From *SIAM News*, Volume 44, Number 1, January/February 2011

People Who Read this Article Also Read . . . : Part I

By Desmond J. Higham, Peter Grindrod, and Ernesto Estrada

Many key ideas in the booming field of network science can be traced to the 1950s and 1960s, when researchers began to formalize the study of social interactions. The language and tools of mathematics, especially graph theory and linear algebra, offered a natural framework, and soon mathematicians were viewing social network analysis as a viable application area. By 1984, the social scientist Linton Freeman [7] observed that "There's a whole lot of really high powered math types running around in the social networks arena."

Fast forward to 2008; Freeman [8] now argued that

"When ideas and tools move from one field to another, the movement is generally from the natural to the social sciences. In recent years, however, there has been a major movement in the opposite direction. The idea of centrality and the tools for its measurement were originally developed in the social science field of social network analysis. But currently the concept and tools of centrality are being used widely in physics and biology."

Whereas social scientists had painstakingly collected data in the field to build up networks link by link (see, for example the UCINET IV collection at http://vlado.fmf.uni-lj.si/pub/networks/data/ucinet/ucidata.htm), advances in experimental techniques and computing power opened the possibility of studying and visualizing large-scale networks in nature and technology. The landmark small-world paper of Watts and Strogatz [17] raised the profile of network science outside the social science community; according to Freeman [8], "In the five years between 1998 and 2003 physicists turned out more publications on the subject than members of the social network community had produced over a period of 45 years."

Networks were being "discovered" everywhere, and analysed with tools derived in the social sciences. Links could represent, for instance, scholarly co-authorships, Hollywood co-starring roles, WWW hyperlinks, Internet connections, linguistic similarities, electronic circuitry, online co-purchases, food web connections, physical protein interactions, coordinated gene expression, metabolic regulation, amino acid residue similarities, transportation channels, electric power cables, and common street location. In-depth discussions of many examples can be found in [4].

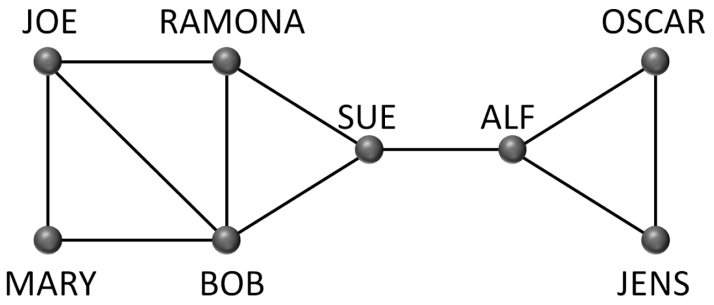

Figure 1. *A simple friendship network.*

The concept of *centrality*, first developed to identify important actors in various types of human social interaction networks [8], has now been applied extensively across almost every conceivable network scenario. Consider the simple network shown in Figure 1, in which each undirected edge records a friendship between a pair of individuals. Who is the best person to invite to a "bring-a-bottle-and-some-friends" party? Who would start a rumour most effectively? Who is most likely to have heard the latest rumour? We can address such questions by assigning a centrality measure to each node. Some of the main path-based measures are defined precisely in Box A. More comprehensive treatments can be found in, for example, [13, 16]. Loosely:

■ *Degree centrality* simply records the degree (number of incident edges) of each node. High-degree nodes are good for the "bring-your-friends" events.

■ *Closeness centrality* records the reciprocal of the sum of the shortest path lengths between a node and all other nodes in the network. Nodes that score high are smart places to start a rumour.

■ *Betweenness centrality* records the propensity of a node to be involved in shortest paths. Nodes with high betweenness are more likely to learn the latest rumour.

Table 1 shows results for the simple network of Figure 1. We see that although Bob has the highest degree, Sue is at the top for closeness and betweenness. This is intuitively reasonable, as Sue appears to occupy a more central position in the network. Similarly, Alf, who ranks below Bob in terms of degree and closeness, jumps into a very clear second place in terms of betweenness; to form a path to Oscar or Jens, the others have no choice but to go through Alf.

Box A: Centrality Measures

The *degree centrality* of node i, denoted k_i, is simply the degree of that node—the number of incident edges.

The *distance* between nodes i and j, denoted d_{ij}, is the length of the shortest path connecting them.

The *distance sum* of node i, denoted s_i, is defined as $s_i = \sum_{j=1}^{N} d_{ij}$.

The *closeness centrality* of node i, denoted CC_i, is the normalized reciprocal of the distance sum; $CC_i = (N-1)/s_i$.

The *betweenness centrality* of node i, denoted BC_i, has the form

$$BC_i = \sum_{r \neq s \neq i} \frac{\rho_{rs}(i)}{\rho_{rs}},$$

where ρ_{rs} is the total number of shortest paths connecting nodes r and s and $\rho_{rs}(i)$ is the number of such shortest paths that pass through i.

These centrality measures are based on the concept of the shortest paths between pairs of nodes. Considering more general routes through the network leads to a more relaxed view. As discussed by Borgatti [3], depending on the type of process being considered, it may be appropriate to account for:

■ *Trails*, where nodes can be revisited during the excursion, but edges cannot be reused. In Figure 1, for example, Mary–Bob–Sue–Ramona–Bob–Joe is a trail but not a path. A piece of gossip typically spreads along a trail. Bob might hear the gossip from both Mary and Ramona. But if Bob did hear it from Mary, it is unlikely that he and Mary would repeat the gossip back to each other.

Node i	Degree k_i	Closeness CC_i	Betweenness BC_i
Joe	3	0.4375	0.5
Mary	2	0.4118	0
Ramona	3	0.5385	2.0
Bob	4	0.5833	6.5
Sue	3	0.6364	12.0
Alf	3	0.5385	10.0
Oscar	2	0.3889	0
Jens	2	0.3889	0

Table 1. *Centrality measures for the network shown in Figure 1.*

■ *Walks*, which allow the use of both nodes and edges more than once. For example, in Figure 1, Mary–Bob–Sue–Ramona–Bob–Mary–Joe is a walk, but not a trail or a path. A particular dollar bill flows through the person–person network along walks. Mary might give the bill to Bob in one transaction, and Bob might return it to Mary in another.

The walk scenario is attractive from a linear algebra point of view. Suppose that an undirected network has adjacency matrix A, where $a_{ij} = 1$ if nodes i and j are connected and 0 otherwise. The basic identity

$$\left(A^n\right)_{ij} = \sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\cdots\sum_{k_{n-1}=1}^{n} a_{i,k_1}a_{k_1,k_2}\ldots a_{k_{n-1},j}$$

then shows that $(A^n)_{ij}$ counts the total number of walks of length n from node i to node j. To summarize this information with a single number, we could take a weighted sum over all walk lengths. A very influential 1953 paper [11,13] suggested that the number of walks of length n could be scaled by a factor a^n, for some suitably chosen parameter $0 < a < 1$. We could then construct a closeness-style centrality score for node i by summing over all j the weighted total number of walks between i and j. Using the expansion $(I + aA)^{-1} = I + aA + a^2A^2 + a^3A^3 + \ldots$, we obtain a centrality score of

$$\sum_{j\neq i}\left((I+aA)^{-1}\right)_{ij}$$

for node i. Intuitively, summarizing over walks of all lengths, rather than using the all-or-nothing shortest-path convention, should make the measures less sensitive to spurious or missing information. This type of uncertainty is inherent in most data sets. In the case of "who e-mailed whom" information, for example, false negatives might arise because we are overlooking other forms of communication; false positives can arise when, say, a manager's e-mails are heavily filtered by a personal assistant.

In Box B we show how walk-based centrality measures can be constructed, using the general approach of [5].

In part II of this article we will discuss the recent boom in social network analysis that is being driven by the desire of businesses and governments to exploit the tell-tale fingerprints of our digital behavior. We will also show that the dynamic nature of this data, for example, the time-stamp that accompanies e-mails and cell phone communications, throws up fascinating challenges for the applied mathematician.

Box B: Walk-based Centrality

Suppose that we have an appropriate sequence of non-negative real numbers $\{c_n\}_{n\geq 1}$, where c_n is the scaling factor that we intend to apply to the count for walks of length n. Defining the function $f(x)$ through the Maclaurin series $\sum_{n\geq 1} c_n x^n$, we can define the following measures in terms of the adjacency matrix A; see [5] for more details.

The f-centrality of node i is given by $f(A)_{ii}$.

The f-communicability between nodes i and j is given by $f(A)_{ij}$.

The f-betweenness of node i is given by

$$\frac{1}{(N-1)^2-(N-1)}\sum\sum_{p\neq q,p\neq i,q\neq i}\frac{f(A)_{pq}-f(A-E(i))_{pq}}{f(A)_{pq}},$$

where the matrix $E(i)$ has nonzeros only in row and column i, and its row and column i have 1 wherever A has 1. $(A - E(i))$ is thus the adjacency matrix for the network that arises when we remove all edges involving node i.)

The particular case of $c_n = 1/n!$ was introduced and studied in [6], and the corresponding f-centrality for $f(x) = e^x$ has come to be known as the *Estrada index*.

Acknowledgments

The authors are supported by the UK Engineering and Physical Sciences Research Council Mathematical Sciences programme and the Research Council UK Digital Economy programme.

References

[1] E. Acar, D.M. Dunlavy, and T.G. Kolda, *Link prediction on evolving data using matrix and tensor factorizations*, in LDMTA'09: Proceedings of the ICDM'09 Workshop on Large Scale Data Mining Theory and Applications, IEEE Computer Society Press, December 2009, 262–269.

[2] J. Bohannon, *Counterterrorism's new tool: 'Metanetwork' analysis*, Science, 325 (2009), 409–411.

[3] S.P. Borgatti, *Centrality and network flow*, Social Networks, 27 (2005), 55–71.

[4] E. Estrada, M. Fox, D.J. Higham, and G.-L. Oppo, eds., *Network Science: Complexity in Nature and Technology*, Springer, Berlin, 2010.

[5] E. Estrada and D.J. Higham, *Network properties revealed through matrix functions*, SIAM Rev., 52 (2010), 696–714.

[6] E. Estrada and J.A. Rodríguez-Velázquez, *Subgraph centrality in complex networks*, Phys. Rev. E, 71 (2005), 056103.

[7] L.C. Freeman, *Turning a profit from mathematics: The case of social networks*, J. Math. Sociol., 10 (1984), 343–360.

[8] L.C. Freeman, *Going the wrong way down a one-way street: Centrality in physics and biology*, J. Social Structure, 9 (2008).

[9] P. Grindrod and D.J. Higham, *Evolving graphs: Dynamical models, inverse problems and propagation*, Proc. Roy. Soc., Series A, 466 (2010), 753–770.

[10] P. Grindrod, D.J. Higham, M.C. Parsons, and E. Estrada, *Communicability across evolving networks*, Mathematics and Statistics Research Report 32, University of Strathclyde, 2010.

[11] L. Katz, *A new index derived from sociometric data analysis*, Psychometrika, 18 (1953), 39–43.

[12] P.J. Mucha, T. Richardson, K. Macon, M.A. Porter, and J.-P. Onnela, *Community structure in time-dependent, multiscale, and multiplex networks*, Science, 328 (2010), 876–878.

[13] M.E.J. Newman, *Networks: An Introduction*, Oxford University Press, Oxford, UK, 2010.

[14] J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia, *Analysing information flows and key mediators through temporal centrality metrics*, in SNS 2010: Proceedings of the 3rd Workshop on Social Network Systems, New York, NY, 2010, ACM, 1–6.

[15] Technology Quarterly, *Untangling the social web. Software: From retailing to counterterrorism, the ability to analyse social connections is proving increasingly useful*, The Economist, September (2010).

[16] S. Wassermann and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, UK, 1994.

[17] D.J. Watts and S.H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature, 393 (1998), 440–442.

Desmond J. Higham is a professor in the Department of Mathematics and Statistics, University of Strathclyde, UK. Peter Grindrod is a professor of mathematics in the Department of Mathematics and Statistics at the University of Reading, UK. Ernesto Estrada is a professor in the Departments of Mathematics and Statistics, and Physics, at the University of Strathclyde, UK.