7-10-2017

Coauthorship in Regional Science: A Case Study of the WVU RRI Research Community

Jing Chen
West Virginia University, jechen@mail.wvu.edu

Randall Jackson
West Virginia University, randall.jackson@mail.wvu.edu

Article

International Regional Science Review
2018, Vol. 41(4) 387-409

DOI: 10.1177/0160017617719124
journals.sagepub.com/home/irx
SAGE

Coauthorship in Regional Science: A Case Study of the WVU RRI Research Community

Jing Chen[1,2] and Randall Jackson[1,2]

Abstract

The year 2015 marked the fiftieth anniversary of West Virginia University's (WVU) Regional Research Institute (RRI), which has played an important role in many scientific collaboration networks. Through social network analysis (SNA) focusing on the RRI research community since its inception in 1965, this article illustrates the role that organizations and the networks they promote can play in scientific problem domains, promoting scholarly collaborations and coauthorship in the field of regional science. We analyzed an evolving WVU RRI coauthorship network that has grown and gained in complexity over time in terms of (1) global metrics, (2) components and cluster analysis, (3) centrality, and (4) PageRank and AuthorRank. The results of these analyses depict a well-developed and influential scientific collaboration structure within both WVU and the regional science research community.

Keywords

coauthorship, regional science, social network analysis, Regional Research Institute

[1] Regional Research Institute, West Virginia University, Morgantown, WV, USA
[2] Department of Geology and Geography, West Virginia University, Morgantown, WV, USA

Corresponding Author:
Jing Chen, Regional Research Institute and Department of Geology and Geography, West Virginia University, Morgantown, WV 26505, USA.
Email: jechen@mail.wvu.edu

Introduction

Research on publication patterns within regional science has been scarce relative to related fields such as economics and geography (Rey and Anselin 2000). Early contributions came from Rey and Anselin (2000), Durden and Knox (2000), and Isserman (2004). With the development of information and communication technologies, research publications are being shared and catalogued online worldwide, and more recent examples including Maier (2007); Haddad, Mena-Chalco, and Sidone (2015); and Rickman and Winters (2016) are beginning to appear. Table 1 presents a brief comparison of these articles in terms of method, data source, scope, and scale. In this context, scope refers to substantive and topical boundaries, while scale identifies the minimum unit for their discussion or conclusion.

Research methods on regional science publication patterns have ranged from statistical analysis on page count and number of academic outputs to more advanced methods like social network analysis (SNA) and citation analysis. In addition to the challenges that analysts face in such studies within academic disciplines, regional scientists must confront the greater scope that characterizes regional science and regional science publications. Some advantage can be gained by focusing attention on coauthorship relationships among regional scientists, as demonstrated recently by Haddad, Mena-Chalco, and Sidone (2015), who analyzed scholarly collaboration among the Brazilian Regional Economics Application Laboratory (REAL) network as a case study focused further still on the patterns of evolving regional science collaboration networks in developing countries. Coauthorship networks are interesting in that they can reflect cooperation between and among individual scholars who share similar research interests. There are also other formats of social networks for analyzing publication patterns like citation and cocitation networks. Compared with these networks, coauthorship networks reveal stronger social relations as citations frequently occur without any social ties (Liu et al. 2005). From a broader perspective, coauthorship has been associated with a higher quality of scientific production (Haddad, Mena-Chalco, and Sidone 2015).

In this article, we assess fifty years of West Virginia University's (WVU) Regional Research Institute (RRI) research community activity to reveal the roles that coauthorship networks and research organizations can play in regional science. The WVU RRI research community is an ideal candidate for such an analysis because it was among the earliest research groups in regional science, and given its continuous operation and long history, it can reflect some of the evolution character of regional science research. In the fifty years since William H. Miernyk came to West Virginia University (WVU) as its founding director, the RRI has employed and engaged numerous scholars, including faculty research associates (FRAs) and graduate research assistants (GRAs). It has provided a rich array of opportunities to

Table 1. A Comparison between Articles on Publication Pattern within Regional Science.

Articles	Main method	Scope	Scale	Source
Rey and Anselin (2000)	Citation analysis	Five core regional science journals	Authors	Journal articles
Durden and Knox[a] (2000)	Descriptive statistics on length and number of papers	*The Review of Regional Studies*	Authors and institutions	Journal articles
Isserman (2004)	Citation analysis	Thirteen regional science journals	Authors	Journal articles
Maier (2007)	Statistical analysis on survey results	Regional science and related journals	Journals	Survey
Haddad, Mena-Chalco, and Sidone (2015)	Social network analysis	Brazilian Regional Economics Application Laboratory alumni	Authors	Authors' CVs on Lattes
Rickman and Winters (2016)	Weighed ranking	Ten regional science journals	Authors and institutions	Journal articles

[a]This article is not exclusively on the publication patterns.

network and learn new ideas in various forms, ranging from seminars and workshops to the Web Book of Regional Science and team-oriented research. Moreover, the institute is recognized nationally and internationally in regional science and regional economics, and it has participated in these and other research domains within and beyond regional science throughout its half-century existence. Our examination of publication patterns via the state of the art in SNA among the RRI research community is designed to detect key authors, their collaboration networks and subnetworks, and research fields for the purpose of understanding coauthorship and organizational structures in regional science.

In the second section, we describe the methods used to collect and compile the RRI research community publication records. In the third section, the basics of SNA on coauthorship networks are introduced, and the results of the analysis are presented in terms of endogenous, exogenous, and entire networks. Specifically, we use the term endogenous network to identify those researchers who have been in residence at and employed by WVU at some time in their careers, while the exogenous network includes the researchers that have coauthored with members of the endogenous network. We refer to the union of these networks as the entire network. The last two sections present discussion and conclusions.

Data Collection and Preprocessing

The Research Community

The first step in this research is to define the WVU RRI research community, which includes RRI directors, research assistant professors, research associates, research associate professors, FRAs, RRI GRAs, and visiting scholars. We define the *endogenous network* to include all of these except for GRAs and visiting scholars, who are excluded in part due to incomplete records on these individuals, and because of the impracticality of identifying and tracing all the publication records for more than 100 visiting scholars and 130 GRAs. The effect of this exclusion, however, is to understate the role of this community in regional science, as many of these individuals have gone on to highly productive and active regional science careers.

Nine regional researchers have been RRI director, acting director, or interim director and have played important roles in leading the institute. They are William Miernyk (1965–1983), Robert Saunders (1969–1970), Patrick Mann (1983–1984), Andrew Isserman (1985–1997), Brian Cushing (1991), Luc Anselin (1997–1998), Scott Loveridge (1999–2000), Ronald Lewis (2000–2001), and Randall Jackson (2002 to present). All directors have had secondary appointments in academic departments.

Research assistant professor and research associate positions have most commonly been non-tenure-track, two- to three-year appointments for junior and senior PhD researchers, respectively. Their appointments have been joint with other departments in some cases and fully within the RRI in others. They usually conduct research of their own design, participate in collaborative research projects, and sometimes teach courses. Two representative examples of this group of scholars are Emily Talen (1995–1998) and Attila Varga (1997–1998), both of whom have become leading scholars in regional science and related fields.

FRAs are specially designated faculty members who have had appointments in six colleges and eleven departments, with interests in regional research. Their disciplinary backgrounds have most often been in geography, economics, and agricultural and resource economics, but law, sociology, history, political science, and others have also been represented. Although FRAs do not have formal appointments with the RRI, they constitute an interdisciplinary scholarly community and participate actively in RRI seminars, workshops, conferences, and other events.

The exogenous network refers to coauthors of members of the endogenous network. In total, there are thirty-nine RRI researchers in the endogenous network and 204 coauthors in the exogenous network, resulting in a total of 243 researchers in the entire analysis.

Data Source

With the boundaries of the network defined, the next task is identifying a suitable source for relevant publication data. Other studies have acquired publication records from various sources like researchers' curriculum vitae (CVs) on the Lattes Platform (Haddad, Mena-Chalco, and Sidone 2015), and directly from academic journals (Rey and Anselin 2000; Durden and Knox 2000; Isserman 2004; Rickman and Winters 2016), but these sources are not well suited to our research for the following reasons:

- there is no existing platform like the Lattes for all the individual researchers and institutions working in the United States,
- there is no source for electronic comparable and up-to-date CVs for the individuals in our analysis, and
- we did not have access to the large number of journals or databases listing their contents.

We considered using Google Scholar as a possible information source, in part because it often includes gray literature like manuscripts and presentations. Yet this potential advantage is also a drawback, because many of these research products end up being published in mainstream outlets, their presence in the database creates issues of double counting that would be extremely difficult if not impossible to resolve.

Hence, we extracted our citation data for every RRI scholar in the endogenous network from the Web of Science, an online scientific citation indexing service provided by Thomson Reuters. According to the description of the journal impact factor calculation process, the Web of Science has selected various types of academic literature, such as scholarly books, peer-reviewed journals, reviews, and editorials from diverse disciplines including since the beginning of the twentieth century. Because the Web of Science data source may unintentionally omit several publications, its databases cannot capture fully the contributions that the RRI research community have made to their disciplines. Despite its shortcomings, the Web of Science was selected as the best choice, given our research objectives.

Article Selection

The next step was to identify the set of publications that fall within the scope of our analysis. Many individuals involved have published both within and beyond regional science. While they might be engaged in regional science research, some of their research is unrelated to regional science or its parent disciplines like geography, regional economics, or planning. An example problem domain is epidemiology, in which many geographers and spatial analysts participate actively. Epidemiological research departs from the core of regional science although epidemiological studies are often published in journals of cognate disciplines, they only rarely appear in core regional science journals. Following publication networks into medical journals, for example, would have the benefit of capturing more comprehensively the extent and impact of the scholar network, but it would also have the drawback of opening the exogenous network definition to include a much larger set of scholars with only a very few meaningful ties to the endogenous network or to regional science more generally. It therefore becomes necessary to define a set of journals to include in and exclude from the analysis.

This is, of course, a common step in such analyses. Within regional science, Haddad, Mena-Chalco, and Sidone (2015) used the entire set of papers published in peer-reviewed journals and chapters of published books by Brazilian University of Illinois Urbana-Champaign (UIUC) REAL alumni, but excluded those published by coauthors outside of the alumni group. Rey and Anselin (2000) focused on papers from five core regional science journals exclusively: *Regional Science and Urban Economics*, *Journal of Regional Science*, *Annals of Regional Science*, *Papers in Regional Science*, and *International Regional Science Review*. Rickman and Winters (2016) likewise focused on a specified set of journals, adding to the list of journals above those considered to be "extended" regional science journals, ones containing regional in the title and others that extensively cite regional science journals: *Journal of Urban Economics*, *Regional Studies*, *the Review of Regional Studies*, *Growth and Change*, and *Journal of Economic Geography*. While defining a relevant set of journals is difficult at best in publication pattern research, it is perhaps even more so in regional science because the boundaries of regional science are often dynamic, fuzzy, and ambiguous (Schaeffer, Jackson, and Bukenya 2011).

In Rey and Anselin's (2000) analysis, the topic of spatial analysis received more attention than any other in the 1990s regional science publications. We therefore chose to add to Rickman and Winter's list of ten core journals spatial analytic nonregional science journal publications in core journals from cognate disciplines. Hence, articles that have been published in journals like *Geographical Analysis* or *International Journal of Geographical Information Science* are included because of their centrality to regional science. Anselin, Talen, and Rey (2000) compiled a list of sixty-six such journals. For the current research, we have chosen to include all sixty-six publications on their list, plus the *Journal of Economic Geography,* a newer journal that Rickman and Winters (2016) included that was not in Anselin, Talen, and Rey (2000). In addition, we encountered one article that has exceedingly large numbers of authors (i.e., Goodchild et al. 1992) and deleted it from the data set because of its potential to overwhelm and distort impacts on the overall SNA results.

The resulting set of 437 articles was drawn from these sixty-seven journals. Table 2 provides a brief overview of publications by journal for the top ten journals. Two-thirds of the publications appeared in these journals, so the other fifty-seven journals published only 145 of the 437 articles. The top three journal, *Journal of Regional Science*, *International Regional Science Review,* and the *Annals of the Association of American Geographers*, accounted for nearly 30 percent of the publications.

Establishing Coauthorships

Once all the publication records were retrieved from the Web of Science and pre-processed, establishing coauthorship links was the next analytical step. We

Table 2. Top Ten Journals by Number of Publications by the Endogenous Network Members.

Ranking	Journals	Category	Count
1	*Journal of Regional Science*	Regional science	55
2	*International Regional Science Review*	Regional science	37
3	*Annals of the Association of American Geographers*	Geography	34
4	*The Professional Geographer*	Geography	30
5	*Journal of the American Planning Association*	Planning	28
6	*Economic Development Quarterly*	Economics	22
7	*Growth and Change*	Regional science	22
8	*Papers in Regional Science*	Regional science	22
9	*Environment and Planning A*	Planning	22
10	*Geographical Analysis*	Geography	20

Source: Calculated by the authors.

developed and used a Python program to parse the publication records and to establish coauthorship links for all researcher pairs. The result is a spreadsheet listing of the coauthorship link records that can be directly imported into Gephi, an open-source software application for SNA (Bastian, Heymann, and Jacomy 2009). We corrected or modified author names manually to account for name ambiguity. For example, names like “Peter V. Schaeffer” and “Peter Schaeffer” that refer to the same person were reconciled in the database manually.[1]

Coauthorship Network Analysis

SNA applies graph theory to study social interactions, where nodes are equivalent to social actors and edges represent the interactions between actors (Wasserman and Faust 1994). The method has been extensively applied to bibliometric research from different perspectives like coauthorship and cocitation networks. A coauthorship network is an important class of social networks for illustrating scientific collaboration and the status of individual researchers (Garfield 1979). For this analysis, we have created a series of weighted undirected networks, where nodes and edges represent individual authors and coauthorship linkages, respectively. The weight of an edge is proportional to the frequency of coauthorship links.

The main software for network visualization and analysis is Gephi (Bastian, Heymann, and Jacomy 2009), in which characteristics of these social networks can be identified directly through such metrics as number of nodes, number of edges, average degree, and network density. We define these metrics in the following subsections.

Global Metrics

Several global metrics that are widely used in SNA are presented below in terms of their algorithms and usage, including the *average degree*, the *network density*, the *clustering coefficient*, the *diameter*, and the *average path length*. The first of these—average degree—refers to the ratio of twice the number of edges to the number of nodes on average. In our case, it is the expected number of coauthors that any researcher in the network will have, reflecting the overall degree of network collaboration over the five decades (although different researchers have been active for different numbers of years).

The second metric is the density of an undirected network, which can be calculated as the ratio of the current number of edges to the maximum possible number of edges between nodes. This ratio is often interpreted as the probability of collaboration for two given authors; larger density networks have a higher probability that a collaboration between two given authors occurs.

The third metric, the overall clustering coefficient, corresponds to three times the number of triangles in the network divided by the number of connected triplets, where a connected triplet refers to a connected subgraph consisting of three vertices and two or three edges. This characteristic measures the extent to which a friend of my friend is also my friend in SNA; in our case, it tells us how likely it is that a coauthor of my coauthor would also be my coauthor.

Finally, the diameter denotes the longest distance of all the shortest paths in the network. It can be also interpreted as the shortest distance between the two most distant nodes. Networks that have smaller diameters are more intensive than extensive. Finally, the average path length measures the average distance between two nodes and is calculated as the ratio of the sum of the shortest paths between all pairs of nodes to the total number of pairs.

Component and Clustering Analysis

A component of a network refers to a subset of the network in which all nodes are connected directly or indirectly. Coauthorship networks most often contain disconnected nodes and therefore form different components. Among these, the component with the greatest number of nodes is called the giant component and can be revealing in terms of overall network structure. Indeed, the giant component is quite often extracted from the original network to enable analysis by methods that can only be used on connected networks (e.g., closeness centrality).

By comparison, Knoke and Yang (2008) identify a cluster or clique as "a group of actors that are more closely related than their fellows." As such, although clusters and components are related, they are not synonymous conceptually. Actors within a cluster are mutually tied to one another, and no other actor is tied directly to all of these actors. A cluster identifies a group of authors that might well be mutually engaged in common research, while a component represents a group of authors that are connected via coauthorship and includes more indirectly connected coauthors. Authors in a component might be separated by paths that span several indirect links, but within a clique, authors are mutually and directly connected.

For our purposes, we use modularity to detect clusters as suggested by Newman (2006). Modularity measures the strength of a subnetwork division. Networks with high modularity are composed of subnetworks with authors that are more likely to publish papers together, while authors in low modularity networks are less likely to collaborate. Mathematically, modularity is calculated as the sum of the difference between the actual number of edges between nodes v and w and the expected number of edges for all node pairs as illustrated in equation (1).

$$Q = \frac{1}{2m} \sum_{vm} \left[A_{vw} - \frac{k_v k_w}{2m} \right] \frac{s_{vw} + 1}{2}, \tag{1}$$

where

- m is the number of nodes in the network;
- A_{vw} equals 1 if there is an edge between nodes v and w and 0 otherwise;
- k_v is the weighted degree of node v; and
- s_{vw} equals 1 if nodes v and w belong to the same community and 0 otherwise.

Communities are determined based on modularity scores using various algorithms.[2]

Centrality Metrics

There is no consensus on a general optimal measure for centrality; each method has its own distinct virtues and utility (Wasserman and Faust 1994). Knoke and Yang (2008) have introduced several local metrics to measure the characteristics of a coauthorship network, and here we discuss three of them for undirected coauthorship networks including (1) degree, (2), betweenness, and (3) closeness.

Degree centrality calculates the number of nodes connected to a given node. Relating this to a coauthorship network, this feature captures the number of coauthors that an author has, so authors with higher degree centralities have more collaborators, irrespective of the number of times collaborators are linked. Hence, this measure does not consider edge weights.

Betweenness centrality measures the number of times a node can be a bridge between two other nodes along the shortest path connecting them in the network, so authors with higher betweenness centralities control the flow of information in the network. A high betweenness might suggest that this researcher plays an important role in connecting other researchers to one another in the network. These individuals may be helping others form coauthorship linkages, which will serve to intensify the level of connectivity in the network. In extreme cases, networks might not be fully connected and nodes with zero betweenness centrality can dominate the entire network.

Finally, closeness centrality informs us the average distance for a given node to any other node in a connected network (e.g., the giant component). It is calculated as the inverse of the average shortest path length from a given node to all nodes, and an author with a high degree of closeness centrality can reach more researchers within a given distance. An author with a high closeness metric is characterized by many short connections to other authors in the network (Liu et al. 2005) and can be closer to the network's "inner circle" than others who might be only weakly involved in the network.

PageRank and AuthorRank

PageRank is an algorithm originally developed by Page and his associates (1999) to measure the importance of Web pages by counting the number and quality of links to a page. It is used to rank Web pages in Google searching results. The entire web (Internet) consists of billions of Web pages, and hyperlinks can be regarded as a network. PageRank can measure the relative importance of Web pages. In our case, a high-quality link is a link to a coauthor that also has a high PageRank, so PageRank not only considers the number of coauthors but also the PageRanks of these coauthors.

PageRank, however, fails to include edge weight, which refers to the number of coauthorships that had occurred between two given authors (Liu et al. 2005). PageRank can falsely imply that all of a given researcher's each coauthors have the same number of publications with the researcher. To offset this PageRank weakness, we also used AuthorRank, a modifiable version of PageRank that incorporates the impact of edge weight (Liu et al. 2005) to evaluate authors' contributing roles in the research community. AuthorRank is often considered to be a more comprehensive measure of the status of an author in the network, especially when some authors frequently collaborate with only few coauthors. Examples in the current network include Giarratani and Rephann, both of whom have their own set of frequent collaborators.

In the next section, a series of collaboration networks have been generated by considering different groups of scholars: we first concentrate on coauthorship in the RRI research community exclusively and build the endogenous network covering the past five decades; then, we assess the entire network by adding the exogenous network of scholars that coauthored with the RRI scholars.

Result and Analysis

Endogenous Network

Our database includes the Web of Science publication records for all the RRI scholars employed and working at WVU at some time in their careers. Connections established while at WVU can be seen in the database to have persisted even for those whose times at WVU were limited. Figure 1 shows the collaborators in the endogenous collaboration network. Not every scholar in the endogenous network has coauthored with others within the RRI community—though they may have

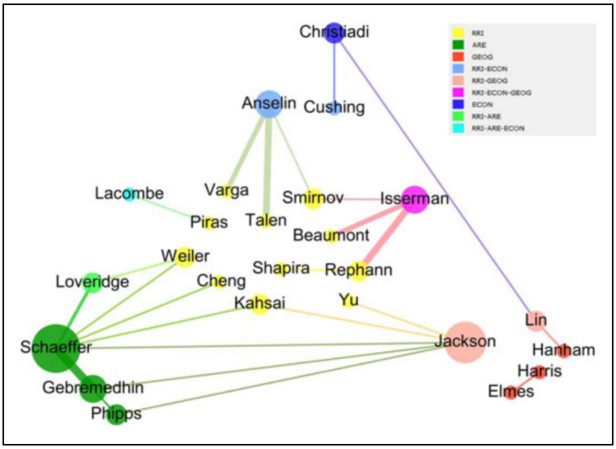

Figure 1. Endogenous collaboration network, 1965–2014.

collaborated with others outside the community—and published articles on regional science and related journals. In Figure 1, we included twenty-five researchers that meet our requirements. Thus, the endogenous collaboration network reveals the coauthorship among WVU RRI scholars.

In Figure 1, labeled nodes are sized based on their degrees, which identify the number of coauthors for individual researchers. Edges are also sized based on their weights, which represent the frequency of coauthorships between scholar pairs. Nodes are also color coded to indicate departmental affiliation. Those with the darkest colors and positioned at the edges of the diagram have a single departmental affiliation, those with lighter colors, Loveridge, Lacombe, Anselin, Cushing, Isserman, Lin, and Jackson, have or had multiple affiliations. Those shown in yellow were supported solely by the RRI. This diagram reveals a number of interesting network traits. First, disciplinary ties influence coauthorships quite strongly, as can be seen by the clusters at the three edges. Second, without exception, interdisciplinary ties involved RRI supported scholars, suggesting the RRI's critical role in fostering such collaborations. And third, except for Schaeffer, who collaborated with two directors, most of the active collaborators have been the RRI directors themselves. In terms of edge strength, several pairs stand out, especially Schaeffer—Gebremedhin, Schaeffer—Loveridge, Anselin—Varga, Anselin—Talen, Isserman—Beaumont, and Isserman—Rephann.

While there are two larger interconnected networks, there are also three isolated links in the outer part of the figure (Lacombe–Piras, Elmes–Harris, and Cushing–Christiadi–Lin–Hanham). Despite attempts to stimulate collaborations, Lacombe and Piras did not expand their networks beyond the RRI, although had their time at RRI been longer such links might have well developed. The Elmes–Harris collaboration is within geography, and indeed, within Geographical Information Science (GIScience).

Table 3. Characteristics of the Noncumulative Subnetworks by Period.

Network metrics	1965–1984	1985–1994	1995–2004	2005–2014
Number of nodes	27	48	79	146
Number of edges	20	45	99	278
Average degree	1.480	1.876	2.538	3.808

Source: Calculated by the authors.

The Cushing–Christiadi–Lin–Hanham isolation is, however, interdisciplinary, explained largely by a common problem domain focus. While interdisciplinary ties have formed between RRI and non-RRI authors, the institute might in future take more steps to help researchers in different disciplines identify areas with higher potential for cross-fertilization. From a broader perspective, potential clearly exists to expand applications of regional science theories and methods to other fields.

One of the two large interconnected networks revolves around Isserman, Rephann, Beaumont, and Anselin, and Varga, while the other includes Loveridge, Schaeffer, Phipps, and Jackson. These two networks correspond strongly to temporal periods at the RRI. The former network reflects earlier research collaborations and makes it clear that collaborations that began at RRI can and have continued long after the researchers leave WVU. For example, the initial coauthorship link between Luc Anselin and Attila Varga, both of whom left RRI in the late 1990s, continued well into the 2000s (e.g., Anselin, Varga and Acs 2000; Acs, Anselin, and Varga 2002; Varga, Anselin, and Acs 2005).

The Entire Collaboration Network

The entire network includes the endogenous plus exogenous network or all RRI researchers and their coauthors. To understand the evolution of the entire coauthorship network, we assess its features initially by period: for 1965–1984, 1985–1994, 1995–2004, and 2005–2014. The first two decades were combined because there were almost no collaborations between 1965 and 1974. Table 3, which describes the basic network characteristics for each period, reveals that the number of nodes and edges grows dramatically, as would be expected, but the growth in edges outpaces the growth in number of nodes so that the average degree more than doubles. In part, this reflects the increasing trend toward collaboration throughout academia. The number of nodes reflects the accumulation of authors included in the whole research community, but the edges refer only to coauthoring relationships during each specified period.

The rest of this section is based on the entire network over the 1965–2014 period as shown in the network graph in Figure 2 and explore some of its interesting features along these dimensions: (1) global metrics, (2) components and clusters, (3) centrality, and (4) PageRank and AuthorRank.

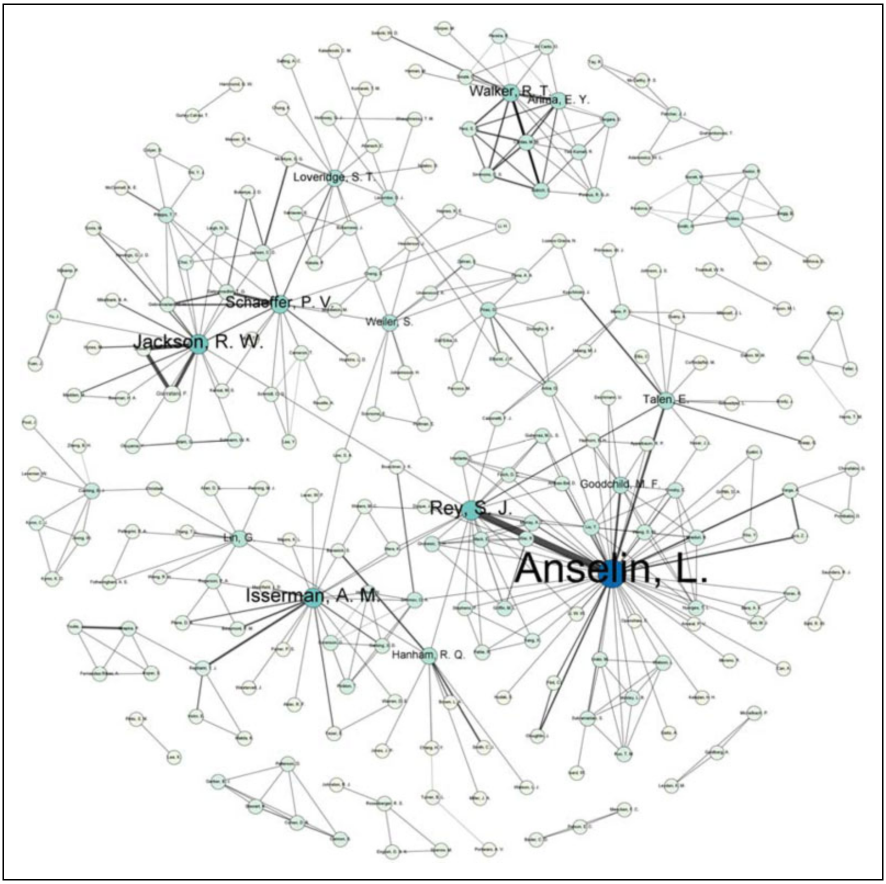

Figure 2. Entire coauthorship network, 1965–2014.

General Statistics

The entire network includes the 39 researchers in the endogenous network and the 204 coauthors in the exogenous network, for a total of 243 individual scholar nodes and 424 edges through the past five decades. Our network graphs display connections among coauthors, with thicker links indicating stronger connections based on numbers of collaborations. Several scholars stand out in Figure 2 because of their comparative importance: notably Luc Anselin, Andrew Isserman, Scott Loveridge, Randall Jackson, Sergio Rey, and Peter Schaeffer. The thickest edge is the one between Luc Anselin and Sergio Rey, identifying them as the most frequent collaborators.

The average degree of 3.49 indicates that in the last fifty years, a given network scholar has had between three and four coauthors on; the density of the network is

0.014, which—combined with the network graph—might suggest that there are still collaboration opportunities between all authors involved in the network. With no other network for comparison, however, this generalization is subjective.

The next three characteristics focus on the giant component, which is the connected subdivision of the network with the largest number of nodes. The high value of the clustering coefficient (0.821) indicates that there is a high chance that a coauthor of a coauthor will often publish together. A diameter of eleven denotes the longest shortest distance in the giant component. Considering "six degrees of separation," which claims that everyone is less than six steps away from anyone else (Guare 1990), there is a high distance between some of the authors. This might reflect, however, the large number of subfields covered by network members. With this large number, there is a high potential for subfield intersections that collectively tie together very different topical problem domains. By comparison, the average path length (4.694) is much shorter than the diameter (11), indicating a short distance on average between authors in the giant component. In other words, the giant component is well connected and there is a high chance that authors in the giant component cooperate with each other.

Components and Clusters

The entire RRI coauthorship network does not form a completely connected graph; there are thirteen components of mutually collaborating researchers. Among these components, the largest component of the network has 188 authors with 333 edges, the second largest component has fifteen authors and the third has eight authors. As with the endogenous network, the institute might take a more proactive stance toward establishing links to strengthen the network.

The giant component accounts for more than 75 percent of all nodes and edges in the entire network. This aspect of the coauthorship network and its component structure is visualized in Figure 3, where nodes and edges represent individual authors and coauthorships, respectively. The giant component and relative sizes and connectivity of the other components are clearly identifiable. The giant component overwhelmingly dominates the graph, indicating that its member researchers play more central and more active roles in tying together the research community.

To better understand the constitution of the giant component, we have analyzed its clusters. There are eleven clusters or communities in the giant component (an interconnected component can be composed of more than one cluster). Authors are more likely to publish papers with other members of the same cluster, but note that not all coauthors will appear in SNA clusters. As an example, although Schaeffer and Jackson have a coauthorship relationship, Schaeffer does not appear in the Jackson cluster, because of the low likelihood that he will publish with Jackson's other coauthors. The three largest clusters are illustrated in Figure 4, where nodes and edges are once again sized based on authors' publication numbers and numbers of coauthorships. The key authors in these three clusters are RRI directors Isserman,

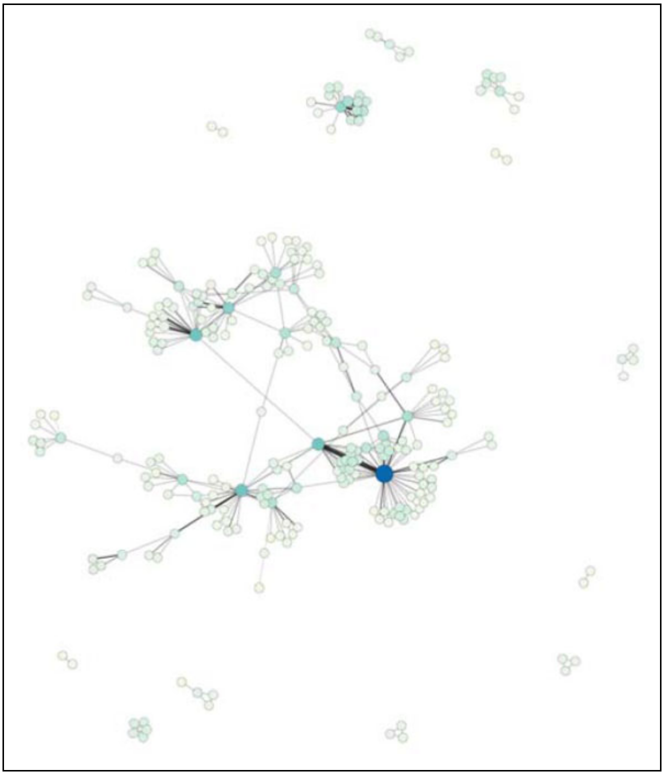

Figure 3. Component analysis for the entire network.

Anselin, and Jackson. In addition to links with the endogenous members of the research community (e.g., Smirnov and Rephann), these three authors have many links to authors outside WVU and form three center–periphery models. Although there are no direct links between these three researchers, partly because of distinct regional science research domains, there is a small number of researchers with more diverse skill sets and or topical interests, like Rey and Smirnov, that link these three research clusters.

Informed readers will recognize the dominant role that Anselin plays in these analyses as a reflection of his prominent position in academia more generally. As such, the Anselin cluster, while dominant, might justifiably be held in contrast to the Isserman and Jackson clusters, both of which more strongly reflect the importance of their relationships with the WVU RRI network than their general academic standing. In both cases, the number of links with exogenous authors reflects, in part, their enhanced visibility to other authors by virtue of their RRI directorial roles. For a contrasting view that might clarify the nature of the remainder of the network, we have generated a network graph shown in Figure 5 that dampens the Anselin effect

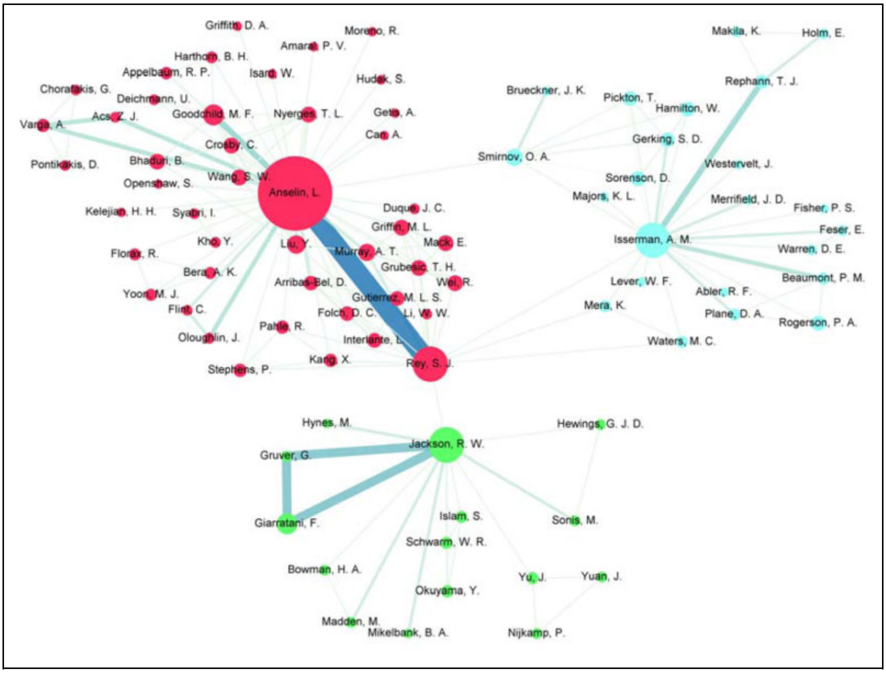

Figure 4. Three clusters in the giant component.

by eliminating all Anselin edges that connect to the exogenous network. In this graph, the roles of Isserman, Jackson, Schaeffer, Walker, Loveridge, Hanham, Weiler, and Talen are much more discernable.

Node Centrality Analysis

The lists of authors who have the largest degree, closeness, and betweenness centralities are illustrated in Table 4, where the WVU RRI endogenous members are in bold. The degree centrality of a node identifies the number of researchers who have published paper(s) together in a coauthorship network. Anselin (fifty-one), Isserman (twenty-one), Rey (twenty-one), Jackson (twenty-one), and Schaeffer (sixteen) are the top five with the highest degree centrality in the coauthorship network, and this can be interpreted that they have more coauthors than any other researchers in the network. Except for Rey, all of them come from the WVU RRI endogenous research community, and three of them have been RRI directors, emphasizing the role that the institute plays in linking regional science researchers.

Betweenness centrality identifies connections to others not directly connected and thus identifies the influence of a member in spreading information through the network. It is calculated as the number of times that a node belongs to the shortest

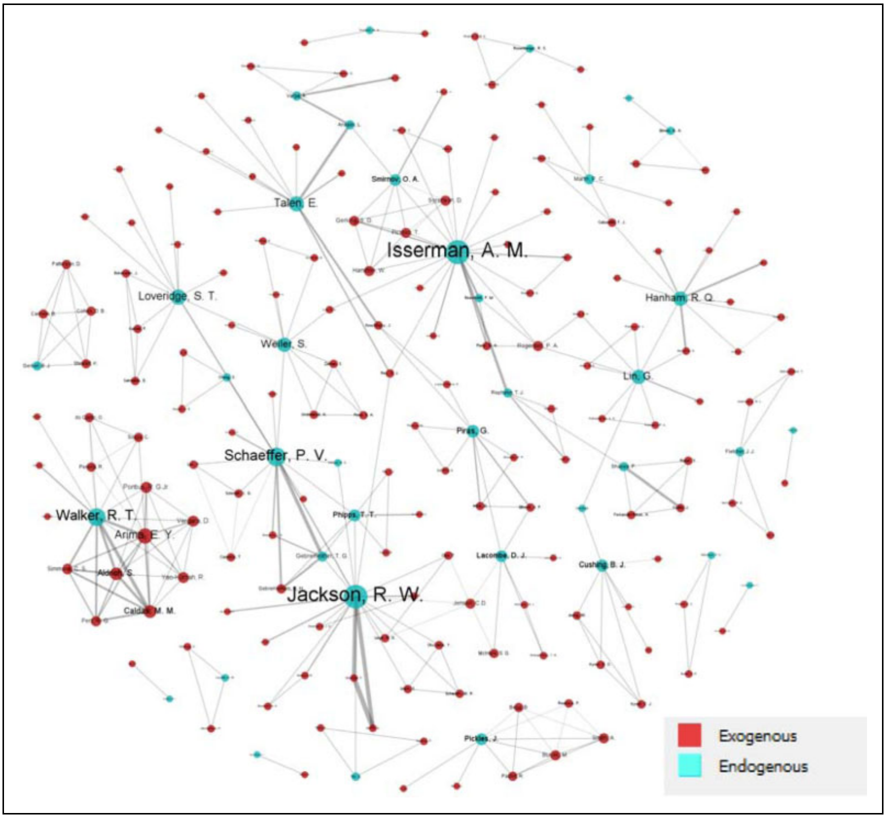

Figure 5. Entire network absent Anselin—exogenous edges.

Table 4. Top Authors with Centrality Metrics in the Entire Network.

Author	Degree	Author	Betweenness	Author	Closeness
Anselin, L.	51	Rey, S. J.	8,335	Rey, S. J.	.351
Isserman, A. M.	21	**Isserman, A. M.**	8,147	**Isserman, A. M.**	.324
Rey, S. J.	21	**Jackson, R. W.**	6,798	**Jackson, R. W.**	.316
Jackson, R. W.	21	**Anselin, L.**	6,174	**Anselin, L.**	.311
Schaeffer, P. V.	16	**Lin, G.**	4,838	**Smirnov, O. A.**	.289
Walker, R. T.	15	Rogerson, P. A.	4,836	**Talen, E.**	.288
Loveridge, S. T.	11	**Schaeffer, P. V.**	3,431	Mera, K.	.287
Talen, E.	11	**Hanham, R. Q.**	2,497	Waters, M. C.	.287
Arima, E. Y.	11	**Weiler, S.**	2,211	Liu, Y.	.279
Lin, G.	10	**Talen, E.**	1,662	Murray, A. T.	.279

Note: The endogenous Regional Research Institute members are in boldface.
Source: Calculated by the authors.

Table 5. Author Ranked According to PageRank and AuthorRank.

Rank	PageRank	AuthorRank
1	**Anselin, L.**	**Anselin, L.**
2	**Jackson, R. W.**	**Jackson, R. W.**
3	**Isserman, A. M.**	**Isserman, A. M.**
4	**Schaeffer, P. V.**	Rey, S. J.
5	Rey, S. J.	**Schaeffer, P. V.**
6	**Hanham, R. Q.**	**Hanham, R. Q.**
7	**Loveridge, S. T.**	**Talen, E.**
8	**Talen, E.**	**Loveridge, S. T.**
9	**Lin, G.**	**Lin, G.**
10	**Weiler, S.**	**Walker, R. T.**

Note: The endogenous Regional Research Institute members are in boldface.
Source: Calculated by the authors.

path between two other nodes. The top ten lists for Degree and Betweenness share seven names in common, although the rank ordering differs. Rey (8,335), Isserman (8,147), Jackson (6,798), Anselin (6,174), and Lin (4,838) have the highest betweenness centralities. Eight of the top ten belong to the endogenous RRI research community, and therefore it is assumed that these authors play significant roles in channeling the flow of information in the network.

Closeness is calculated as the inverse of the average shortest path length between this node and all nodes in a connected network, and its ranking is also illustrated in Table 4. High closeness centrality nodes are characterized by many short connections, indicating that these individuals tend to work most closely with the greatest number of other network members. As such, they can be seen to be heavily involved in the publication activities of the network. Rey (0.351) once again tops the list, and the next four of the top five belong to the endogenous RRI network. There are four names on the closeness top ten list that do not appear on the first two lists.

PageRank and AuthorRank

We have calculated both PageRank and AuthorRank for all 243 researchers; the ten highest scoring researchers are listed in Table 5. As expected, there are many names in common in these two rankings. For example, Anselin (1/1), Jackson (2/2), Isserman (3/3), Schaeffer (4/5), and Rey (5/4) all appear in the top five for both metrics (PageRank/AuthorRank). The differences between these two rankings are subtle and arise because edge weights—the number of collaborations between two authors—are considered in AuthorRank but not in PageRank. So, for example, because Rey and Anselin have collaborated frequently, this greater edge weight results in Rey's AuthorRank being higher than his PageRank. Finally, more than seven of the ten names in these lists are RRI endogenous members.

Discussion

Our focus in this article is similar to the work of Haddad, Mena-Chalco, and Sidone's (2015) in method and scale but differs in source and scope methodologically. Our scope includes the overall RRI research publication patterns in regional science and related journals at individual level for the past five decades, while the main data source of publication records is the Web of Science. Because of these differences, our case illustrates an interesting example to reflect the evolution of WVU RRI–centered coauthorship in regional science. However, like others, our SNA cannot account for several important parameters that could help characterize other dimensions of coauthorship networks, like key words and citations in analyzing publication patterns. Ideally, future analyses will integrate the use of SNA on coauthorship with methods that do incorporate citation, cocitation, key word analysis, and other dimensions of scientific network structures.

Our analysis confirms Haddad, Mena-Chalco, and Sidone's (2015) argument on key factors that contribute to successful collaboration networks including (1) time, (2) key researchers, (3) external links, and (4) the existence of institutional mechanisms to facilitate interaction. We agree with others that building scientific collaboration networks requires at least these four factors. In our case, the fifty-year history of the RRI at WVU has benefited from leadership from key researchers with ties to the larger regional science community, and the institute has continually striven to provide opportunities for researchers within and outside its local research community to interact with one another and has promoted collaborative and interdisciplinary research.

The institute and the role that it plays in the regional science publication arena have been independently acknowledged as well. Described above, Rickman and Winters (2016) reported on publication characteristics in recent years in core regional science journals. Six WVU authors, Anselin, Piras, Jackson, Schaeffer, Lacombe, and Ross, were listed among the top 50 US authors by a number of regional science publications, all of whom are or have been directly affiliated with the RRI. Exogenous network coauthors in the top fifty included Murray, Hewings, Florax, Lesage, and Rey. Similarly, five RRI authors are listed in the world top 100 authors by a number of regional science publications, namely, Anselin, Piras, Jackson, Schaeffer, and Lacombe. In addition to these internal members, there are ten more exogenous members in the world top 100 list. In both rankings, more than 15 percent of the total authors have direct or indirect connections with the RRI.

Rickman and Winters's (2016) analysis, covering the period of 2010–2014, also ranked WVU second in the United States and sixth all over the world in terms of numbers of regional science publications. In their analysis, forty-two journal articles in ten core regional science journals can be attributable to the endogenous RRI-networked researchers: Anselin published eight articles, Randall Jackson and Gianfranco Piras each published seven articles, Peter Schaeffer and Donald Lacombe each published five articles, and others accounted for the remainder. All of the authors of these forty-two publications are members of the RRI research community and most of the publications are coauthored. Finally, eighty-three more journal articles are attributed to our exogenous network.

As a local regional research community in the North American context, our case study also reflects the early lack of and more recently growing diversity in regional science and cognate disciplines. In the endogenous network, Talen was the only female presence early on, and female scholars like Ross have begun to show up only recently. We expect that female researchers will continue to be more prevalent and play more active roles in our research community. Within the institute, we have recruited and welcomed numerous international scholars and postdoctoral researchers of both genders in recent decades from many countries in an explicit attempt to expand diversity. These scholars enrich the experience of our research community and bring their own unique perspectives to solutions to regional problems within and outside the United States.

As regional researchers, we may well expect to see increasing numbers of opportunities to engage in interdisciplinary research. US federal agencies began in the 1990s to require socioeconomic dimensions in large-scale research initiatives, and this trend continues today. Our own experience at the RRI suggests that others in engineering, natural, and physical sciences are recognizing the competitive funding advantages that come from building into their proposals scholars from our community who add depth of understanding and experience in regional socioeconomic modeling. As these ties strengthen, opportunities to further diversify our research networks should continue to grow.

Conclusion

We have constructed and visualized collaboration networks in the WVU RRI research community in terms of endogenous, exogenous, and entire coauthorship networks to reveal the role of the RRI and its network of scholars in regional science. While there is no obvious reference for the entire regional science community against which to benchmark our findings, as the longest surviving regional research center or institute in regional science, and given the active role of its directors from Miernyk through especially Isserman and Anselin to the present, few would question whether the institute and its associated community of scholars have been impactful on the field. Miernyk was highly influential in the early days of the RSAI and continued to participate in regional science organizations and activities through his entire career. The WVU RRI network includes many individuals who have contributed in various ways to regional science including launching the first Regional Science Association International (RSAI) Web site, editing the RSAI Newsletter, serving as RSAI treasurer, editing its journals (including *Papers in Regional Science* and *International Regional Science Review*), and serving on sectional boards and councils. Indeed, its service includes seven past Southern Regional Science

Association presidents. This article takes us one step further in assessing the RRI impact, this time in the context of its coauthorship publications network.

Specifically, the identification of the endogenous and exogenous network in this analysis helps us understand the structure of a specific regional science coauthorship network; identify the key researchers, clusters, and associated problem domains; and potentially identify areas in which supporting organizations like the RRI might take additional steps to enrich the network. The analysis showed that the endogenous network is not fully connected, which allows us to focus on these gaps as potential collaboration opportunities. The overall collaborative activities in the entire network, as measured by the numbers of nodes and edges, have clearly increased and become more complex over the past half century. Meanwhile, network centrality measures allowed us to identify scholars who have played particularly significant roles in linking both endogenous and exogenous members of the research community.

Collaborations and coauthorship networks facilitate cross-fertilization of ideas and the spread of knowledge within and across problem domains. Collaboration networks are formed not only because of individual scholars' own research interests but also as a result of continual organizational effort in guiding interactions, with activities such as seminars, workshops, symposia, and conferences, all of which promote scientific collaboration. Of course, scientific collaboration networks can emerge in the absence of such organizational structures, but with the results of analyses like Rickman and Winters (2016), a case can clearly be made that centers and institutes like the RRI effectively promote research collaborations.

Like others, as the RRI collaboration network continues to both expand and deepen, it includes and involves new and existing members from diverse fields in its research community. These relationships can and should continue to extend advances in regional science to related fields and to leverage research from developments in related fields to more rapidly build and develop the body of regional science knowledge.

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) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Had the data set been substantially larger, this step would likely have been automated.
2. For more information on modularity-based clustering algorithms, see Blondel et al. (2008) and Lambiotte, Delvenne, and Barahona (2009).

References

Acs, Z., L. Anselin, and A. Varga. 2002. "Patents and Innovation Counts as Measures of Regional Production of New Knowledge." *Research Policy* 31:1069–85. doi:10.1016/s0048-7333(01)00184-6.

Anselin, L., E. Talen, and S. J. Rey. 2000. "Subject and Author Indexes to 1999: Regional Science and Related Journal Literature." *International Regional Science Review* 23: 351–57.

Anselin, L., A. Varga, and Z. Acs. 2000. "Geographic and Sectoral Characteristics of Academic Knowledge Externalities." *Papers in Regional Science* 79:435–43. doi:10.1007/pl00011486.

Bastian, M., S. Heymann, and M. Jacomy. 2009. "Gephi: An Open Source Software for Exploring and Manipulating Networks." *ICWSM* 8:361–62.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. 2008. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* 2008:P10008.

Durden, G. C., and H. W. Knox. 2000. "The Southern Regional Science Association and the Review of Regional Studies: History and Characteristics." *The Review of Regional Studies* 30:93–114.

Garfield, E. 1979. *Citation Indexing: Theory and Application in Science, Technology and Humanities*. New York: Wiley.

Goodchild, M.F., R. Haining, S. Wise, G. Arbia, L. Anselin, E. Bossard, C. Brunsdon, P. Diggle, R. Flowerdew, M. Green, M. Green, D. Griffith, L. Hepple, T. Krug, R. Martin, and S. Openshaw 1992. "Integrating GIS and Spatial Data-analysis-problems and Possibilities." *International Journal of Geographical Information Systems* 6:407–23. doi:10.1080/02693799208901923.

Guare, J. 1990. *Six Degrees of Separation: A Play*. New York, NY: Vintage.

Haddad, E. A., J. P. Mena-Chalco, and O. J. G. Sidone. 2015. "Scholarly Collaboration in Regional Science in Developing Countries." *International Regional Science Review* 40: 500–529.

Isserman, A. M. 2004. "Intellectual Leaders of Regional Science: A Half-century Citation Study." *Papers in Regional Science* 83:91–126. doi:10.1007/s10110-003-0178-7.

Knoke, D., and S. Yang. 2008. *Social Network Analysis*, Vol. 154. Thousand Oaks, CA: Sage.

Lambiotte, R., J.-C. Delvenne, and M. Barahona. 2009. "Laplacian Dynamics and Multiscale Modular Structure in Networks." Accessed May 15, 2016. http://arxiv.org/abs/0812.1770.

Liu, X., J. Bollen, M. L. Nelson, and H. Van de Sompel. 2005. "Co-authorship Networks in the Digital Library Research Community." *Information Processing & Management* 41: 1462–80.

Maier, G. 2007. "What Do We Think are the Most Important Journals for Regional Science?" *Scienze Regionali* 6:5–34.

Newman, M. E. 2006. "Modularity and Community Structure in Networks." *Proceedings of the National Academy of Sciences* 103:8577–82.

Page, L., S. Brin, R. Motwani, and T. Winograd. 1999. *The Page Rank Citation Ranking: Bringing Order to the Web*. Accessed May 15, 2016. http://ilpubs.stanford.edu:8090/422/.

Rey, S. J., and L. Anselin. 2000. “Regional Science Publication Patterns in the 1990s.” *International Regional Science Review* 23:323–44.

Rickman, D. S., and J. V. Winters. 2016. “Ranking Authors and Institutions by Publications in Regional Science Journals: 2010-2014.” *Growth and Change* 47:312–36.

Schaeffer, P., R. Jackson, and J. Bukenya. 2011. “Regional Science Reconsidered.” *The Review of Regional Studies* 41:161–77.

Varga, A., L. Anselin, and Z. Acs. 2005. “Regional Innovation in the US over Space and Time.” In *Spillovers and Innovations: Space, Environment and the Economy*, edited by G. Maier and S. Sedlacek, 93–104. Vienna: Springer.

Wasserman, S., and K. Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.