Social Network Analysis for Virtual Communities

Francisco Ignacio Revuelta Domínguez
University Of Salamanca, Spain

SOCIAL NETWORK ANALYSIS FOR VIRTUAL COMMUNITIES

The social network analysis (SNA) method is a set of methods considered these days as a paradigm or a method that offers "the mapping and measuring of relationships and flows between people, groups, organizations, animals, computers or other information/knowledge processing entities. The nodes in the network are the people and groups while the links show relationships or flows between the nodes. SNA provides both a visual and a mathematical analysis of human relationships. Management consultants use this methodology with their business clients and call it organizational network analysis (ONA)." (Krebs, 2005)

As Krebs says,

a method to understand networks and their participants is to evaluate the location of actors in the network. Measuring the network location is finding the centrality of a node. These measures help determine the importance, or prominence, of a node in the network. Network location can be different than location in the hierarchy, or organizational chart.

SNA uses several concepts to evaluate different network properties such as centrality, connectivity, cliques, and so on, each of which pertain to specific dimension of the network.

SNA allow us to measure the three most popular individual network parameters: degree centrality, betweenness centrality, and closeness centrality.

Centrality Measures in Social Network Analysis

At this point we would like to define a number of descriptive measures, each of them, in a global sense, help us understand the actors' situation in a social network. It is not necessary to know the mathematical formulas to calculate these measures. Computer programs have the formulas built in. Should the reader be interested in the formulas (Wasserman & Faust, 1994) many papers and SNA handbooks describe them. We offer an approach.

Density

Density is a measure of the connectedness in a network. "Density is defined as the actual number of ties in a network, expressed as a proportion of the maximum possible number of ties" (Coulon, 2005). Density is a number between 0 and 1. When density is close to 1.0, the network is said to be dense.

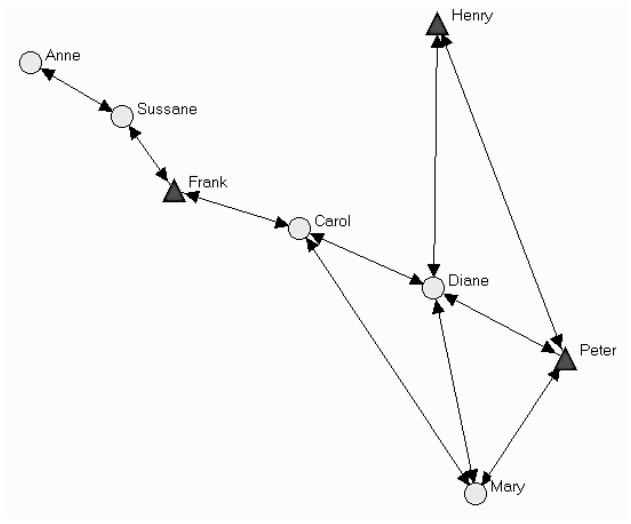
$$D = \frac{2L}{n(n-1)}$$

L = *number of ties*
n = *number of nodes*

Figure 1. Symmetric table of data interactions in example network

	Peter	Mary	Diane	Carol	Henry	Frank	Susanne	Anne
Peter	0	1	1	0	1	0	0	0
Mary	1	0	1	1	0	0	0	0
Diane	1	1	0	1	1	0	0	0
Carol	0	1	1	0	1	1	0	0
Henry	1	0	1	1	0	0	0	0
Frank	0	0	0	1	0	0	1	0
Sussane	0	0	0	0	0	1	0	1
Anne	0	0	0	0	0	0	1	0

Figure 2. Map of example network. Colors and shapes indicate the actors' gender in a social network



Degree Centrality

Social network researchers measure network activity for a node by using the concept of degrees—the number of direct connections a node has. In the example network, Diane has the most direct connections in the network, making hers the most active node in the network. She is a 'connector' or 'hub' in this network. Diane (or node) is denominated 'star.'

$$C_D(n_i) = d(n_i) = \sum_j x_{ij} = \sum_j x_{ji}.$$

$d(n_i)$ = number of connected nodes

Standard mode:

$$C_d(n_i) = \frac{d(n_i)}{g-1}$$

g = number of nodes in network (size)

Betweenness Centrality

While Diane has many direct ties, Carol has few direct connections—fewer than the average in the network. Yet, in many ways, she has one of the best locations in the network—she is between two important constituencies. She plays a 'broker' role in the network. The good news is that she plays a powerful role in the network, the bad news is that she is a single point of failure. Without her, Frank, Susanne and Anne would be cut off from information and knowledge in Diane's cluster. A node with high betweenness has great influence over what flows in the network.

$$C_B(n_i) = \sum_{j<k} g_{jk}(n_i)/g_{jk}$$

g_{jk} = number of geodesic (shorter ways) connect j and k

$g_{jk}(n_i)$ = number of geodesic between j and k and go across i

Standard mode:

$$C'_B(n_i) = \frac{C_b(n_i)}{\left[(g-1)(g-2)/2\right]}$$

Closeness Centrality

Carol has fewer connections than Diane, yet the pattern of their direct and indirect ties allow them to access all the nodes in the network more quickly than anyone else. They have the shortest paths to all others—they are close to everyone else. She is in an excellent position to monitor the information flow in the network—she has the best visibility into what is happening in the network.

$$C_c(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j)\right]^{-1}$$

$d(n_i)$= geodesic distance between n_i and n_j

It is a farness measure (more distance smaller centrality).

Krebs describes the most recent applications of SNA as the following:

- Determine influential journalists and analysts in the IT industry
- Unmask the spread of HIV in a prison system
- Map executive's personal network based on e-mail flows
- Discover the network of Innovators in a regional economy
- Analyze book selling patterns to position a new book