

Summary

This article introduces automated network analysis and discusses its purpose. Algorithms for performing automated network analysis are also discussed, along with associated software and applications. This is relevant because many people want to answer questions about the high-level properties of networks such as who are the most active members or what cohesive subgroups of members exist within the network. Currently, network analysis requires users to understand what network analytic measures are available and how they should be used. The methods discussed here automate the process of network analysis so that more advanced applications can be built based on the information provided by network analyses.

Automating Network Analysis

Once a social network has been extracted from data sources such as **blogs, wikis, newsgroups, business networks, organizational networks**, etc., **network analysis software** such as UCINET, Pajek and NodeXL can be used. However, if one needs to perform a specific analysis such as finding the cohesive subgroups in a network, or finding the most important or relevant people in a network, the tools required for such tasks are generally not available. Which network analytic measures should be used to address different questions asked of social networks, and how can multiple measures be combined appropriately in asked complex or challenging questions about patterns of activity in social networks? Current network analysis tools provide calculation of particular network analysis measures such as measures of network centrality. For some measures and tools such as k-cores, the situation is further complicated by a need to choose appropriate parameter settings, leading to a time consuming interactive search space through the parameter space of interest. In addition, if one wants to perform network analysis for different snapshots of the network at different times, the steps in the analysis must be repeated manually and exported to other tools (e.g., spreadsheets, or statistical packages) in order to compare the values of the network analysis measures over time. Thus, it is beneficial to automate the network analysis process. Aside from making network analysis less tedious for users, automation also has the advantage of making the results of network analysis readily available to other software processes, so that network analysis can be a building block or component within other applications.

In the absence of more comprehensive network analysis packages, researchers and practitioners have automated network analysis for their own purposes on a case by case basis by creating their own scripts to run individual network analyses using scripting languages such as Perl, Python, R or MATLAB. However, this requires users to be technically savvy and have the ability to use these programming languages, and the overall process is wasteful since the same problems may be repeatedly solved by different researchers in different locations. Another problem is that some of the main network analysis software packages are not open source, so that it is not possible to call those features as functions from a new application. This means that if measures are to be

included in the automated network analysis, the algorithms underlying them to calculate the measures have to be re-implemented, resulting in unnecessary additional time and effort.

Algorithms for Automated Network Analysis

There are algorithms that can be used for automating network analysis. If one wants to find cohesive subgroups and relevant members, then the SCAN and DISSECT methods from Alvin Chin and Mark Chignell are recommended. They use a combination of network analysis techniques consisting of network centrality and hierarchical clustering to identify any cohesive subgroups that exist at different time periods, followed by similarity analysis of cohesive groupings between time periods to find the most cohesive subgroups in a social network over time. Once cohesive subgroups have been identified, their evolution can then be tracked over time in terms of the changing membership of subgroups as members enter and leave. For finding the subgroups and communities in a social network in real time, one can use community identification algorithms such as Newman's algorithm, density-based community detection algorithm, or other techniques reviewed by Alvin Chin and Mark Chignell. Network analysis measures such as indegree and outdegree can be used in recommending friends as in the LJMiner algorithm that recommends friends in LiveJournal.

Software Tools for Automated Network Analysis

While automated network analysis is in its infancy with few comprehensive but customizable tools available, some academic institutions, industry labs, and companies have created their own software to perform particular tasks which automate some aspects of network analysis.

One example is a tool developed by academics for visualizing online social networks and exploring the communities and connections (**network visualization**). Vizster, developed by Jeff Heer and danah boyd, uses spring-embedding algorithms from network analysis for visualizing nodes and their links, and allows users to use a slider to highlight the different types of communities, where the communities are found based on Newman's community identification algorithm. Community Detector is a software tool for mining and visualizing the evolution of subgroups that does not abstract the network analysis details of clustering, allowing the subgroups that form over different periods in time to be visible. Network analysis can also be used to determine the interactions and subgroups of genes as in **biological networks** such as NetBox. As an example of its use, NetBox was used to perform automated network analysis of a human gene network and visualize partitions in the network using the Newman-Girvan algorithm.

Besides academic software, there are also commercial products that can be used. One commercial product for performing automated network analysis is Sentinel Visualizer which can generate a wide range of centrality values. A somewhat similar package is UCINET and Pajek. Social network analysis can also be used for finding expertise in a company such as IBM's SmallBlue using clustering algorithms and analysis of degree proximity.

However, the above software may not be adequate for all purposes and some types of

automated social network analysis require programming languages and open-source software for obtaining the network analysis results. For calculating network analysis measures and visualizing social networks, NodeXL, a template plug-in for Microsoft Excel 2007, and JUNG (Java Universal Network Graph) can be used. Other useful software tools can be found listed in the Wikipedia entry for “Social network analysis software”. Network analysis can be supported by follow on statistical analyses (used for instance to find which trends are significant) obtained from open-source statistics software called R as well as using proprietary packages such as SPSS and SAS. Even though automated network analysis is not yet widely used, further research and development will help automated network analysis to become a customized and rigorous method for automatically analyzing social networks and retrieving the desired results.

Applications of automated network analysis

A wide range of applications for the tools discussed in this article are envisaged. These include community tracking tools, community-based search engines, and community based recommender systems. With the rapid growth in social computing it seems likely that automated network analysis will be an important tool for online applications potentially rivaling other major tools such as search engines.

Further Reading

- Borgatti, S. P., Everett, M. G., and Freeman, L. C. *Ucinet for Windows: Software for Social Network Analysis*. Analytic Technologies, Harvard, USA, 2002.
- Cerami, E., Demir, E., Schultz, N., Taylor, B.S., and Sander, C. *Automated Network Analysis Identifies Core Pathways in Glioblastoma*. PLoS ONE 5(2): e8918. doi:10.1371/journal.pone, 2010, 0008918.
- Chin, A. and Chignell, M. *DISSECT: Data-Intensive Socially Similar Evolving Community Tracker*, Computational Social Network Analysis: Trends, Tools and Research Advances, Series: Computer Communications and Networks, Abraham, Ajith; Hassanien, Aboul-Ella; Snášel, Václav (Eds.), 2010, 81-105.
- Chin, A., and Chignell, M. *Automatic detection of cohesive subgroups within social hypertext: A heuristic approach*. New Review of Hypermedia and Multimedia 14, 1 (2008), 121–143.
- de Nooy, W., Mrvar, A., and Batagelj, V. *Exploratory Social Network Analysis with Pajek*. Cambridge University Press, New York, USA, 2005.
- Falkowski, T., Bartelheimer, J., and Spiliopoulou, M. 2006. *Mining and Visualizing the Evolution of Subgroups in Social Networks*. In Proceedings of the 2006 IEEE/WIC/ACM international Conference on Web intelligence (December 18 - 22, 2006). Web Intelligence. IEEE Computer Society, Washington, DC, 52-58.
- Falkowski, T., Barth, A., and Spiliopoulou, M. 2007. *DENGRAP: A Density-based Community Detection Algorithm*. In Proceedings of the IEEE/WIC/ACM international Conference on Web intelligence (November 02 - 05, 2007). Web Intelligence. IEEE Computer Society, Washington, DC, 112-115.
- Heer, J., boyd, d. *Vizster: Visualizing online social networks*. In Proceedings of the 2005 IEEE Symposium on information visualization, 2005, 33–40.

Hsu, W. S., King, A. L., Paradesi, M. S. R., Pydimarri, T. and Weninger, T. *Collaborative and Structural Recommendation of Friends using Weblog-based Social Network Analysis* (2006). In AAAI Spring Symposia 2006 on Computational Approaches to Analysing Weblogs.

Lin, C. Y., Cao, N., Liu, S. X., Papadimitriou, S., Sun, J., and Yan, X. *SmallBlue: Social Network Analysis for Expertise Search and Collective Intelligence*, Data Engineering. In Proceedings of IEEE 25th International Conference on Data Engineering (ICDE '09), March 29 2009-April 2 2009, 1483-1486.

Newman, M.E.J., and Girvan, M. *Finding and evaluating community structure in networks*. Phys Rev E Stat Nonlin Soft Matter Phys 69, 2004, 026113.

Newman, M.E.J. *Fast algorithm for detecting community structure in networks*. Physical Review E, 69, 2004, 066133.

Alvin Chin, Mobile Social Networking Group, Nokia Research Center, Beijing, China

Mark Chignell, Department of Mechanical and Industrial Engineering, University of Toronto, Canada