Social network analysis: a powerful strategy, also for the information sciences

Evelien Otte

UIA, Rectoraat – Academische Planning, Wilrijk, Belgium

Ronald Rousseau

UIA, IBW and KHBO, Belgium

Received 2 April 2002
Revised 28 August 2002

Abstract.

Social network analysis (SNA) is not a formal theory in sociology but rather a strategy for investigating social structures. As it is an idea that can be applied in many fields, we study, in particular, its influence in the information sciences. Information scientists study publication, citation and co-citation networks, collaboration structures and other forms of social interaction networks. Moreover, the Internet represents a social network of an unprecedented scale. In all these studies social network analysis can successfully be applied. SNA is further related to recent theories concerning the free market economy, geography and transport networks. The growth of SNA is documented and a co-author network of SNA is drawn. Centrality measures of the SNA network are calculated.

1. Introduction

Network studies is a topic that has gained increasing importance in recent years [1]. The fact that the Internet is one large network is not unrelated to this [2–4]. Social network theory directly influences the way researchers nowadays think and formulate ideas on the Web and other network structures, such as those shown in enterprise interactions [5]. Even within the field of sociology, network studies are becoming increasingly important.

In this article we will study social network analysis (SNA) and show how this topic may be linked to the information sciences. Of course, Internet studies will also be mentioned, as the World Wide Web represents a social network of a scale unprecedented in history [5].

Interest in networks, and in particular in social network analysis, has only recently bloomed in sociology [6, 7]. There are, however, many related disciplines where networks play an important role. Examples are computer science and artificial intelligence (neural networks), recent theories concerning the Web and free market economy [8], geography and transport networks [9, 10]. In informetrics, researchers study citation networks, co-citation networks, collaboration structures and other forms of social interaction networks [11–19]. Underlying any concrete network lies a graph, a structure studied by mathematicians since Euler solved the problem of the Königsberg bridges.

2. What is social network analysis?

Social network analysis, sometimes also referred to as 'structural analysis' [20], is not a formal theory, but rather a broad strategy for investigating social structures. The traditional individualistic social theory and data analysis considers individual actors making

Correspondence to: Professor R. Rousseau, Universiteitsplein 1, B- 2610 Wilrijk, Belgium. E-mail: ronald.rousseau@kh.khbo.be

choices without taking the behaviour of others into consideration. This individualistic approach ignores the social context of the actor [21]. One could say that properties of actors are the prime concern here. In SNA, however, the relationships between actors become the first priority, and individual properties are only secondary. Relational data are the focus of the investigations. It should be pointed out, however, that individual characteristics as well as relational links are necessary in order to fully understand social phenomena [21]. Wetherell *et al.* [22, p. 645] describe SNA as follows:

> Most broadly, social network analysis (1) conceptualises social structure as a network with ties connecting members and channelling resources, (2) focuses on the characteristics of ties rather than on the characteristics of the individual members, and (3) views communities as 'personal communities', that is, as networks of individual relations that people foster, maintain, and use in the course of their daily lives.

Another important aspect of SNA is the study of how structural regularities influence actors' behaviour. It is clear that ideas originating in SNA can offer added value to investigations in many disciplines, in particular those mentioned in the Introduction.

One distinguishes two main forms of SNA: the ego network analysis, and the global network analysis. In 'ego' studies the network of one person is analysed. An example in the information sciences is White's description of the research network centred on Eugene Garfield [23]. In global network analyses one tries to find all relations between the participants in the network.

SNA, although considered here mainly within the field of sociology, is an interdisciplinary technique developed under many influences, the most important ones coming from mathematics and computer science. In sociology itself SNA can be described as originating from sociometrics (important names here are Lewin and Moreno), the Harvard School (with W. Lloyd Warner) and the Manchester anthropological school (with Barnes, Mitchell and Bott).

3. Some notions from graph theory

3.1. Directed and undirected graphs

A directed graph G, a digraph, consists of a set of nodes, denoted as $N(G)$, and a set of links (also called arcs or edges), denoted as $L(G)$. In this text the words 'network' and 'graph' are synonymous. In sociological research nodes are often referred to as 'actors'. A link e, is an ordered pair (i,j) representing a connection from node i to node j. Node i is called the initial node of link e, $i = \text{init}(e)$, and node j is called the final node of the link: $j = \text{fin}(e)$. If the direction of a link is not important, or equivalently, if existence of a link between nodes i and j necessarily implies the existence of a link from j to i, we say that this network is an undirected graph. A path from node i to node j is a sequence of *distinct* links $(i, u_1), (u_1,u_2), \ldots, (u_k,j)$. The length of this path is the number of links (here k+1). In this article we only use undirected graphs. Consequently, the following definitions are only formulated for that case. A co-authorship network is an example of an undirected graph: if author A co-authored an article with author B, automatically author B co-authored an article with A. An undirected graph can be represented by a symmetrical matrix $M = (m_{ij})$, where m_{ij} is equal to 1 if there is an edge between nodes i and j, and m_{ij} is 0 if there is no direct link between nodes i and j.

3.2. Components

A component of a graph is a subset with the characteristic that there is a path between any node and any other one of this subset. If the whole graph forms one component it is said to be totally connected.

Next we define some indicators describing the structure (cohesion) of networks and the role played by particular nodes [9]. Many more are described in the literature, but we will restrict ourselves to these elementary ones.

3.3. Definition: density

The density is an indicator for the general level of connectedness of the graph. If every node is directly connected to every other node, we have a complete graph. The density of a graph is defined as the number of links divided by the number of vertices in a complete graph with the same number of nodes. For an undirected graph G with N nodes, the density D is defined as:

$$D = \frac{2 * (\#L(G))}{N(N-1)}$$

The density is sometimes called the gamma index [9].

3.4. Definition: centrality [24, 25]

The most important centrality measures are: degree centrality, closeness centrality and betweenness centrality.

Degree centrality of a node is defined as the number of ties this node has (in graph-theoretical terminology, the number of edges adjacent to this node). In mathematical terms degree centrality, $d(i)$, of node i is defined as:

$$d(i) = \sum_j m_{ij}$$

where $m_{ij} = 1$ if there is a link between nodes i and j, and $m_{ij} = 0$ if there is no such link. In a co-author graph the degree centrality of an actor is just the number of authors in the graph with whom she has co-authored at least one article. The degree centrality in an N-node network can be standardized by dividing by N–1: $d_S(i) = d(i)/(N(1)$.

Closeness centrality of a node is equal to the total distance (in the graph) of this node from all other nodes. As a mathematical formula closeness centrality, $c(i)$, of node i can be written as:

$$c(i) = \sum_i d_{ij}$$

where d_{ij} is the number of links in a shortest path from node i to node j. Closeness is an inverse measure of centrality in that a larger value indicates a less central actor while a smaller value indicates a more central actor. For this reason the standardized closeness is defined as $c_S(i) = (N–1)/c(i)$, making it again a direct measure of centrality.

Individual closeness measures can be combined to define global measures, characterizing the cohesion of the total network. The best-known ones are the Wiener index [26, 27] and the BRS compactness [28–30].

Finally, betweenness centrality may be defined loosely as the number of times a node needs a given node to reach another node. Stated otherwise, it is the number of shortest paths that pass through a given node. As a mathematical expression the betweenness centrality of node i, denoted as $b(i)$ is obtained as:

$$b(i) = \sum_{j,k} \frac{g_{jik}}{g_{jk}}$$

where g_{jk} is the number of shortest paths from node j to node k ($j, k \neq i$), and g_{jik} is the number of shortest paths from node j to node k passing through node i. According to Borgatti [25], the purpose is to provide a weighting system so that node i is given a full centrality point only when it lies along the only shortest path between j and k. Betweenness gauges the extent to which a node facilitates the flow in the network. It can be shown that for an N-node network the maximum value for $b(i)$ is $(N^2-3N+2)/2$. Hence the standardized betweenness centrality is:

$$b_s(i) = \frac{2b(i)}{N^2 - 3N + 2}$$

Besides the Wiener index and the BRS compactness measure mentioned above, every centrality measure can be used to derive a centrality measure, C, for the whole network. This is done as follows:

$$C_{\text{network}} = \frac{\sum_j \left(C_{\max} - C_j\right)}{\text{max value possible}}$$

This formula can be applied for determining degree, closeness and betweenness centrality. The summation goes over all nodes of the network; $C_{\max}$ is the largest value obtained in the network under study, and 'max value possible' refers to the maximum value possible for the numerator, given the total number of nodes. It can be shown that the total network C-measure is 1 for a star (one central point and all other nodes connected only to this central node).

3.5. Definition: cliques

A clique in a graph is a subgraph in which any node is directly connected to any other node of the subgraph.

Figure 1 and Table 1 present a simple example of three networks and differences in their characteristics.

The density index, D, indicates that the networks a, b and c (considered in this order) become increasingly dense. All centrality measures show that node u is the centre, and that the other nodes become increasingly central (that is to have a larger centrality value) when considering graphs a, b and c (in this order). The networks themselves, taken as a whole, show less and

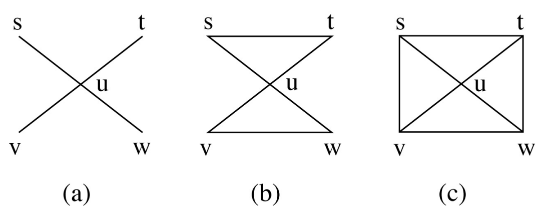

Fig. 1.

Table 1
Density and centrality measures for the networks of Fig. 1

	Network a	Network b	Network c
D	(2*4)/(5*4) = 0.4	(2*6)/(5*4) = 0.6	(2.8)/(5.4) = 0.8
$d_S(s) = d_S(t) = d_S(v) = d_S(w)$	1/4 = 0.25	2/4 = 0.5	3/4 = 0.75
$d_S(u)$	4/4 = 1	4/4 = 1	4/4 = 1
d_{netw}	12/12 = 1	8/12 = 0.667	4/12 = 0.333
$c_S(s) = c_S(t) = c_S(v) = c_S(w)$	4/7 = 0.571	4/6 = 0.667	4/5 = 0.8
$c_S(u)$	4/4 = 1	4/4 = 1	4/4 = 1
c_{netw}	1	0.778	0.467
$b_S(s) = b_S(t)$			
$b_S(v) = b_S(w)$	0	0	1/3 = 0.333
$b_S(u)$	1	4/6 = 0.667	2/3 = 0.667
b_{netw}	1	16/24 = 0.667	(4/3)/24 = 0.056

less centrality. The relative values also illustrate the difference between the three centrality measures considered here.

For more information on graphs the reader is referred to References [31–36].

4. The development and growth of social network analysis

It is often stated that SNA has recently experienced rapid growth, but this statement has rarely been proved. In order to corroborate the latter, three databases were consulted: CSA Sociological Abstracts Database (SA), Medline Advanced and PsycINFO. It is clear that the first one would have sufficed for the investigation, but it was interesting to find out whether related fields used this technique and, if so, whether a similar growth could be detected there. The yearly number of articles related to SNA was counted, as well as the number of subjects within the field that were discussed. For the latter aspect the subject headings of Sociological Abstracts were used (see Appendix, Table A1).

4.1. Growth in the number of published articles

Searching in SA (for the period 1963–2000) 1601 articles were retrieved having 'social network analysis' in the 'Subject heading' field. In Medline Advanced 308 articles were retrieved and in PsycINFO 105. The 1601 articles found in SA have publication dates between 1969 and 2000. There are, however, only two articles from the year 1969 and two from 1971. These are omitted from the graph (Fig. 2). Similarly, when the search was performed (beginning 2001) data for the year 2000 were not yet complete. Hence these are also not shown in the graph.

Figure 2 clearly shows the fast growth of the field in recent years. More specifically, the real growth began around 1981, and there is no sign of decline. This is most obvious in SA, but Medline Advanced also shows a modest increase. This proves that other fields, besides sociology, have used the term and the techniques.

Next, quantification of this growth was attempted. A linear regression analysis was performed on the SA data. This led to the equation

$$p(t) = -20.74 + 5.958t \qquad (1)$$

where time $t = 1$ in 1974 and $p(t)$ denotes the number of published articles as a function of t. The correlation coefficient is 0.956, which is highly significant. An attempt to fit an exponential distribution yielded unsatisfactory results. From this it was concluded that the field has experienced a linear growth over the last 25 years.

A similar fitting exercise on the Medline data resulted in the following regression line:

$$p(t) = -5.018 + 1.232t \qquad (2)$$

with a correlation coefficient of 0.927.

The cumulative number of articles published in the field (SA data only) was next investigated, from the year 1974 on. This is illustrated in Fig. 3.

Applying non-linear regression leads to the equation:

$$P(t) = -34 + 0.96t^{2.28} \qquad (3)$$

where $P(t)$ denotes the cumulative number of articles in the SA database, with $t = 1$ for the year 1974 ($r^2 = 0.998$).

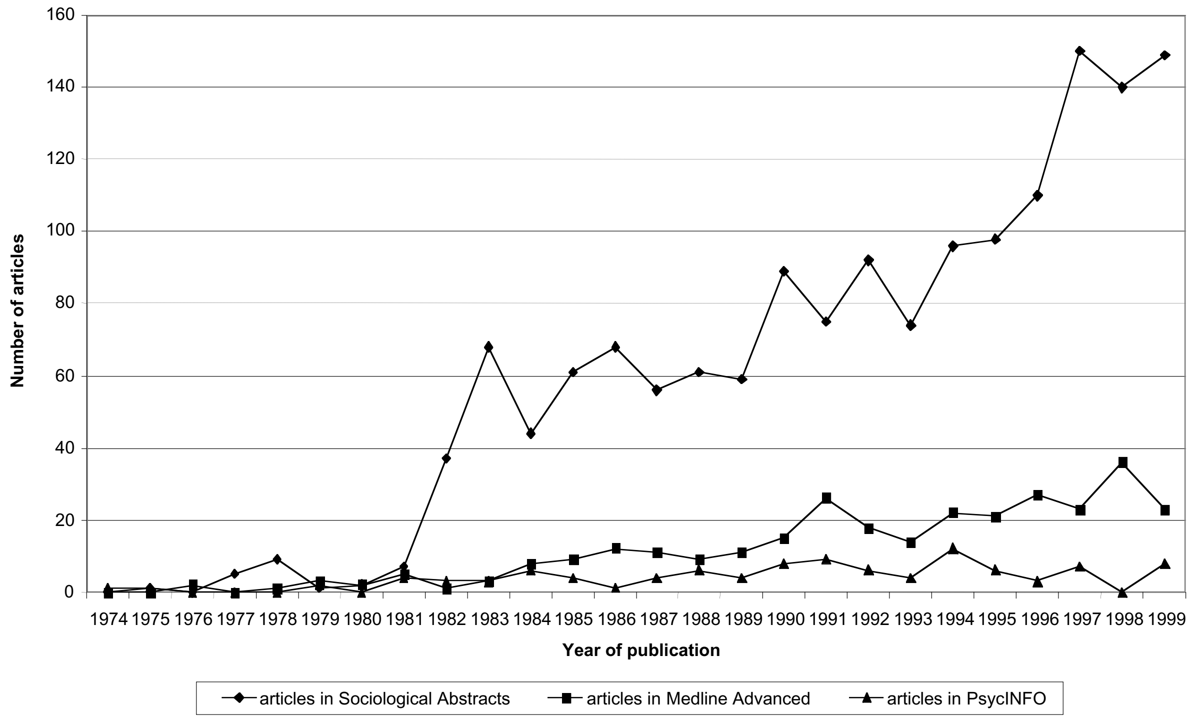

Fig. 2. Growth of social network analysis.

Note that a linear equation for year-by-year data mathematically leads to a square law for cumulative data. Statistical fitting yields an exponent of 2.28 instead of the theoretically expected value of 2, a result that falls within our expectations.

The sum of the results from the three databases (Fig. 4) shows that the development of SNA began later in

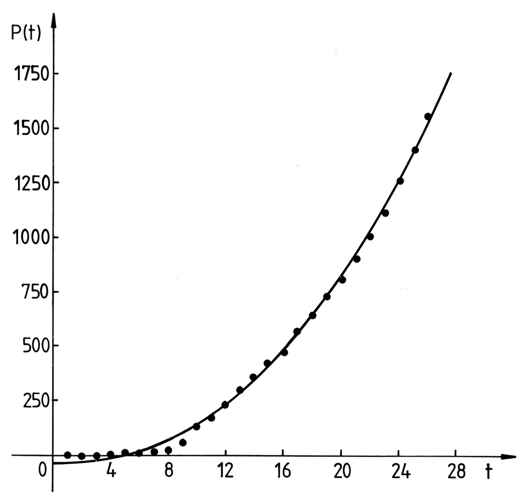

Fig. 3. Cumulative number of articles on social network analysis.

the fields of medicine and psychology (the graphs split up only from 1984 on). This is no surprise as the method was developed in sociology (at least under that name), and only later adopted in other fields.

The three graphs (Figs 2–4) demonstrate the fact that it was only in the early 1980s that SNA started its career. The main reasons for this are the institutionalization of social network analysis since the late 1970s, and the availability of basic textbooks and computer software.

The institutionalization of the field began with the foundation in 1978 by Barry Wellman of the International Network for Social Network Analysis (INSNA). This is the professional association for researchers interested in social network analysis. Its principal functions are the publication of the informal bulletin *Connections*, containing news, scholarly articles, technical columns, abstracts and book reviews; sponsoring the annual International Social Networks Conference (also known as Sunbelt) and maintaining electronic, web-based services for its members. The society also publishes, in association with Elsevier, the peer-reviewed international quarterly *Social Networks*.

The earliest basic text that the authors know of dealing exclusively with social network analysis is Knoke and Kuklinski's *Network Analysis*, published in

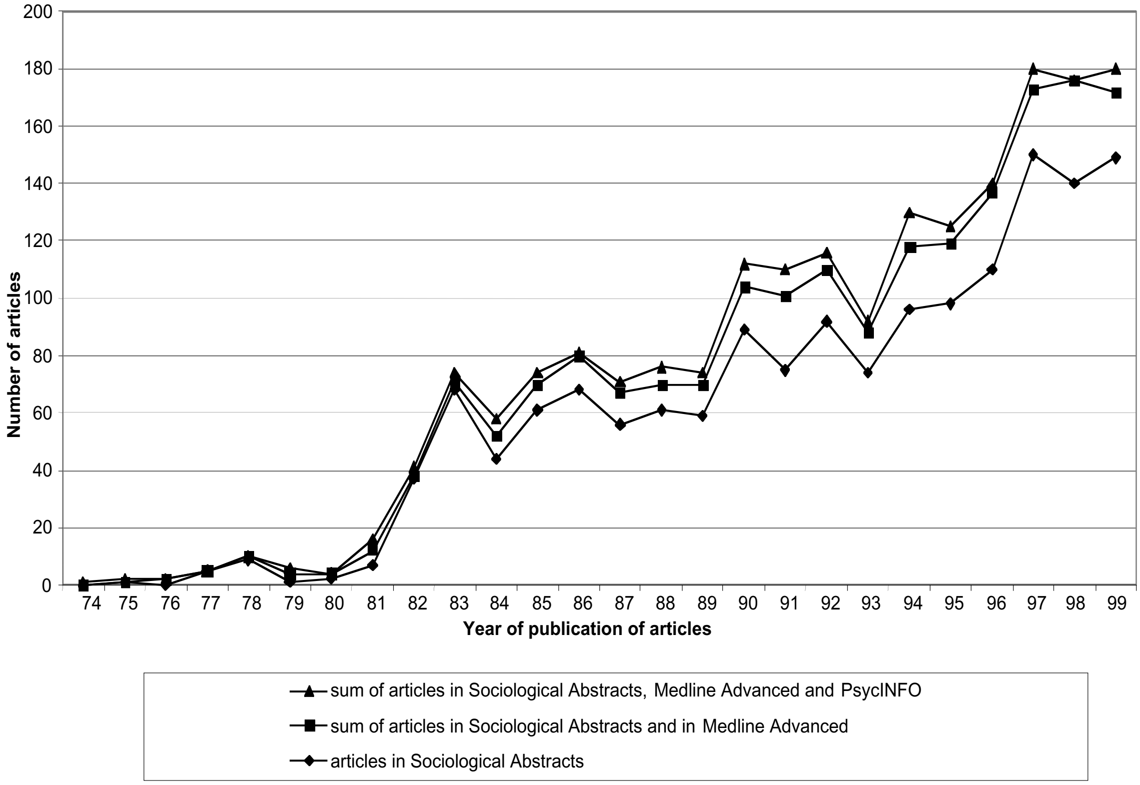

Fig. 4. Sum of SNA articles from the three databases.

1982. Other important books having influenced the growth of the discipline are Wellman and Berkowitz' *Social Structures: a Network Approach* (1988), Scott's *Social Network Analysis: a Handbook* (1991), and Wasserman and Faust's *Social Network Analysis: Methods and Applications* (1994).

The development of dedicated software also led to an increase in interest in the field and its methods. The best-known (and very user-friendly) program for the analysis of social networks is UCInet (a free evaluation version can be downloaded from www.analytictech.com/ucinet_5_description.htm). UCInet can easily be combined with Krackplot, a well-known program for drawing social maps. Other examples of computer programs for social network analysis are Gradap, Multinet, Negopy and Pajek.

4.2. Articles dealing with social network analysis and relationships with other subjects

It is to be expected that growth in the number of studied subjects follows the growth in the number of articles on social network analysis. In order to study this the following method was applied based on the subject classification scheme of Sociological Abstracts. This scheme divides the field of sociology into 33 subjects, each subdivided into a variable number of subfields (see Appendix, Table A1). Social network analysis is a subfield under the main heading 'Complex Organization'.

Articles may be assigned to different subject headings. Most articles found under the heading SNA are also classified under other headings. The fact that an article is classified with the code for social network analysis (0665) and one or more other codes indicates that the author of that article has either discussed relationships between SNA and that other subfield, has applied SNA together with techniques from that other subfield, or has applied SNA in that other subfield.

From 1984 on (for each year) the total number of different subfield codes that were assigned together with the code 0665 for SNA were counted. Figure 5 shows the linear best-fitting function of the number of additional (i.e. together with 0665) codes (Pearson correlation coefficient = 0.89). Further, a striking change was observed in the subjects studied using SNA, or related to it: in the early 1990s most articles dealt with family and socialization, while at the end of

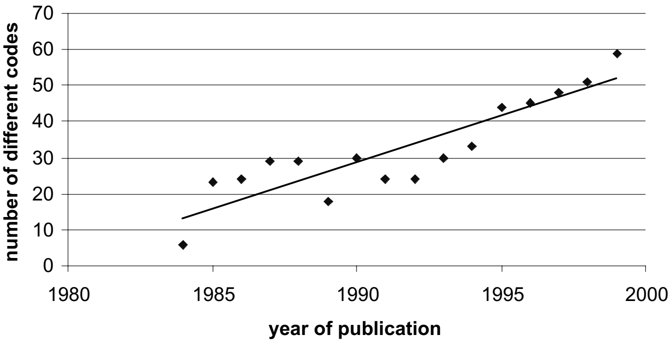

Fig. 5. Number of different subfield codes assigned together with the code for SNA.

this period the SNA articles mostly dealt with the sociology of health and medicine. Indeed, social network analysis is now often applied in AIDS and drug abuse studies.

5. An SNA co-authorship network

In this section a network analysis was performed of authors in the field of social network analysis. The central players are pointed out as are the underlying collaborative relationships between authors. Co-authorship, a (strong) form of collaboration, is not the only way to describe relationships between scientific authors. Citation network, for instance, could reveal other relationships, but these are not studied in this article.

In the 1601 articles dealing with SNA there were 133 authors occurring three times or more. Forming an undirected co-authorship graph (of these 133 authors) led to a big connected component of 57 authors, two components of four authors, two components of three authors, seven small components consisting of two authors and 48 singletons. The central cluster of 57 authors will be concentrated on. Most important scientists in the field belong to this cluster. There are, however, exceptions, the most notable one being Ronald S. Burt (University of Chicago), who has 17 articles in the Sociological Abstracts database. As these articles are either written as a single author or with authors who have only one article in the database, they were excluded. Presumably these collaborators were students.

Network analysis was performed using UCInet while the map was drawn with Pajek (Package for Large Network Analysis). Figure 6 shows the network of network analysts.

5.1. Network analysis of the central cluster

Density. The density is an indicator for the level of connectedness of a network. It is given as the number of lines in a graph divided by the maximum number of lines (the case where every author is connected to every other one). Hence it is a relative measure with values between 0 and 1. The density for the central network of network analysts is 0.05, so this network is clearly not dense at all, but very loose.

Degree centrality. Degree centrality is equal to the number of connections that an actor (a node) has with other actors. In this network being a central author means that this scientist has collaborated (in the sense of co-authored) with many colleagues. The author with the highest degree centrality is Barry Wellman (University of Toronto), who has a degree centrality of 9. The degree centrality of the whole network is 11%, indicating that many authors are not connected.

Closeness. Another way of studying centrality is using the closeness indicator. This indicator is more general than the previous one, because it takes the structural position of actors in the whole network into account. A high closeness for an actor means that he or she is related to all others through a small number of paths. The most central author in this sense is Patrick Doreian (University of Pittsburgh). The closeness of the whole network is 14%.

Betweenness. This measure is based on the number of shortest paths passing through an actor. Actors with a high betweenness play the role of connecting different groups, as ‘middlemen’. Again Patrick Doreian has the highest betweenness. The betweenness of the whole network is 47%.

Cliques. UCInet found 16 cliques, meaning 16 subgraphs consisting of three or more nodes. The largest one consists of six authors: Bernard, Johnsen, Killworth, McCarthy, Shelley and Robinson. The second largest one consists of the five authors: Erger, Lovaglia, Markovsky, Skvoretz and Willer.

6. A bibliometric analysis of the SNA database

Barry Wellman is the most prolific author in the field of social network analysis, based on the Sociological Abstracts database. He published 31 articles in the investigated period (21 as first author). Table 2 shows the most prolific authors (using total counts) in the field of SNA. More details can be found in Otte [7].

The author publication frequencies can easily be described by a Lotka distribution, which is a power

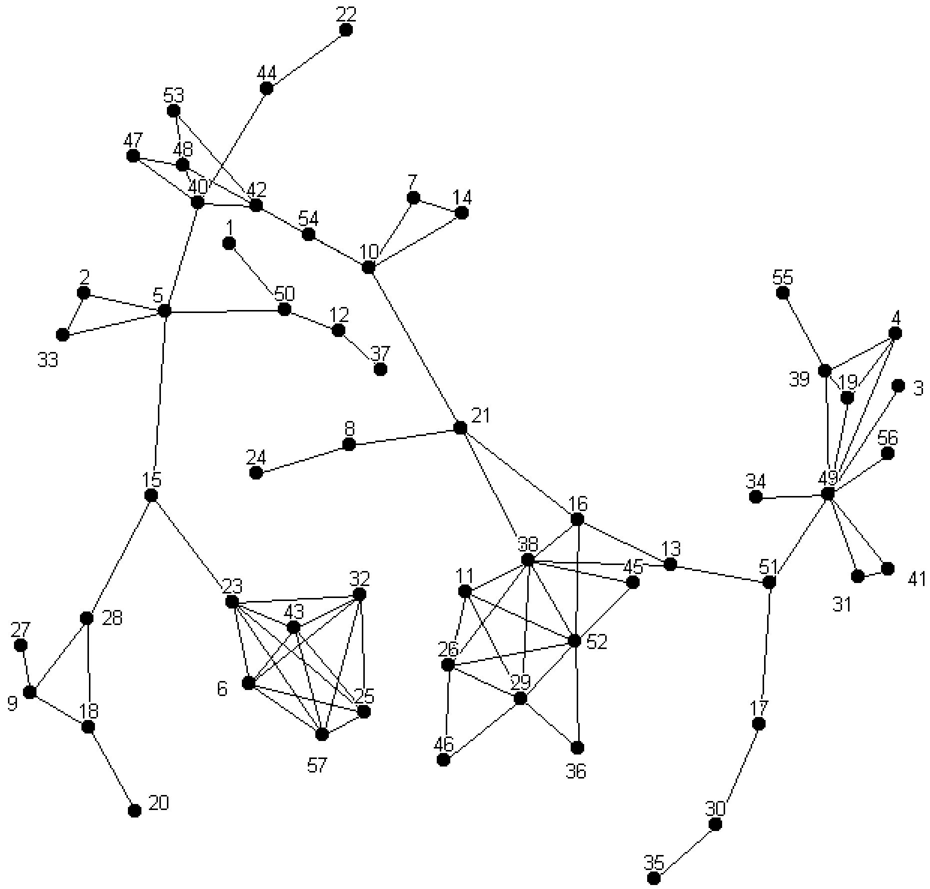

Fig. 6. The network of network analysts. 1, D.D. Brewer; 2. E.J. Bienenstock; 3. S.D. Berkowitz; 4, M. Gulia; 5, P. Bonacich; 6, H.R. Bernard; 7, V. Batagelj; 8, K. Carley; 9, K.E. Campbell; 10, P. Doreian; 11, J.S. Erger; 12, L.C. Freeman; 13, K. Faust; 14, A. Ferligoj; 15, N.E. Friedkin; 16, T.J. Fararo; 17, J. Galaskiewicz; 18, J.S. Hurlbert; 19, C. Haythornthwaite; 20, V.A. Haines; 21, N.P. Hummon; 22, I. Jansson; 23, E.C. Johnsen; 24, D. Krackhardt; 25, P.D. Killworth; 26, M.J. Lovaglia; 27, B.A. Lee; 28, P.V. Marsden; 29, B. Markovsky; 30, M.S. Mizruchi; 31, D.L. Morgan; 32, C. McCarthy; 33, M. Oliver; 34, S. Potter; 35, B. Potts; 36, T. Patton; 37, D. Ruan; 38, J. Skvoretz; 39, J.W. Salaff; 40, T.A.B. Snijders; 41, J.J. Suitor; 42, F.N. Stokman; 43, G.A. Shelley; 44, M. Spreen; 45, J. Szmatka; 46, S.R. Thye; 47, M.A.J.Van Duijn; 48, G.G. Van de Bunt; 49, B. Wellman; 50, C. Webster; 51, S. Wasserman; 52, D. Willer; 53, E.P.H. Zeggelink; 54, K.L. Woodard; 55, S.L. Wong; 56, N.S. Wortley; 57, S. Robinson.

law. Using the software program available in the journal *Cybermetrics* [37], it was found that:

$$f(y) = \frac{0.790}{y^{2.727}} \quad (4)$$

where $f(y)$ denotes the relative number of authors with y publications. According to the Kolmogorov–Smirnov statistic the fit is excellent ($D_{max} = 0.009$). Counting only first authors (see Appendix, Table A2, for data) the following frequency distribution was obtained:

$$f(y) = \frac{0.802}{y^{2.800}} \quad (5)$$

where, now, $f(y)$ denotes the relative number of authors with y first authorships, $y > 0$. Again the fit is excellent (the Kolmogorov–Smirnov $D_{max} = 0.004$). For the role of Lotka's law in informetrics the reader is referred to Wilson's [38] review article.

As it is the thesis of this article that network analysis is a field equally important to sociology as to the information sciences, the database LISA was consulted in order to find out whether the top scientists in SNA had also published in journals covered by this library and information science database. It turned out that, of the 47 most prolific SNA authors (that is those who wrote at least six articles), 12 had articles in LISA (not necessarily as first author). Articles were published in

Table 2
Most prolific authors in SNA

Author	Number of published articles	Author	Number of published articles
Wellman, Barry	31	Klovdahl, Alden S.	9
Skvoretz, John	24	Lovaglia, Michael J.	9
Bonacich, Phillip	20	Snijders, Tom A.B.	9
Everett, Martin G.	20	Stokman, Frans N.	9
Willer, David	19	Wasserman, Stanley	9
Burt, Ronald S.	17		
Friedkin, Noah E.	16	9 authors	8
Borgatti, Stephen P.	14	8 authors	7
Johnsen, Eugene C.	14	11 authors	6
Faust, Katherine	13	13 authors	5
Markovsky, Barry	13	20 authors	4
Doreian, Patrick	12	57 authors	3
Marsden, Peter V.	12	223 authors	2
Mizruchi, Mark S.	10	1308 authors	1

journals such as *Scientometrics*, *JASIS(T)*, *Journal of Classification, Information Processing and Management* (and its precursor, *Information Storage and Retrieval*) and *Social Science Computer Review*. Table 3 gives the list of these authors. The fact that Patrick Doreian heads this list is perhaps not surprising in view of his prominent role in the SNA network. Note that even this mini-list yields a perfect Lotka distribution (exponent equal to 1.97).

7. Use of network analysis in the information sciences

In this section a short, and hence necessarily incomplete, overview of articles and authors in the information sciences (or at least in information science and multidisciplinary journals) that have used the network approach in their investigations is given.

In information science studies publications, citations, co-citations [39, 40] as well as collaborations give rise to networks [9]. Recently other collaborations, such as movie actor collaborations, have also inspired fellow scientists [41]. These authors and others link their research to the so-called small-world phenomenon or 'six degrees of separation' phenomenon [42–45]. A small-world network is then characterized as a network exhibiting a high degree of clustering and having at the same time a small average distance between nodes. Moreover, the 'hubs' and 'authorities' approach is related to the Pinski–Narin influence weight citation measure [46] and mimics the idea of 'highly cited documents' (authorities) and reviews (hubs) [1].

As early as 1972 Nance and co-workers [47] studied information networks as directed graphs. In their article the message transfer structure was the central notion. These authors defined measures of network structure such as the accessibility and the flexibility in message transfer.

Shaw [48] used the random graph hypothesis (lines of a graph are randomly selected from the set of all possible ones) to study the validity of thresholded co-citation graphs. Logan and Pao [49, 50] investigated the structure of co-author graphs and determined central authors based on their position in the co-author graph. The presence of these central authors created order and

Table 3
SNA authors and number of articles in LISA (1969–2001)

Author	Number of articles in LISA
Patrick Doreian	8
Nan Lin	5
Barry Wellman	4
Kathleen Carley	2
H. Russell Bernard	2
Douglas R. White	2
Barry Markovsky	1
Tom A.B. Snijders	1
Michael J. Lovaglia	1
Linton C. Freeman	1
Thomas J. Fararo	1
Karen E. Campbell	1

structure in the graph. They were, moreover, especially important for the information transfer and exchange within the graph. These 'middlemen', however, were very often not represented in first author relationships.

In his thesis Pritchard [9] investigated the question whether it is possible to classify information transfer networks on the basis of their topological structure. In his work he noted the close relationship between transport geography and information transfer, and made successful use of graph-theoretic measures. Applications were given for five citation networks: four patent networks and a comprehensive bibliography on bibliometrics. It was found that the bibliometrics network and the patent networks had different patterns.

Martinsons *et al.* [51] have recently shown that the field of strategic management has entered the mainstream of social science. They studied the network of journals in the field, using an asymmetric theory (journal-to-journal citations are not symmetric), where the notions of feeder and receiver journals are central.

Network sociologists Haythornthwaite and Wellman [52] used a social network approach in their *JASIS* article studying how work and friendship ties in a university research group were related to the kind of media used for information exchange. They found that face-to-face means of communication were preferred, supplemented primarily by e-mail. Further, those persons who had the most frequent contacts used more different media. In a somewhat similar vein, Kretschmer studied the structure of interpersonal relationships based on co-authorships [53]. She found an increasing social distance with declining similarity (as measured by co-authorships) expressed succinctly as 'birds of a feather flock together'.

8. Conclusion

Social network analysis is a typical example of an idea that can be applied in many fields. With mathematical graph theory as its basis it has become a multidisciplinary approach with applications in sociology, the information sciences, computer sciences, geography etc. In this article the growth of social network analysis within sociology is documented. Based on the study of a collaboration network it was found that Barry Wellman and Patrick Doreian are the most central scientists in the field. These authors have also published several articles in information science journals. It has been shown where social network analysis can be linked to work in the information sciences. Now that not only computer scientists but also more and more information scientists are becoming interested in the Internet (under the names webometrics, cybermetrics), it is clear that social network analysis will find more and more of a place in the information sciences. A perusal of Manfred Kochen's book *The Small World* [44] clearly shows that more than 10 years ago he was already fully aware of the importance of network theory for sociology as well as for the information sciences. Finally, in conclusion, the relationship between networks, percolation theory and the 'informetric laws' is pointed out [54].

Acknowledgements

The first author thanks Peter Van Aelst (UIA) for guidance during her master's work. The authors thank M. Dekeyser for editorial help.

References

[1] J. Kleinberg, Authoritative sources in a hyperlinked environment, *Journal of the ACM* 46 (1999) 604–632.

[2] C. Borgman, *From Gutenberg to the Global Information Infrastructure. Access to Information in the Networked World* (MIT Press, Cambridge, MA, 2000).

[3] C. Lynch, The evolving Internet: applications and network service infrastructure, *Journal of the American Society for Information Science* 49 (1998) 961–972.

[4] L. Björneborn and P. Ingwersen, Perspectives of webometrics, *Scientometrics* 50 (2001) 65–82.

[5] P. Raghavan, Social networks on the Web and in the enterprise. In: N. Zhong, Y.Y. Yao, J. Liu and S. Ohsuga (eds), *Web Intelligence: Research and Development.* Lecture Notes in Artificial Intelligence 2198 (Springer, Berlin, 2001), pp. 58–60.

[6] L. Garton, C. Haythornthwaite and B. Wellman, Studying online social networks, *Journal of Computer-Mediated Communication* 3(1) (1997). Available at: www.ascusc.org/jcmc/vol3/issue1/garton.html

[7] E. Otte, Social network analysis: the origin, the development and the structure of the field (in Dutch). Master's thesis (UIA, Department of Political and Social Sciences, Belgium, 2001).

[8] J. Markoff, On the Web, as elsewhere, popularity is self-reinforcing, *The New York Times* 21 June 1999.

[9] A. Pritchard, On the structure of information transfer networks. M.Phil. thesis (School of Librarianship, Polytechnic of North London, London, 1984).

[10] S. Démurger, Infrastructure development and economic growth: an explanation for regional disparities in China? *Journal of Comparative Economics* 29 (2001) 95–117.

[11] H.D. White and B. Griffith, Author cocitation: a literature measure of intellectual structure, *Journal of the American Society for Information Science* 32 (1981) 163–171.

[12] R.J.W. Tijssen, A quantitative assessment of interdisciplinary structures in science and technology: co-classification analysis and energy research, *Research Policy* 21 (1992) 27–44.

[13] R. Todorov, Displaying content of scientific journals: a co-heading analysis. *Scientometrics* 23 (1992) 319–334.

[14] G. Melin, R. Danell and O. Persson, A bibliometric mapping of the scientific landscape on Taiwan, *Issues and Studies* 36 (2000) 61–82.

[15] H. Kretschmer, Distribution of co-author couples in journals: 'continuation' of Lotka's law on the 3rd dimension. In: M. Davis and C. Wilson (eds), *Proceedings of the 8th International Conference on Scientometrics and Informetrics* (BIRG, University of NSW, Sydney, 2001), pp. 317–325.

[16] M. Mymoon and K.S. Raghavan, Mapping biotechnology research using co-classification analysis. In: M. Davis and C. Wilson (eds), *Proceedings of the 8th International Conference on Scientometrics and Informetrics* (BIRG, University of NSW, Sydney, 2001), pp. 447–457.

[17] P.S. Nagpaul, Visualizing the cooperation networks of elite institutions in India. In: M. Davis and C. Wilson (eds), *Proceedings of the 8th International Conference on Scientometrics and Informetrics* (BIRG, University of NSW, Sydney, 2001), pp. 459–468.

[18] Y. Fang and R. Rousseau, Lattices in citation networks: an investigation into the structure of citation graphs, *Scientometrics* 50 (2001) 273–287.

[19] L. Egghe and R. Rousseau, Co-citation, bibliographic coupling and a characterization of lattice citation networks, *Scientometrics* 55 (2002) 349–361.

[20] B. Wellman and S.D. Berkowitz, *Structural Analysis in the Social Sciences 2: Social Structures: a Network Approach* (Cambridge University Press, Cambridge, 1988).

[21] D. Knoke and J.H. Kuklinski, *Network Analysis*. Sage University Paper Series on Quantitative Applications in the Social Sciences, no. 07–028. (Sage, Newbury Park, CA, 1982).

[22] C. Wetherell, A. Plakans and B. Wellman, Social networks, kinship, and community in Eastern Europe, *Journal of Interdisciplinary History* 24 (1994) 639–663.

[23] H.D. White, Toward ego-centered citation analysis. In: B. Cronin and H.B. Atkins (eds), *The Web of Knowledge* (Information Today, Medford, NJ, 2000), pp. 475–496.

[24] L.C. Freeman, Centrality in social networks: I. Conceptual clarification, *Social Networks* 1 (1978) 215–239.

[25] S.P. Borgatti, Centrality and AIDS, *Connections* 18 (1995) 112–115.

[26] H. Wiener, Structural determination of paraffin boiling points, *Journal of the American Chemical Society* 69 (1947) 17–20.

[27] D.H. Rouvray, Predicting chemistry from topology, *Scientific American* 255(3) (1986) 36–43.

[28] R.A. Botafogo, E. Rivlin, and B. Shneiderman, Structural analysis of hypertexts: identifying hierarchies and useful metrics, *ACM Transactions on Information Systems* 10 (1992) 142–180.

[29] L. Egghe and R. Rousseau, BRS-compactness in networks: theoretical considerations related to cohesion in citation graphs, collaboration networks and the Internet, *Mathematical and Computer Modeling* (in press).

[30] L. Egghe and R. Rousseau, A measure for the cohesion of weighted networks. *Journal of the American Society for Information Science and Technology* 54 (2003) (in press).

[31] W.K. Chen, *Applied Graph Theory* (North-Holland, Amsterdam, 1971).

[32] A. Gibbons, *Algorithmic Graph Theory* (Cambridge University Press, Cambridge, 1985).

[33] F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1969).

[34] J. Scott, *Social Network Analysis: a Handbook* (Sage, London,1991).

[35] N. Trinajstić, *Chemical Graph Theory* (CRC Press, Boca Raton, FL, 1992).

[36] R.J. Wilson, *Introduction to Graph Theory* (Longman, London, 1972).

[37] B. Rousseau and R. Rousseau, LOTKA: a program to fit a power law distribution to observed frequency data, *Cybermetrics* 4 (2000) 4. Available at: www.cindoc.csic.es/cybermetrics/articles/v4i1p4.html

[38] C.S. Wilson, Informetrics. *Annual Review of Information Science and Technology* 34 (1999) 107–247.

[39] D.J. de Solla Price, Networks of scientific papers. *Science* 149 (1965) 510–515.

[40] R. Garner, *A Computer Oriented, Graph Theoretic Analysis of Citation Index Structures* (Drexel University Press, Philadelphia, PA, 1967).

[41] A.-L. Barabási and R. Albert, Emergence of scaling in random networks. *Science* 286 (1999) 509–512.

[42] S. Milgram, The small world problem, *Psychology Today* 2 (1967) 60–67.

[43] M.E.J. Newman and D.J. Watts, Scaling and percolation in the small-world network model, *Physical Review E* 60 (1999) 7332–7342.

[44] M. Kochen (ed.), *The Small World* (Ablex, Norwood, NJ, 1989).

[45] L. Björneborn, Small-world linkage and co-linkage, *Proceedings of the 12th ACM Conference on Hypertext* (Aarhus, 2001).

[46] G. Pinski and F. Narin, Citation influences for journal aggregates of scientific publications: theory, with applications to the literature of physics, *Information Processing and Management* 12 (1976) 297–312.

[47] R.E. Nance, R.R. Korfhage and U.N. Bhat, Information networks: definitions and message transfer models, *Journal of the American Society for Information Science* 23 (1972) 237–247.

[48] W.M. Shaw Jr, Critical thresholds in co-citation graphs, *Journal of the American Society for Information Science* 36 (1985) 38–43.

[49] E. Logan and M.L. Pao, Analytic and empirical measures of key authors in schistosomiasis, *Proceedings of the American Society for Information Science* 27 (1990) 213–219.

[50] E. Logan and M.L. Pao, Identification of key authors in a collaborative network, *Proceedings of the American Society for Information Science* 28 (1991) 261–266.

[51] M.G. Martinsons, J.E. Everett and K. Chan, Mapping the scholarly development of strategic management, *Journal of Information Science* 27 (2001) 101–110.

[52] C. Haythornthwaite and B. Wellman, Work, friendship, and media use for information exchange in a networked organization, *Journal of the American Society for Information Science* 49 (1998) 1101–1114.

[53] H. Kretschmer, Patterns of behavior in coauthorship networks of invisible colleges. In: B.C. Peritz and L. Egghe (eds), *Proceedings of the Sixth Conference of the International Society for Scientometrics and Informetrics* (School of Library, Archive and Information Studies, Jerusalem, 1997), pp. 197–208.

[54] J. Bogaert, R. Rousseau and P. Van Hecke, Percolation as a model for informetric distributions: fragment size distribution characterised by Bradford curves. *Scientometrics* 47 (2000) 195–206.

Appendix

Table A1
Sociological Abstracts Classification Scheme (*sa* Classification codes), with details for the field of Complex Organization

0100 Methodology and research technology
0200 Sociology: history and theory
0300 Social psychology
0400 Group interactions
0500 Culture and social structure
0600 Complex organization
  0621 jobs, work organization, workplaces and unions
  0623 military sociology
  0624 bureaucratic structure/organizational sociology
  0665 social network analysis
  0671 sociology of business and entrepreneurism
  0674 voluntary organizations/philanthropy
0700 Social change and economic development
0800 Mass phenomena
0900 Political sociology/interactions
1000 Social differentiation
1100 Rural sociology and agriculture
1200 Urban sociology
1300 Sociology of language and the arts
1400 Sociology of education
1500 Sociology of religion
1600 Social control
1700 Sociology of science
1800 Demography and human biology
1900 The family and socialization
2000 Sociology of wealth and medicine
2100 Social problems and social welfare
2200 Sociology of knowledge
2300 Community/regional development
2400 Policy, planning, forecasting
2500 Radical sociology
2600 Environmental interactions
2700 Studies in poverty
2800 Studies in violence
2900 Feminist/gender studies
3000 Marxist sociology
3100 Clinical sociology
3200 Sociology of business
3300 Visual sociology

Table A2
First authorships in the SNA database

Author	Number of articles published as first author	Author	Number of articles published as first author
Wellman, Barry	21	Borgatti, Stephen P.	8
Burt, Ronald S.	17	Doreian, Patrick	8
Skvoretz, John	15	Knoke, David	7
Bonacich, Phillip	14		
Friedkin, Noah E.	13	8 authors	6
Everett, Martin G.	11	13 authors	5
Mizruchi, Mark S.	10	19 authors	4
Faust, Katherine	9	35 authors	3
Markovsky, Barry	9	132 authors	2
Marsden, Peter V.	9	870 authors	1

There were, moreover, 578 authors who were never first author.