

Analyzing Social Media Networks with NodeXL

Derek Hansen¹, Cody Dunne², Ben Shneiderman²

¹iSchool, ²Department of Computer Science

Contact: dlhansen@umd; {cdunne, ben}@cs.umd.edu

<http://nodexl.codeplex.com/>

Social media services, such as Facebook, Twitter, and Wikis have enabled new forms of collaboration and interaction in nearly every imaginable human endeavor. And we have only begun to realize the potential of technology-mediated social interaction. Despite numerous success stories, we must remember the countless failures due to social and technical factors. How can we support practitioners in their efforts to cultivate meaningful and sustainable online interaction?

One promising strategy is to provide tools and concepts that help practitioners make sense of social media data. There is precedence to this approach in the development of sophisticated, yet fairly intuitive website analytics tools such as Google Analytics. These tools help non-programmers understand website traffic patterns so they can make more informed design decisions. We envision an equivalent set of social analytics tools to help social media analysts and community administrators make better decisions based on their in-depth understanding of social participation and relationships.

Social network analysis (SNA) provides a set of concepts and techniques for making sense of social data through quantifiable metrics and network visualizations. These complement basic metrics of social participations used in current tools (e.g., number of posts; membership duration) and reveal the patterns in the network that result from social interactions. SNA concepts provide an effective vocabulary to characterize important relational properties of network members, as well as entire network structures. However, SNA also adds significant complexity and imposes obstacles for analysts that lack technical skills. Tools such as Pajek and UCINET have made SNA possible for those with sufficient drive and technical know-how, such as intelligence analysts, computer scientists, and social science doctoral students.

With the prevalence of social media network data, there is a great opportunity to make the powerful concepts and techniques of SNA accessible to a much wider audience of community analysts, participants, and designers. Doing so is hardly trivial, leading to the need to address the pressing research question: How can the complex, sophisticated set of SNA techniques be supported in an intuitive manner? To address this question, we have been working as part of a

team of researchers funded by Microsoft Research to develop NodeXL, an open source add-in for the widely used spreadsheet application Excel 2007 (see <http://nodexl.codeplex.com>). It provides a range of basic network analysis and visualization features [1], that we have refined over time based on user studies [2]. NodeXL uses a highly structured workbook template that includes multiple worksheets to store all the information needed to represent a network graph. Network relationships (i.e., graph edges) are represented as an “edge list”, which contains all pairs of entities that are connected in the network. Complementary worksheets contain information about each vertex and cluster. Data importers allow users to grab networks and user data from popular social media networks such as Twitter, YouTube, Flickr, and email. Visualization features allow users to display a range of network graph representations and map data attributes to visual properties including shape, color, size, transparency, and location (Figure 1).

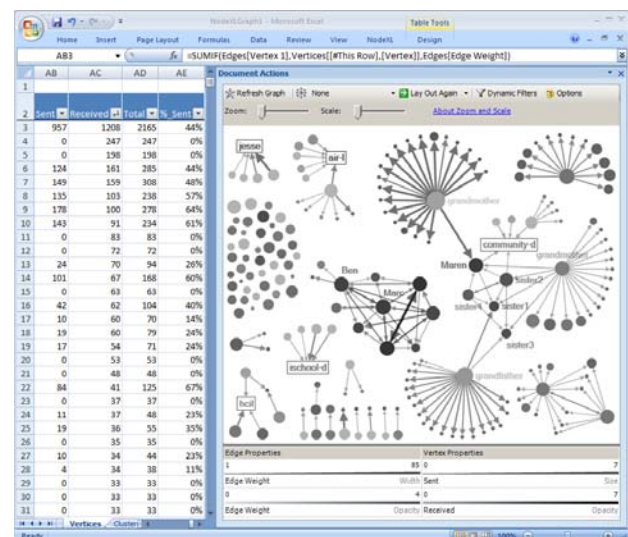


Figure 1 – Derek’s personal email network shown in NodeXL, with data represented in the spreadsheet (left) and a network graph (right). Network and social metrics are mapped onto visual attributes including size, opacity, and edge thickness to highlight important people and relationships. Nodes are positioned to identify clusters such as the NodeXL team shown in the middle.

While there is a need to improve the functionality, efficiency, scalability, and usability of NodeXL, there is also a need to understand how non-technical experts can use NodeXL (and related tools) to understand community interaction. Over the past 2 years we have been studying the process through which non-technical practitioners and students can use NodeXL to make sense of online community data [3] (see Figure 2). Our experience has shown that with minimal training, non-technical students can use NodeXL to generate meaningful graphs and insights from online communities. The tight integration of visualization and data proved key to teaching, understanding, and applying network concepts for beginners [2,3]. NodeXL is now used to teach SNA in dozens of classes around the globe. The ability to automatically capture data was essential for non-programmers, as were the layout algorithms. Our studies have led to important improvements in the tool such as the inclusion of a legend and “binning” of isolate edges, as well as identify additional priorities for network analysis tools such as improved layout algorithms, support for grouping, and working with multi-modal data.

We have also been developing usage scenarios that highlight how SNA can be applied to a variety of social media networks in order to derive actionable insights. This work stands in contrast to the majority of SNA work by computer scientists and computational social scientists who characterize the mathematical properties of social media networks, but fail to speak to community administrators trying to gain practical insights. Many of these insights are captured in a forthcoming Morgan-Kaufmann book titled “Analyzing Social Media Networks with NodeXL: Insights from a Connected World” by Derek Hansen, Ben Shneiderman, and Marc Smith. The book introduces social media networks, social network analysis, and NodeXL to those unfamiliar with these concepts and then applies them to a number of social media including email, discussion forums, Twitter, YouTube, Flickr, wikis, and websites.

This work has been funded by Microsoft Research, with significant contributions from Natasa Milic-Frayling (Microsoft Research Cambridge), Marc Smith (Connected Action Consulting Group), Tony Capone (Microsoft Research), Eduarda Mendes Rodrigues (University of Porto), and Jure Leskovec (Stanford University). University of Maryland HCIL contributors include Ben Shneiderman, Derek Hansen, Cody Dunne, Dana Rotman, Elizabeth Bonsignore, Udayan Kourana, and Puneet Sharma.

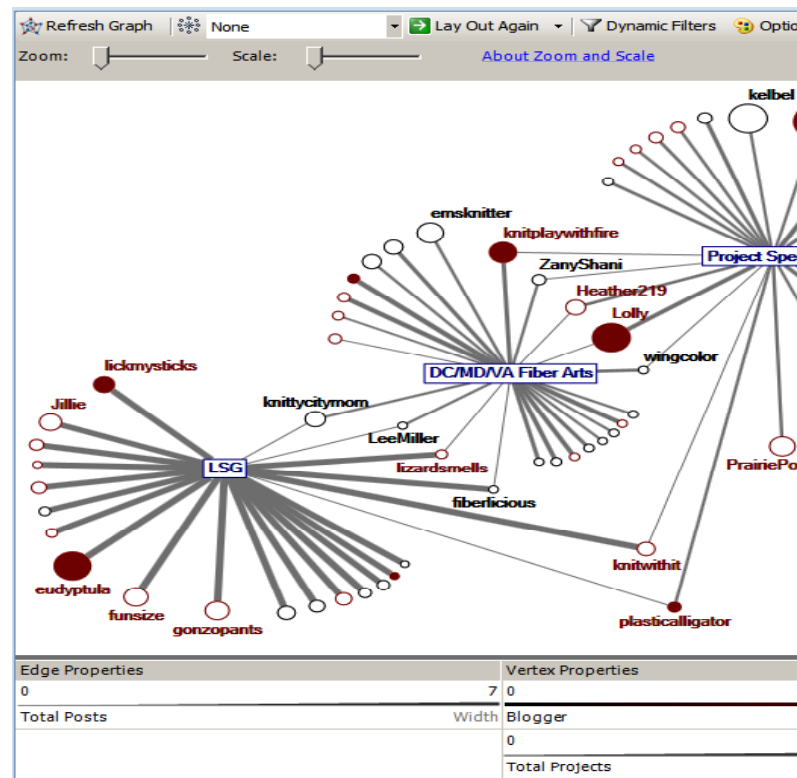


Figure 2 – A network created by Rachel Collins, a student new to SNA in Derek Hansen’s Communities of Practice class.

Bimodal network connecting three Ravelry groups (i.e., forums) represented as blue text boxes to contributors represented as circles. Edge width is based on number of posts (with logarithmic mapping). Vertex size is based on number of completed Ravelry projects. Maroon vertices have a blog and solid circles are either community moderators or volunteer editors. The network helps identify important boundary spanners (e.g., those connected to multiple groups) as well as compare groups.

PAPERS

1. Smith, M., Shneiderman, B., Milic-Frayling, N., Rodrigues, E. M., Barash, V., Dunne, C., et al. Analyzing Social (Media) Network Data with NodeXL. Forthcoming In *Proc. C&T 2009*.
2. Bonsignore, E.M., Dunne, C., Rotman, D., Smith, M., Capone, T., Hansen, D.L. & Shneiderman, B. (2009), "First steps to NetViz Nirvana: evaluating social network analysis with NodeXL", In *SIN '09: Proc. international symposium on Social Intelligence and Networking*. IEEE Computer Society Press.
3. Hansen, D., Rotman, D., Bonsignore, E., Milic-Frayling, N., Rodrigues, E., Smith, M., Shneiderman, B. (2009), Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Data. In *University of Maryland Tech Report: HCIL-2009-17*.