite multiple authors—hardly a surprising finding from what we know about article bibliographies. The positive author k-star parameter estimate, on the other hand, signifies a tendency for authors to be on the receiving end of citations from multiple papers. Net of other structural effects, a positive k-star parameter alongside a negative edge parameter has been interpreted to reflect *preferential attachment* at the local level leading to a *core–periphery* structure at the global level (Robins et al., 2007b, 2009).

The negative paper k-two path parameter indicates that there is not a very strong tendency towards high bibliographic coupling, so two papers do not tend to cite many of the same authors. The positive k-star estimates and the negative k-two path estimate in conjunction imply that while papers tend to cite some of the same authors (preferential attachment towards authors in the core), this tendency is limited in its scope for closure. This implies that, despite or in addition to preferential attachment towards the core, citation profiles do not simply converge on a single set of star predecessors. They actually tend to vary from one paper to the next. This indicates some degree of *differentiation* built into individuals' selection of citations, and/or processes or norms endemic to the field that invite limitations on the convergence normally associated with preferential attachment. Finally, the positive homophily (McPherson et al., 2001) estimates suggest that papers of the same region tend to cite the same authors, leading to regional grouping at the global level.[9]

5. Discussion and conclusion

It is important to emphasize that this analysis pertains to a single area of literature, and consequently generalizability to other cases should be interpreted with caution. Nevertheless, if the analysis has external validity, what could we learn from it?

At the aggregate level, the descriptive network analysis demonstrates that the data exhibit features of all three paradigmatic models of knowledge production discussed in previous research. The set of 25 authors with high degree centrality are collectively a crucial source of interconnectivity, conforming to the star model. At the same time, the comparatively higher internal density and external negative correlations of the regional components are indicative of small-world like properties. Lastly, the continued connectedness of the papers despite the removal of high indegree authors, as well as it being a bicomponent, accords well with a structural cohesiveness argument.

Exponential random graph modeling reveals that three sets of structural configurations – the presence of star structures and regional homophily, and the absence of paper-two-paths – are responsible for inducing the global network structure. While the first two of these local configurations seem to be consistent with the global core–periphery and small-world structures respectively, the absence of two-paths appears to be at odds with a cohesive structure. Recall that a cohesive structure requires a relatively

[9] In addition to the multi-cited data, I also ran models on the complete dataset including single citations. It is important to bear in mind that on the basis of current technology, the time costs of fitting ERGM to very large datasets are prohibitively large. In fact, large fitting ERGM to very large datasets is often intractable. Initial runs on the complete dataset revealed that models similar to the one above (with higher star parameter values and lower two-path effects) were suitable for the complete dataset, as well.

even distribution of ties and many paths linking nodes together. The positive star parameters and regional homophily defy the first requirement, while the negative paper k-two-path parameter challenges the validity of the second. Clearly, there is a discrepancy between the structural properties of the network in its aggregate form and those operating at the micro-level. In what follows, I first offer substantive interpretations of the three local configurations generally, as well as in the context of emergent literature on EPZs. Next, I link the micro- and macro-level analyses in light of those substantive interpretations.

A tendency towards the formation of author-stars can be interpreted simply: scholars are more likely to cite reputable authors that are visibly also being by cited by other scholars, a process termed preferential attachment in network literature. Yet, preferential attachment need not be operative uniformly over the life-course of an area of research. In the early years, when a literature is marked by low consensus (Hill and Carley, 1999) and high task uncertainty (Fuchs, 1992), authors with lower prestige have a greater motivation to cite more reputable ones in order to legitimate their own work (Evans, 2005) thereby accelerating preferential attachment tendencies. This strategy serves the dual purpose of signaling as well as drawing substantive or ideational links. Furthermore, Evans (2007) argues that low consensus is associated with high rates of publication rejection, which creates yet another basis for uncertainty, further driving up the need for legitimation. Fuchs also argues that task uncertainty is expected to be higher in the event of disintegrating paradigms. Furthermore, Whitley (1984) points out that uncertainty is positively related to audience plurality. Given that literature on feminization in Asia is an emergent area of research, that it has arisen largely in response to changing state policies and in opposition to rooted economistic models of labor and globalization, and that it is considerably politically charged with appeal to diverse audiences, it should be characterized by high task uncertainty and consequently greater need for legitimation.

Another possible explanation for the emergence of star structures is that certain authors have been around longer. That is, preferential attachment effects could be due to time in the discipline rather than reputation. Without explicitly controlling for tenure, it is nearly impossible to distinguish between these two effects in a static analysis. Nevertheless, the difference between 'pure' reputation and temporal effects is likely to be much more fluid in the case of emergent literatures. This is because, as Crane (1988) argues, early researchers are also the innovators setting the research agenda well as training and recruiting students. These early developers are highly influential, both socially and intellectually. Consequently, as the evidence suggests here, most papers publishing in this area cite this small pool of pioneering scholars almost as a matter of necessity.

Positive propensities towards regional homophily configurations imply papers within the same regions are more likely to cite many of the same authors. On the one hand, this is indicative of differentiation away from the general subject matter and towards the specifics of the case. In this instance, authors aim to produce findings that speak not just to the subject matter of feminization of the labor force in general, but also to its regional particularities. This tendency to attend to specifics produces configurations of the regional homophily type. On the other hand, regional homophily may also obtain from the very processes that lead to preferential attachment to core authors. Successfully publishing in area-specific journals may require citing local luminaries, who may well also have considerable social and intellectual influence.

Lastly, the alternating-paper-k-two-path configuration measures the extent to which papers cite the same authors. ERGM reveals that a tendency *against* the formation of such configurations is likely to have produced the network. Substantively, this tendency against citation overlap, which I call differentiation, can be explained in multiple ways. One, given the paucity of time as a resource, and the continuously changing nature of research, one cannot possibly hope to read all pertinent material before producing new knowledge (Crane, 1988). The opportunity cost of time imposes decision-making about what to incorporate into one's work and what to ignore. Camic (1992) refers to this process as 'predecessor selection.' This locally instantiated deliberation over relevance leads to the emergence of diversity of citations as different scholars make distinctive choices over and above the core authors. Crane argues that this constraint is likely to be stronger during periods of rapid change (pp. 116–117), making for lower citation overlap in such situations. Paisley (1990) similarly argues that the effect of eclectic influences is likely to be amplified during the formative years of a field of literature.[10] Another possible explanation is that low overlap in citations signifies innovative thinking, some of which entails citing research lying outside the customary literature's realm, interpreted as novelty by Gilbert (1977), a preference for diversity by Powell et al. (2005) and competition by Hill and Carley (1999).[11] Finally, heterogeneity could be attributed to uncertainty regarding boundaries, which, as argued above, is to be expected of an emergent area.

Having linked local structures with social logics of attachment, we are now in a position to relate the formation of micro-configurations with the macro-level structural properties of the EPZ citation network. Local star structures and the associated preferential attachment logic suggest that at early stages when a literature is not yet established, referencing reputable others can perhaps be regarded as that much more crucial for situating one's work and convincing gatekeepers and evaluators. A corollary of this notion is that star structures indicate the considerable social and intellectual influence wielded by innovators in the field. In the aggregate, either form of 'preferential attachment' – authors' risk-avoidance strategies, or the relative concentration of innovation in the hands of a few cited experts – should create a core of oft-cited authors according the network with a center-heavy structure.

Second, periphery-mediated structural cohesion alongside a negative k-two-path parameter presents an intriguing paradox. Careful consideration of the structural properties of the data helps to reconcile this apparent contradiction. A positive k-two-path parameter alongside positive star parameters would likely result in (and consequently be obtained from modeling) a highly center-heavy, 'black hole' type network structure. Recall that a positive k-two-path parameter is linked with highly overlapping citation profiles and a positive star parameter with preferential attachment to the core. In conjunction these effects would result in a network structure where papers were citing very many of the same authors making for a highly dense large core (the authors everyone cites) and comparatively small periphery (the few authors that are cited infrequently). Following the same logic, the presence of positive homophily parameters would likely also create dense regional clusters that are nonetheless interconnected via the large core. But, as the actual network has a small core and a multi-layered and large periphery, the negative k-two-path parameter estimate con-

[10] I acknowledge that if the data are not exhaustive, the addition of papers may transform some of the single citations to multiple ones and so on. However, given the magnitude of singular citations, the marginal effect of adding more papers should not be very large. Moreover, the addition of more papers is also likely to increase the absolute number of singular citations.

[11] Admittedly some of these idiosyncratic citations are attributable to factors somewhat orthogonal to the idea of agency as defined here. For instance, Kaplan (1965:182) suggests that authors systematically cite the work of colleagues and superiors, indicative of professionalization/training. One way to mitigate this issue would be to exclude all idiosyncratic citations whose authors appear in the acknowledgements, or are known to have close social ties with the citing author.

sequently represents a tendency towards citing non-core authors, indicative of scholars' innovative ideas in a competitive environment. Yet, as a considerable number of the non-core authors are cited multiply, those citation choices are not entirely idiosyncratic. Recall that the multiple citations of periphery authors lends substantial connectivity to the network despite the removal of more frequently authors. Those cited only twice each, for instance, interconnect 73 of the 76 papers. Thus, in addition to the propensity to cite non-core authors, the negative-k-two-path parameter simultaneously also signifies the tendency of papers to be interconnected via 'overlapping chains' rather than through the formation of dense clusters (towards the core or in regions). Overlapping chains of the sort paper A cites authors T1 and T2; paper B cites authors T1, T2 and T3; paper C cites authors T1, T3, and T4; paper D cites T1, T4 and T5; and so on create a network that is densely connected via the core authors (such as T1) and diffusely connected via the periphery ones. This appearance of a diffusely connected network can be attributed to eclectic influences especially characteristics of an uncertain, emergent setting.

Lastly, the hybrid goals of a researcher to produce findings of sufficient generality while also attending to case specifics and/or the effects of localized preferential attachment as measured by local homophily configurations result in clustering. Yet, in this case of an emergent literature, the effects of homophily are not so strong as to dissolve the regional components into pluralistic worlds connected only via a few links as argued by Daipha (2001). To the contrary, consistent with Carolan (2008), the core- as well as periphery-based inter-region connectivity points towards a type of *cohesive clustering* such that the network simultaneously displays characteristics of small-worlds as well as cohesiveness.

To summarize, the findings suggest that the citation network of an emergent area may be characterized by the combined structural features of a small densely connecting core and diffuse connectivity. Even though the stars do not hold the network together in a strictly structural sense (that is, their elimination does not lead to disconnection of the graph), a connecting core nevertheless serves as a 'centralized anchor' providing an emergent area of research with an identity. It materializes as new research within the field cites it almost as a matter of necessity. The periphery, in contrast, reflects demands for innovativeness, competition, eclecticism, and/or uncertainty, leading to low overlap of non-core citations between any two papers but significant overlap across the space of the network. Furthermore, such concentration and diffusion may be expected to span clusters that can be defined along different dimensions.

As the field expands, however, two quite distinctive outcomes may obtain. Innovations may spread out across the breadth of the network diminishing its central coreness. As long as sufficient central citations persist, however, this distribution of innovations across the system should lead to a network that has some stars but is diffusely densely connected as Moody (2004) finds. Alternatively, if overall centrality is more or less completely replaced by localized centrality, then the network might split up into disconnected fragments as found by Daipha (2001). At its current stage, the network under study exhibits tendencies towards both these states—non-core actors are a source of connectivity, and key local figures form the basis of regional clustering. Future research may indicate where the field ends up.

Recall that previous studies emphasized diffuse connectivity and/or fragmentation in the absence of centrality. The differences between my findings and those from previous research can be explained on the basis of a number of factors. First, they could simply be an artifact of the data construction. Some previous research is based on one-mode networks or one-mode projections of two-mode data, the structural features of which are not directly comparable to those of two-mode networks. However, the measures used in this study are based on recent advances in the descriptive and statistical analysis of two-mode data (Latapy et al., 2008; Wang et al., 2009), which go a long way in rendering unipartite and bipartite analyses comparable. Second, as argued previously, these differences could be the result of differences in the structure of emergent and mature areas of research. There is yet a third possibility—methodology. Recall that at the global level, my findings were approximately consistent with those of mature areas of research. It is possible that the application of ERGM to other knowledge networks may reveal a stronger role of preferential attachment and differentiation than is visible from a macro-perspective. Of course, this is an empirical question that bears examination. Thus, even though these precise findings may not apply as well to established areas of research, they are nonetheless promising for linking micro-level action with macro-level outcomes. Additional research that systematically investigates such linkages may help to shed light on the microscopic processes that contribute towards the generation and perpetuation of the global topology in more mature literatures.

Acknowledgements

I am grateful to Ethel Brooks, Phaedra Daipha, Paul McLean and Eviatar Zerubavel for their thoughtful and constructive criticism and guidance on several earlier drafts of this article. I am also thankful to Pip Pattison, Garry Robins, and Peng Wang for their guidance on exponential random graph modeling.

Appendix A. Exponential random graph model fits and goodness of fit

The model shown in the body of the article was selected after fitting numerous models with several different combinations of structural parameters. Convergence was obtained on five models:

1. A Bernoulli model (only edge parameter).
2. A model containing edge, author-k-star and paper-k-two-path parameters.
3. Model 2 supplemented by paper-k-star.
4. Model 2 supplemented by homophily parameters for Bangladesh, China, and Sri Lanka.
5. Model 3 supplemented by Bangladesh, China and Sri Lanka homophily parameters.

Among these models, the fifth was chosen on the basis of three factors: it had the smallest Mahalanobis distance (showing how far a particular network is from the center of a distribution of networks), displayed better goodness of fit (GOF) statistics (i.e. had *t*-ratios of less than 0.1 for the included parameters and a larger number of graph statistics that were not included in the model had *t*-statistics that were less than 2), and offered the most theoretically plausible interpretation of the data (Wang et al., 2009). The GOF of the models was tested using 3000 samples from 5,000,000 iteration simulations after a burn-in of 100,000 iterations. In the GOF, some graph statistics such as lower order star parameters, four-cycles and consequently clustering coefficient and the distribution of degrees were not fit very well by the model. This can be attributed to fact that both indegree and outdegree distributions have a large range of 2–29 and 3–87, respectively. While the k-star parameters capture some of the effects of this degree distribution, they fail to adequately explain the lower star parameters. Encountering similar problems on one of their datasets, Wang et al. argue that such models can nevertheless be used for interpretation as they offer fits superior to the independent assumption of the Bernoulli model. The fitting results and GOF statistics for models 1, 3 and 5 are shown in

Table A1
Parameter estimates for the Bernoulli, realization dependent, and attribute effects models for the multi-cited data.

Parameters	Model 1			Model 3			Model 5		
	EST	*SE*	*T-S*	*EST*	*SE*	*T-S*	*EST*	*SE*	*T-S*
Edge	−3.05*	0.02	0.05	−12.71*	1.05	−0.05	−9.44*	0.84	−0.03
Alternating-paper-k-star				4.64*	0.53	−0.05	3.16*	0.42	−0.03
Alternating-author-k-star				0.77*	0.12	−0.04	0.36*	0.12	−0.02
Alternating-author-k-four-cycle				−0.48*	0.05	−0.05	−0.38*	0.04	0.00
Homophily Bangladesh							0.21*	0.02	−0.04
Homophily China							0.36*	0.03	−0.02
Homophily Sri Lanka							0.29*	0.03	0.02

EST: estimate; SE: standard error; T-S: *T*-statistic.
* Significant estimate (absolute value of *t*-statistic <0.1).

Table A2
Selected goodness of fit statistics for the models shown in Table A1.

Parameters	Model 1	Model 3	Model 5
Edge	0.094	0.194	0.012
Paper-two-star	8.939	8.450	5.336
Author-two-star	6.531	3.812	2.081
Paper-three-star	38.165	42.271	24.799
Author-three-star	15.186	7.707	4.658
Three-circuit	10.385	7.698	4.580
Four-cycle	56.235	36.166	20.373
Alternating-paper-k-star	0.109	0.172	0.020
Alternating-author-k-star	−0.186	0.186	0.023
Alternating-paper-k-four-cycle	3.195	1.578	0.917
Alternating-author-k-four-cycle	−2.595	0.253	−0.046
Std dev of degree dist for paper	26.272	11.273	7.571
Skew of degree dist for paper	2.988	3.957	3.068
Std dev of degree dist for author	18.359	24.650	16.172
Skew of degree dist for author	30.020	47.981	38.264
Global clustering	78.140	55.107	32.154
Mahalanobis distance	2557.992	1497.232	776.793

Tables A1 and A2, respectively. As can be seen from the latter table, model 5 provides consistently lower statistics indicative of better fit, even for the parameters not fit in the model. The fit statistics for the lower order star parameters (particularly paper three-star) and consequently degree distributions are poorly fit in all models, yet model 5 provides the best fit amongst the models shown. The chosen model fits the k-parameters, both the set of included and excluded, particularly well. Furthermore, the Mahalanobis distance is the smallest for Model 5.

References

Agneessens, F., Roose, H., Waege, H., 2004. Choices of theatre events: p* models for affiliation networks with attributes. Metodoloski Zvezki 1 (2), 419–439.
Babchuk, N., Bruce, K., Peters, G., 1999. Collaboration in sociology and other scientific disciplines: a comparative trend analysis of scholarship in the social, physical, and mathematical sciences. The American Sociologist 30 (3), 5–21.
Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society Series B (Methodological) 36 (2), 192–236.
Barabási, A., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509.
Barnett, G., Danowski, J., 1992. A network analysis of the International Communication Association. Human Communication Research 19 (2), 264–285.
Batagelj, V., Mrvar, A., 1996. "Pajek 1.22".
Berik, G., Dong, X., Summerfield, G., 2007. China's transition and feminist economics. Feminist Economics 13 (3), 1–33.
Borgatti, S., Everett, M., 2000. Models of core/periphery structures. Social Networks 21 (4), 375–395.
Breiger, R., 1974. The duality of persons and groups. Social Forces 53 (2), 181–190.
Camic, C., 1992. Reputation and predecessor selection: Parsons and the institutionalists. American Sociological Review 57 (4), 421–445.
Cappell, C., Guterbock, T., 1992. Visible colleges: the social and conceptual structure of sociology specialties. American Sociological Review 57 (2), 266–273.
Carolan, B., 2008. The structure of educational research: the role of multivocality in promoting cohesion in an article interlock network. Social Networks 30 (1), 69–82.
Collins, R., 2001 (1998). The Sociology of Philosophies: A Global Theory of Intellectual Change. Harvard University Press, Cambridge, MA.
Cozzens, S., 1989. What do citations count? The rhetoric-first model. Scientometrics 15 (5), 437–447.
Crane, D., 1988 (1972). Invisible Colleges: The Diffusion of Knowledge in Scientific Communities. University of Chicago Press, Chicago, IL.
Daipha, P., 2001. The intellectual and social organization of ASA 1990–1997: exploring the interface between the discipline of sociology and its practitioners. The American Sociologist 32 (3), 73–90.
Datta, A., 2005. MacDonaldization of gender in urban India: a tentative exploration. Gender, Technology and Development 9 (1), 125–135.
de Solla Price, D.J., 1965. Networks of scientific papers. Science 149 (3683), 510–515.
Egghe, L., Rousseau, R., 2004. How to measure own-group preference? A novel approach to a sociometric problem. Scientometrics 59 (2), 233–252.
Ennis, J.G, 1992. The social organization of sociological knowledge: modeling the intersection of specialties. American Sociological Review 57 (2), 259–265.
Elson, D., 1999. Labor markets as gendered institutions: equality efficiency and empowerment issues. World Development 27 (3), 611–627.
Evans, J., 2007. Consensus and knowledge production in an academic field. Poetics 35 (1), 1–21.
Evans, J., 2005. Stratification in knowledge production: author prestige and the influence of an American Academic Debate. Poetics 33 (2), 111–133.
Faulkner, R., 1983. Music on Demand: Composers and Careers in the Hollywood Film Industry. Transaction Publications, New Brunswick, NJ.
Fuchs, S., 1992. The Professional Quest for Truth: A Social Theory of Science and Knowledge. State University of New York Press, Albany, NY.
Fujigaki, Y., 1998. The citation system: citation networks as repeatedly focusing on difference continuous re-evaluation, and as persistent knowledge accumulation. Scientometrics 43 (1), 77–85.
Gilbert, G., 1977. Referencing as persuasion. Social Studies of Science 7 (1), 113–122.
Ghosh, J., 2002. Globalization, export-oriented employment for women and social policy: a case study of India. Social Scientist 30 (11/12), 17–60.
Goyal, S., van der Leij, M., Moraga-Gonzalez, J., 2006. Economics: an emerging small-world. Journal of Political Economy 114 (2), 403–412.
Hale, A., 1996. The deregulated global economy: women workers and strategies of resistance. Gender & Development 4 (3), 8–15.
Handcock, M., 2003. In: Breiger, R., Carley, K., Pattison, P. (Eds.), Statistical Models for Social Networks: Degeneracy and Inference. National Academies Press, pp. 229–240.
Hill, V., Carley, K., 1999. An approach to identifying consensus in a subfield: the case of organizational culture. Poetics 27 (1), 1–30.
Hoffman, D.L., Holbrook, M.B., 1993. The intellectual structure of consumer research: a bibliometric study of author cocitations in the first 15 years of the journal of consumer research. The Journal of Consumer Research 19 (4), 505–517.
Holland, P., Leinhardt, S., 1981. An exponential family of probability distributions for directed graphs. Journal of the American Statistical Association 76 (373), 33–50.
Hunter, D., 2007. Curved exponential family models for social networks. Social Networks 29 (2), 216–230.
Jarneving, B., 2005. A comparison of two bibliometric methods for mapping of the research front. Scientometrics 65 (2), 245–263.
Kabeer, N., 2004. Globalization, labor standards, and women's rights: dilemmas of collective (in) action in an interdependent world. Feminist Economics 10 (1), 3–35.
Kaplan, N., 1965. The norms of citation behavior: prolegomena to the footnote. American Documentation 16 (1), 179–184.
Kessler, M., 1963. Bibliographic coupling between scientific papers'. American Documentation 14, 10–25.
Kostoff, R., 1998. The use and misuse of citation analysis in research evaluation. Scientometrics 43 (1), 27–43.
Latapy, M., Magnien, C., Vecchio, N., 2008. Basic notions for the analysis of large two-mode networks. Social Networks 30 (1), 31–48.
Leydesdorff, L., Amsterdamska, O., 1990. Dimensions of citation analysis. Science, Technology & Human Values 15 (3), 305–335.
Lievrouw, L., Rogers, E.M., Lowe, C.U., Nadel, E., 1987. Triangulation as research strategy for identifying invisible colleges among biomedical scientists. Social Networks 9 (3), 217–248.
Makino, J., 1998. Productivity of research groups—relation between citation analysis and reputation within research communities. Scientometrics 43 (1), 87–93.
McPherson, M., Smith-Lovin, L., Cook, J., 2001. Birds of a feather: homophily in social networks. Annual Reviews in Sociology 27 (1), 415–444.

Moody, J., Light, R., 2006. A view from above: the evolving sociological landscape. The American Sociologist 37 (2), 67–86.
Moody, J., White, D., 2003. Structural cohesion and embeddedness: a hierarchical concept of social groups. American Sociological Review 68 (1), 103–127.
Moody, J., 2004. The structure of a social science collaboration network: disciplinary cohesion from 1963 to 1999. American Sociological Review 69 (2), 213–238.
Otte, E., Rousseau, R., 2002. Social network analysis: a powerful strategy, also for the information sciences. Journal of Information Science 28 (6), 441.
Paisley, W., 1990. An Oasis where many trails cross: the improbable cocitation networks of a multidiscipline. Journal of the American Society for Information Science 41 (6), 459–468.
Pattison, P., Robins, G., 2002. Neighborhood-based models for social networks. Sociological Methodology 32, 301–337.
Powell, W., White, D., Koput, K., Owen-Smith, J., 2005. Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences. American Journal of Sociology 110 (4), 1132–1205.
Robins, G., Pattison, P., Kalish, Y., Lusher, D., 2007a. An introduction to exponential random graph (p*) models for social networks. Social Networks 29 (2), 173–191.
Robins, G., Pattison, P., Wang, P., 2009. Closure, connectivity and degrees: new specifications for exponential random graph (p*) models for directed social networks. Social Networks 31 (2), 105–117.
Robins, G., Pattison, P., Woolcock, J., 2005. Small and other worlds: global network structures from local processes. American Journal of Sociology 110 (4), 894–936.
Robins, G., Snijders, T., Wang, P., Handcock, M., Pattison, P., 2007b. Recent developments in exponential random graph (p*) models for social networks. Social Networks 29 (2), 192–215.
Safa, H., 1981. Runaway shops and female employment: the search for cheap labor. Signs: Journal of Women in Culture and Society 7 (2), 418–433.
Skvoretz, J., Faust, K., 1999. Logit models for affiliation networks. Sociological Methodology 29, 253–280.
Small, H., 1978. Cited documents as concept symbols. Social Studies of Science 8 (3), 327–340.
Small, H., Griffith, B., 1974. The structure of scientific literatures I: identifying and graphing specialties. Science Studies 4 (1), 17–40.
Snijders, T., 2002. Markov chain Monte Carlo estimation of exponential random graph models. Journal of Social Structure 3 (2), 1–40.
Snijders, T., Pattison, P., Robins, G., Handcock, M., 2006. New specifications for exponential random graph models. Sociological Methodology 36 (1), 99–153.
Standing, G., 1999. Global feminization through flexible labor: a theme revisited. World Development 27 (3), 583–602.
Strauss, D., Ikeda, M., 1990. Pseudolikelihood estimation for social networks. Journal of the American Statistical Association 85 (409), 204–212.
Tzannatos, Z., 1999. Women and labor market changes in the global economy: growth helps inequalities hurt and public policy matters. World Development 27 (3), 551–569.
Wang, P., Robins, G., Pattison, P., 2004. PNet: Program for the simulation and estimation of p* exponetial random graph models.
Van Den Besselaar, P., 2001. The cognitive and the social structure of STS. Scientometrics 51 (2), 441–460.
Wang, P., Sharpe, K., Robins, G., Pattison, P., 2009. Exponential random graph models for affiliation networks. Social Networks 31 (1), 12–25.
Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge, MA.
Whitley, R., 1984. The Intellectual and Social Organization of the Sciences. Oxford University Press, New York, NY.