BOX D-2
Mapping the Social Network and Expertise of "Network Science" Researchers

This box presents the anonymized results of a bibliometric analysis[1,2] of the social networks and expertise coverage of network science researchers prepared at the committee's request by K. Börner and W. Ke, of the InfoVis Laboratory at Indiana University. All results are based on the self-reported data in the file named "cleaned_survey_as_of_050318_0910a_posted.xls." Subsequently, the authors report the data® cleaning and analyses performed, major results, and their interpretation. They conclude with a set of recommended topics for further study.

Data Set Used, Analysis Results, and Interpretation

The data file "cleaned_survey_as_of_050318_0910a_posted.xls" comprises 499 completed questionnaires that report 923 "collab_with" links reported under Q2c and 376 "invite" links reported under Q4a. To ensure a high quality of automatic data extraction and analysis, all names reported in free-form text as "Other collaborators" under Q2c and all "Other people to invite" reported under Q4a were not considered. Figure D-2-1 illustrates relationships among the initial invitees, respondents, and identified collaborants.

In total, 1,241 unique names of network science researchers were identified. E-mail addresses were used to ensure that these names are truly unique and represent exactly one person. As requested by the National Research Council, author names were replaced by a unique identification number to preserve the anonymity of authors.

In addition, the 22 (checkable) fields of interest as well as the free-form text of "other" fields of interest reported in Q1c were analyzed. In total, 138 unique fields of interest were identified. Fields that were mentioned most often were computer science (mentioned 201 times), information technology (166), and Internet (156).

Data Quality Issues The "collab_with" links are mostly made to researchers in spatial or thematic proximity. Hence, these links help to grow the social network of network science researchers locally. Colleagues reported that they tried to "invite" people who were not yet in the data set. There was no question that asked users to identify "major players" or "gatekeepers." There are many misspellings of names and disciplines in this data set. Information provided in the "other collaborators" and "other people to invite" section could not be used in this automatic analysis.

Data Analysis Results Here we report "major researchers" who are frequently mentioned in the data set, who act as gatekeepers, and who interlink many scientific fields. In addition, we extracted and will present existing social and collaboration networks. Researchers who are frequently mentioned in the complete data set and the number of times they are listed as a collaborator are given inTable D-2-1. Figure D-2-2 shows the major components (size≥10) network of the network science researcher network (NSRN). The Pajek[3] plot shows exactly 630 of the 1,241 unique researchers, and their "collab_with" links and "invite" links are shown. Each researcher is represented by a node. Node color coding is used to identify researchers that submitted (brown) or did not submit (orange) questionnaires. The node areas' size corresponds to the number of times a researcher is mentioned by other researchers. Each

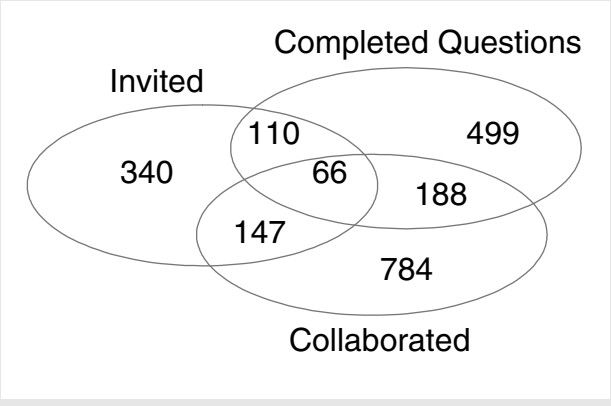

FIGURE D-2-1 Relationships among invitees, respondents, and collaborators.

TABLE D-2-1 Researchers Who Are Frequently Mentioned and Listed as Collaborators

ID	No. Listed	No. Listed as Collaborator
1005	12	8
9	8	7
512	12	6
1009	7	5
139	7	5
1023	8	5
1047	5	4
784	6	4
455	6	4
814	4	4
1238	7	4
925	5	4

BOX D-2 Continued

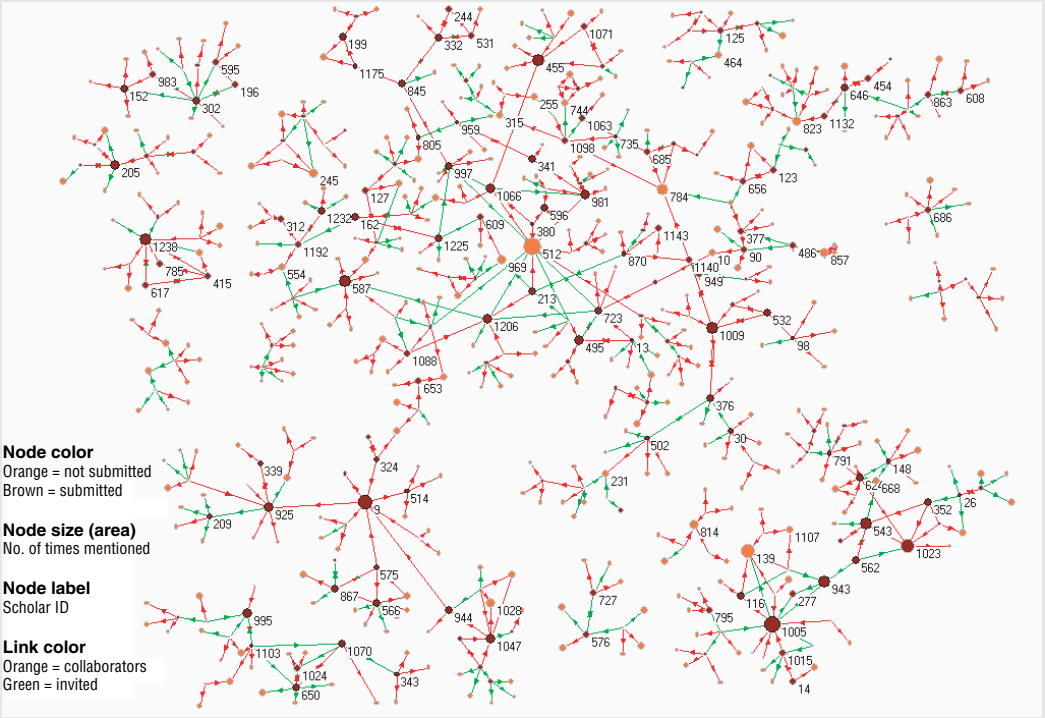

FIGURE D-2-2 Network science researchers network.

node with a betweenness centrality no less than 0.00001 or a size (number of appearances in the data set) ≥3 was labeled with the author's name (ID). Links denote "collab_with" links (in orange) and "invite" links (green).

Subsequently, researchers who act as gatekeepers were identified based on an examination of the betweenness centrality (BC) values[4,5] of nodes in the NSRN. The top 10 researchers are given in Table D-2-2. Figure D-2-3 indicates nodes with a BC value ≥0.00001 by a black ring and shows them in the context of the NSRN.

To examine the community structure of network science researchers, we examined the different components in the NSRN. Table D-2-3 shows the size of existing components, the number of components that have this size, and the total number of nodes in these components. The largest component in the NSRN is shown in Figure D-2-4 using the color coding introduced in Figure D-2-2. It represents the current coherent core of the new field of network science.

TABLE D-2-2 Researchers Who Act as Gatekeepers

ID	No. Mentioned	Betweeness Centrality Value
1066	4	0.00020275
997	2	0.00017878
981	4	0.00015093
9	7	0.00013408
341	2	0.00012502
925	4	0.00010882
845	3	0.00010688
959	1	0.00009716
1225	2	0.00007773
162	3	0.00007060

box continues

BOX D-2 Continued

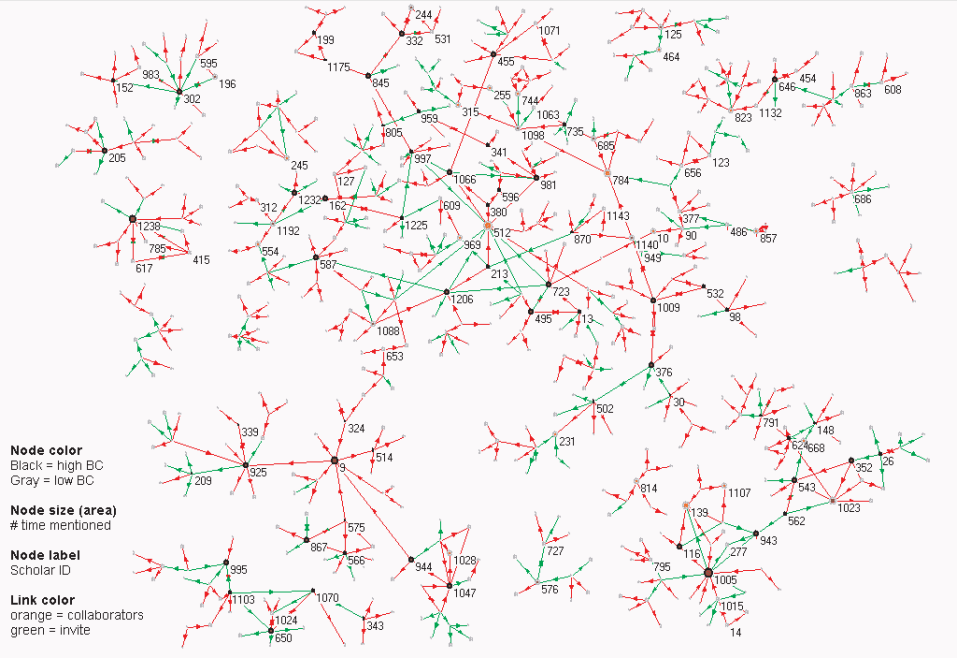

FIGURE D-2-3 Researchers with high BC values (in black) and low BC values (in gray).

TABLE D-2-3 Components in the NSRN

Size	No. of Components	No. of Nodes
1	77	77
2	32	64
3	25	75
4	45	180
5	10	50
6	12	72
7	7	49
8	1	8
9	4	36
10	4	40
11	2	22
13	1	13
14	1	14
15	1	15
17	1	17
18	1	18
30	1	30
33	1	33
73	1	73
355	1	355
Total		1,241

box continues

BOX D-2 Continued

Interpretation Compared with maps of other scientific disciplines, the NSRN clearly exhibits the characteristics of a new and emergent research area: It consists of many unconnected networks of collaborating network science researchers, and the existing networks show a rather heterogeneous coverage of research topics.

Figure D-2-5 is a map of all network science researchers visualized in VxInsight.[6] The map at the left-hand side shows the NSRN. On the right, the very same graph is shown in "landscape" mode, with colored dots representing the self-reported interest profiles of researchers. A white dot denotes that the researcher listed "biology" as a principal field of interest in Q1c. Yellow denotes "computer sciences," light blue "Internet," blue "physics," and green "sociology." As can be seen, there are no groupings of researchers with similar fields of interest. Instead, very different research interests seem to be almost equally distributed over the NSRN.

As the field of network science matures, subareas devoted to the study of specific research fields are likely to emerge, and many of the separate components will exhibit collaboration links, weak or strong and temporary or stable.

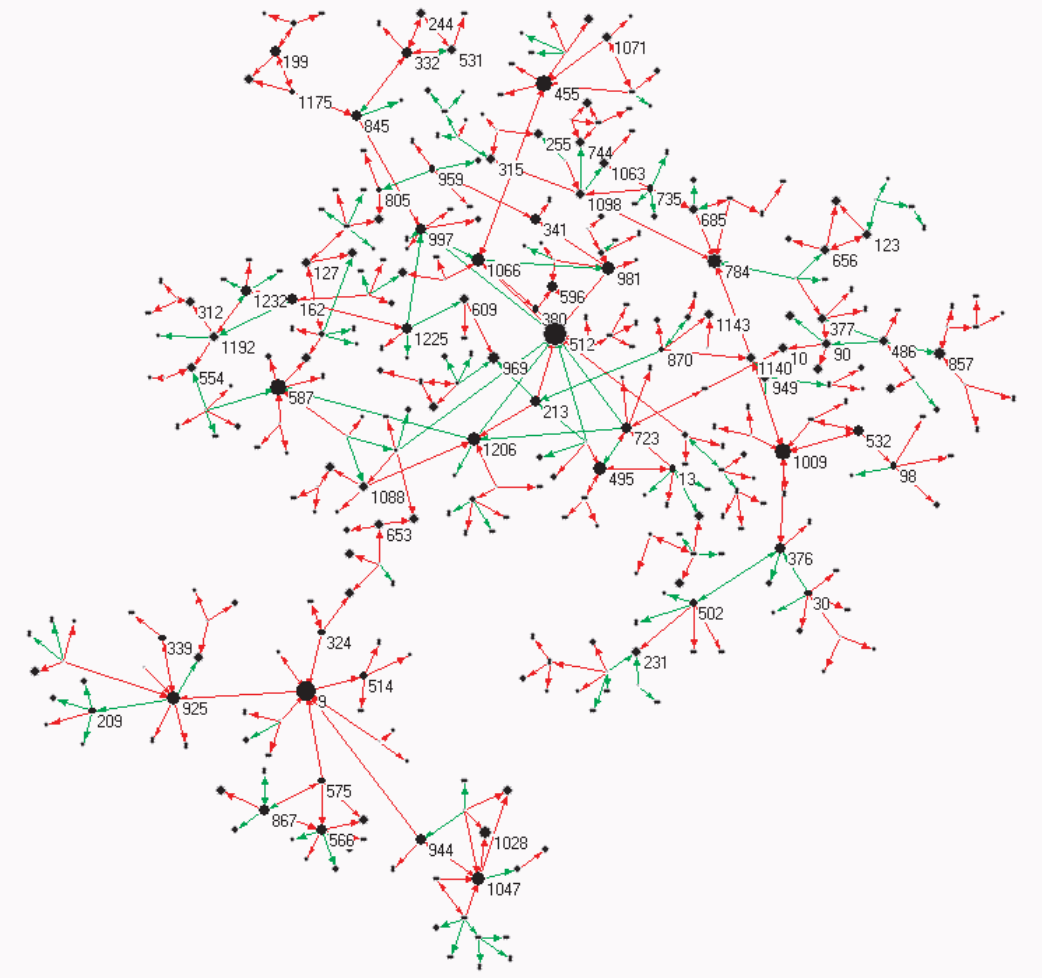

FIGURE D-2-4 Largest component of the NSRN.

box continues

BOX D-2 Continued

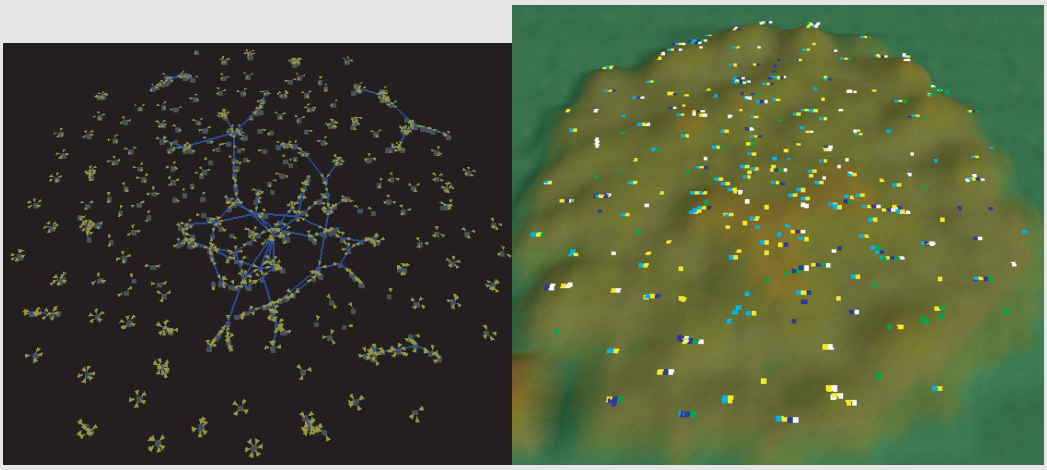

FIGURE D-2-5 Disciplinary heterogeneity of the NSRN.

Social Network of Network Science Researchers: Topics for Further Study

• *Increase our understanding of the interplay of affiliation, thematic, and social interrelations among today's network science researchers.* Invite key network science researchers to identify and label the main research groups key shown in Figure D-2-2.

• *Bibliometric analysis of networkscience publications, patents, and funding data.* The InfoVis Laboratory at Indiana University is developing the sociotechnical infrastructure to analyze the structure and evolution of scientific disciplines and all of science on a large scale.[7] Major publication, patent, and grant databases (covering mostly U.S. research) are available, as are scalable algorithms and compute power. A detailed, objective analysis of scholarly data would complement the self-reported, subjective data and its analysis reported here.

• *Development of an online portal that tracks and communicates the evolution of network science research and results.* Geospatial and semantic maps of network science researchers and publications presented here and proposed in Shiffrin and Börner (2004) can be made available online as a unique interface to data sets, publications, and expertise related to network science research. Researchers interested in being "on the map" should be given the option to submit data about their publications, collaborators, etc. The incentives for researchers to contribute high-quality data can be further increased by using this online map to make funding decisions much as PI's resumes are used today. Assuming that a comprehensive set of high-quality data can be acquired on a continuous basis, an interactive, continuously evolving, weather-forecast-like map of network science research can be made available to grant agencies, researchers, practitioners, and society at large.

NOTE: The authors would like to thank Will E. Leland for compiling the data set used in this study and for insightful feedback on previous results. This work is supported by a National Science Foundation CAREER Grant under IIS-0238261.

[1]Börner, K., C. Chen, and K. Boyack. 2003. Visualizing knowledge domains. In Annual Review of Information Science & Technology, B. Cronin, ed. Medford, N.J.: Information Today, Inc./American Society for Information Science and Technology.

[2]Shiffrin, R.M., and K. Börner, eds. 2004. Mapping knowledge domains. Proceedings of the National Academy of Sciences of the United States 101 (Suppl. 1).

[3]Batagelj, V., and A. Mrvar. 1997. Pajek: Program package for large network analysis. Available at http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

[4]Freeman, L.C. 1997. A set of measuring centrality based on betweenness. Sociometry 40: 35–41.

[5]Brandes, U. 2001. A faster algorithm for betweenness centrality. Journal of Mathematical Sociology 25(1): 163–177.

[6]Davidson, G.S., B. Hendrickson, D.K. Johnson, C.E. Meyers, and B.N. Wylie. 1998. Knowledge mining with VxInsight: Discovery through interaction. Journal of Intelligent Information Systems 11(3): 259–285.

[7]Available at http://iv.slis.indiana.edu.