Analysis and Visualization of Social Networks*

Dorothea Wagner

University of Karlsruhe
Department of Computer Science, Germany

Abstract. Social network analysis is a subdiscipline of the social sciences using graph-theoretic concepts to understand and explain social structure. We describe the main issues in social network analysis. General principles are laid out for visualizing network data in a way that conveys structural information relevant to specific research questions. Based on these innovative graph drawing techniques integrating the analysis and visualization of social networks are introduced.

1 Introduction

Social Network Analysis is a subdiscipline of the social sciences using graph-theoretic concepts to describe, understand and explain, sometimes even predict or design, social structure. It is focused on uncovering the patterning of people's interaction and based on the intuitive notion that these patterns are important features of the lives of the individuals who display them. Starting from social sciences the study of social networks became an interdisciplinary field. On one hand, it is guided by formal theory organized in mathematical terms, on the other hand grounded in the systematic analysis of empirical data. Network analysis has found important applications in organizational behavior, inter-organizational relations, the spread of contagious diseases, mental health, social support, the diffusion of information and animal social organization.

Since the 1980s, a yearly international conference on social network analysis, called *SUNBELT* is organized by the *International Network for Social Network Analysis, INSNA* [1]. A comprehensive, though non-visual, tool for social network analysis is UCINET [2]. For a comprehensive summary of social network analysis, its levels of analysis and its methodological tools see [24].

Also applications such as the analysis of Web graphs, bibliographic data, or financial data, often use similar or identical methods like in social network analysis. Recently, there is growing interest to understand the structure, dynamics and evolution of the Internet and WWW, and this way social network analysis has been rediscovered in other fields. Especially physicists in the complex systems community are interested in the statistical mechanics of complex networks. The very general questions in complex systems are how networks emerge, what

* The author gratefully acknowledges financial support from DFG under grant WA 654/13-1 and from the European Commission within FET Open Project COSIN (IST-2001-33555).

they look like, and how they evolve. This includes networks from such diverse areas as physics, biology, economics, ecology, and computer science. Modeling networks as dynamical systems, network morphogenesis and self-organization, as well as new graph theoretical aspects and network reconstruction from experimental data are considered, [3], [4] and [5]. It seems that because of this new emerging interest in networks at all graph theory and graph algorithms attract increasing attention from other sciences [18], [19].

In 1996, we began a cooperation with researchers from political science, aimed at providing the methodology of social network analysis with tailor-made means of automated visualization. Given the importance of visualizations for scientific development, it is astonishing how little attention the subject had received so far in the analysis of social networks. One of the rare exceptions is [21]. Even though a fair amount of software has been available to facilitate graphical editing, and even automatic layout of networks, the State of the Art that time seemed too heuristic to be satisfactory for supporting network analysis.

One of the first outcomes of our interdisciplinary cooperation was a survey of visualization methods in use at that time [11]. In that paper, general principles are laid out for visualizing network data in a way that conveys structural information relevant to specific research questions. These general principles resulted in innovative uses of graph drawing methods for social network visualization, and prototypical implementations thereof. With the growing demand for access to these methods, we started implementing an integrated tool for public use, the tool visone [15]. The main application area of visone is a methodological approach in the social sciences. Its usage is focused on graphs of small to medium size. As an alternative especially for large graphs, we recommend to try *Pajek* [8].

2 Social Networks

A *social network* consists of entities such as persons, organizations, or things, that are linked by binary relations such as social relations, dependencies, or exchange. These relations may be directed or undirected, weighted or unweighted, and weights, if present, may be interpreted as increasing or decreasing the tie between the two entities. Since data is often gathered by means of questionnaires, not even the existence of an edge is a sure thing. The two respondents corresponding to the end-vertices of a potential edge may have different perceptions regarding the presence of a specific type of tie between them. It is a long-standing debate whether unconfirmed edges should be included in an analysis. Typically, researchers decide to either treat unconfirmed edges like confirmed edges, or to exclude them completely. A crucial feature in many studies is the interrelation between the structure of a social network and the attributes of its elements.

We define a *social network* to be a labeled directed graph $G = (V, E = E_C \cup E_U; \delta, \omega)$, where E_C and E_U are disjoint sets of *confirmed* and *unconfirmed* edges, $\delta : E \rightarrow \mathbb{R}_{\geq 0}$ is a non-negative edge *length*, and $\omega : E \rightarrow \mathbb{R}_{\geq 0}$ a non-negative edge *strength*. A vertex or edge *attribute* is a (partial) function assigning *nominal* or *numerical* values to vertices or edges.

Although we cannot put any restrictions on the class of graphs, typical examples from social science projects are sparse, locally dense, and exhibit a small average distance between pairs of vertices.

3 Analysis

The purpose of *social network analysis* is to identify important vertices, crucial relationships, subgroups, roles, network characteristics, and so on, to answer substantive questions about structures. There are three main levels of interest: the element, group, and network level. On the element level, one is interested in properties (both absolute and relative) of single actors, links, or incidences. Examples for this type of analysis are bottleneck identification and structural ranking of network items. On the group level, one is interested in classifying the elements of a network and properties of subnetworks. Examples are actor equivalence classes and cluster identification. Finally, on the network level, one is interested in properties of the overall network such as connectivity or balance.

While we have an intuitive understanding what makes a vertex important or central, there is no universally accepted definition of importance. *Centrality of a vertex* may for example be measured according to the degree of that vertex, its distance to all other vertices or the number of shortest paths between two other vertices that contain the vertex itself. Similarly, there are different notions of importance or status in a directed graph. We refer to [15] for an unification and overview of such indices. Similarly, mathematical terms that capture to what extend networks tend to build clusters, like the *clustering coefficient*, or how networks evolve, like the *degree distribution*, are of interest [4]. Questions regarding the overall structure ask for example to what extend the network exhibits the *small-world phenomenon* [25].

Algorithmic aspects in network analysis concern the fast computation of such indices. Vertex indices are often easily computable in polynomial time. However, more efficient algorithms that are applicable also for large graphs as the fast algorithm for betweenness centrality presented in [9], are of increasing interest in this context.

4 Visualization

In *graph drawing* algorithms are designed that try to produce what is often termed an "aesthetic" visualization of a graph. In *network analysis* the demand that visualizations are not misleading is maybe even more important. Hence there are two obvious criteria for the quality of social network visualizations:

1. Is the information manifest in the network represented accurately?
2. Is this information conveyed efficiently?

With these criteria in mind, the following three aspects should be carefully thought through when creating network visualizations [11]:

– the *substantive aspect* the viewer is interested in,
– the *design* (i.e. the mapping of data to graphical variables), and
– the *algorithm* employed to realize the design (artifacts, efficiency, etc.).

Depending on the context, actors of high structural importance are interpreted as a being *central* or as having *high status*. With this substantive aspect in mind, we designed visualizations that represent vertex indices by constraining vertex positions to fixed distances from the center or from the bottom of the drawing, in either case depending linearly on the vertex index. These ideas have been further developed and applied in the following projects. We also refer to [15] for a more detailed description and figures illustrating the results.

Drug Policy. This project [20] studies the presence of HIV-preventive measures for IV-drug users in nine selected German municipalities. The substantive question underlying this research is, why municipalities with comparable problem pressure differ significantly in the provision of HIV-preventive measures such as methadone substitution or needle exchange. The policy networks under scrutiny comprise all local organizations directly or indirectly involved in the provision of such measures. The actors included in the study were queried about relations such as strategic collaboration, common activities, or informal communication with other organizations in the same municipality. None of the networks has more than 120 edges of the same type, and typically more than 50% of them are unconfirmed. In [12] a three-stage force-directed method for centrality layouts is presented, and in [6] a simple, purely combinatorial algorithm is developed.

Industry Privatization. The second study [23] deals with networks of public, societal and private organizations that developed during the privatization of industrial conglomerates in East Germany as part of the economic transformation after German unification in 1990. Their privatization is understood as political bargaining processes between actors that are connected by ties such as exchange of resources, command, or consideration of interest. The privatization was foreseen to be carried out by the Treuhandanstalt, a public agency of the federal government. Due to its institutional position and its ownership of all companies, it was generally assumed to be one of the most powerful actors in the transformation of East Germany. As part of the analysis, status indices are used as indicators for the power or influence of actors. In [14] a layered layout algorithm is outlined that visually supports status analyses of networks. A refinement of this algorithm uses the linear-time algorithm of [13] for coordinate assignment.

Topic Identification. Our third example illustrates the use of methods from social network analysis in another domain, namely topic identification in texts by centering resonance analysis [16]. The structure of texts is represented by graphs that have a vertex for each word occurring in a noun phrase and an edge for each pair of words that appear together in the same noun phrase or consecutively in the same sentence. It is argued that words corresponding to nodes with high betweenness centrality in such a graph are important for the structure of the text and thus a proxy for its topic. This method was applied to Reuters news dealing with the terrorist attacks of September 11, 2001 [7] to identify, among other things, the main topics, topic changes, side stories, etc. in the news. Centrality visualizations can then be used to show for example the main topics identified for the very first day of media coverage.

5 visone

The visone software [15] is implemented in C++ using LEDA, the *Library of Efficient Data Types and Algorithms* [22]. While the user interface is a customized version of LEDA's GraphWin class, all graph generation, analysis, and layout algorithms (except for LEDA's force-directed layout routine) have been implemented from scratch.

Starting with version 1.1, the main data format used in visone will be the XML sublanguage GraphML (Graph Markup Language) [10]. GraphML support is implemented in a LEDA extension package which will be made available for public use. It will hence be possible to administer project files with several social networks and any number of attributes. Data attributes can be mapped freely to graphical attributes like color, shape, and so on.

Acknowledgments

First of all, the author indebted to Ulrik Brandes who is a leading part of the project. The author thanks Sabine Cornelsen, Patrick Kenis, Jörg Raab, and Volker Schneider for many years of fruitful cooperation, the participants of POLNET summer schools for their feedback and suggestions, and Michael Baur, Marc Benkert, Marco Gaertler, Boris Köpf, and Jürgen Lerner for their implementation efforts in developing the software tool visone.

References

[1] International Network for Social Network Analysis, INSNA. See http://www.sfu.ca/ insna/

[2] Analytic Technologies. *UCINET V*. Network analysis software. See http://www.analytictech.com/

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

[4] A.-L. Barabási and R. Albert. Statistical mechanics of complex networks *Reviews of Modern Physics*, Vol. 74, January 2002, 47–97.

[5] A.-L. Barabasi. LINKED: The New Science of Networks. Perseus Books, 2002.

[6] M. Baur and U. Brandes. An improved heuristic for crossing minimization in circular layouts. Working Paper, 2003.

[7] V. Batagelj, U. Brandes, J. C. Johnson, S. Kobourov, L. Krempel, A. Mrvar, and D. Wagner. Analysis and visualization of network data. Special Session during *Sunbelt Social Network Conference XXII*, New Orleans, February 2002.

[8] Pajek. See http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

[9] U. Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.

[10] U. Brandes, M. Eiglsperger, I. Herman, M. Himsolt, and M.S. Marshall. GraphML progress report: Structural layer proposal. In: P. Mutzel, M. Jünger, and S. Leipert (eds.) *Proceedings 9th International Symposium on Graph Drawing (GD '01)*, Springer Lecture Notes in Computer Science 2265:501–512, 2002. For up-to-date information see http://graphml.graphdrawing.org/.

[11] U. Brandes, P. Kenis, J. Raab, V. Schneider, and D. Wagner. Explorations into the visualization of policy networks. *Journal of Theoretical Politics*, 11(1):75–106, 1999.

[12] U. Brandes, P. Kenis, and D. Wagner. Communicating centrality in policy network drawings. *IEEE Transactions on Visualization and Computer Graphics*, 9(2), 2003. To appear.

[13] U. Brandes and B. Köpf. Fast and simple horizontal coordinate assignment. In: P. Mutzel, M. Jünger, and S. Leipert (eds.) *Proceedings 9th International Symposium on Graph Drawing (GD '01)*, Springer Lecture Notes in Computer Science 2265:31–44, 2002.

[14] U. Brandes, J. Raab, and D. Wagner. Exploratory network visualization: Simultaneous display of actor status and connections. *Journal of Social Structure*, 2(4), 2001.

[15] U. Brandes and D. Wagner. visoneAnalysis and visualization of social networks. In: P. Mutzel and M. Jünger (eds.) *Special issue on Graph Drawing Software*, Mathematics and Visualization. To appear. For up-to-date information on visonesee http://www.visone.de/.

[16] S.R. Corman, T. Kuhn, R.D. McPhee, and K.J. Dooley. Studying complex discursive systems: Centering resonance analysis of communication. *Human Communication Research*, 28(2):157–206, 2002.

[17] L. Freeman Visualizing Social Networks. *Journal of Social Structure*, 1, 2000, (1).

[18] B. Hayes Computing Science: Graph Theory in Practice: Part I. *American Scientist*, Vol. 88, No. 1, January-February 2000, 9–13.

[19] B. Hayes Computing Science: Graph Theory in Practice: Part II. *American Scientist*, Vol. 88, No. 2, March-April 2000, 104–109.

[20] P. Kenis. An analysis of cooperation structures in local drug policy in germany, 1998. Unpublished Report.

[21] A.S. Klovdahl. A note on images of networks. *Social Networks* 3: 197–214.

[22] K. Mehlhorn and S. Näher. *The LEDA Platform of Combinatorial and Geometric Computing*. Cambridge University Press, 1999.

[23] J. Raab. *Steuerung von Privatisierung*. Westdeutscher Verlag, 2002.

[24] S. Wasserman and K. Faust. Social Network Analysis. Cambridge: Cambridge University Press. 1994.

[25] D.J. Watts and S.H. Strogatz. Collective dynamics of "small-world" networks. *Nature*, 393:440–442, 1998.

Author Index

Althaus, Ernst 1
Andreou, Maria I. 15

Bianchi, Leonora 209
Bilò, Vittorio 33

Cantone, Domenico 47
Caramia, Massimiliano 59
Carlier, Jacques 70
Clautiaux, François 70
Crescenzi, Pilu 81

Daneshmand, Siavash Vahdati ... 1
Díaz, Josep 97
Drake, Doratha E. 107

Faro, Simone 47
Flammini, Michele 33
Fleischer, Rudolf 120
Freivalds, Kārlis 134

Giordan, Stefano 59
Giovannelli, Roberto 33
Grossi, Roberto 81

Hall, Alexander 148
Hoefer, Martin 165
Hougardy, Stefan 107
Hsu, Wen-Lian 195

Iovanella, Antonio 59
Italiano, Giuseppe F. 81

Jiménez, Víctor M. 179

Kochol, Martin 191

Lu, Wei-Fu 195

Marzal, Andrés 179
Mastrolilli, Monaldo 209
Mehlhorn, Kurt 222
Meloni, Carlo 223
Moukrim, Aziz 70

Nikoletseas, Sotiris E. 15
Nègre, Stéphane 70

Pacciarelli, Dario 223
Petit, Jordi 97
Polzin, Tobias 1
Pranzo, Marco 223
Prasanna, Viktor K. 237

Serna, Maria 97
Singh, Mitali 237
Solis-Oba, Roberto 253
Spirakis, Paul G. 15

Täubig, Hanjo 148
Trippen, Gerhard 120

Wagner, Dorothea 261