What else do we want to know about social networks?

MAUREEN O'MARA AND SALLY SCHULZ
National Geospatial-Intelligence Agency
Email: eda_kdd_etl@yahoo.com and schulzs@nga.mil

What is common to all networks and what is unique to particular network types? Are we looking at the sorts of networks that matter? Can we answer questions about membership dynamics in significant types of networks?

Humans live in relation to each other. Sometimes they live in strong relationship, sometimes in weak relationship. Families, friendships, business partnerships, transnational organizations, political parties, clans and tribes, swim teams, charities, drug cartels, and organized crime are all forms of human relatedness, ensuring that the members of these networks are more strongly tied to each other in some time and place than to other networks at the same time and place.

Social network analysis should facilitate understanding of variables governing where and when members enter each network, how, when, and where a network expands and contracts, how relationships alter dynamically within a network, how multiple networks might share members, which network has the strongest claim to member loyalty at any given time, and how networks start and end. Given the number of relationships most people have, an understanding is needed about when a repetitive pattern (buying coffee at the same Starbucks on Thursdays) reaches the level of a relationship in a network. People are usually proximal to multiple networks, but are unlikely to consider themselves in network with all of them. What level of connectedness/role/function is needed before a person is part of a network? What types of boundaries are recognized by people that cause them to see themselves as members of one network, but not another-even if they contact it frequently? (Is an older sister a member of a crime gang if a younger brother who takes her to lunch occasionally is a member?)

With the development of online social media sites and the accessibility of digital data associated with these sites, methods and concepts for visualizing, measuring, and describing network strength and organization have blossomed. A large number of metrics and strategies for characterizing online social media networks have been developed, usually extending graph theory and its focus on links and nodes to individuals who come in contact with each other. This methodology needs to be enhanced to answer questions about the dynamic nature of network creation and membership.

Key to creating a general approach to networks is a better understanding of the varied types of relationship within networks (Wilson et al, 2009). Social network research needs to expand its focus to a range of types of networks and begin to compare and contrast the nature of relationships and roles within each of network. Analyzing the spatial and temporal aspects of networks can provide insight into the patterns of network membership, roles, and involvement not apparent from link and node analysis.

Potential Spatial constraints on Networks: Are networks constrained by space?

While the Internet has changed the impact of physical distance, proximity is well accepted throughout the social sciences as having an effect on strength of relationship. A considerable body of information exists that suggests that the closer individuals are to each other spatially, the more likely they will be close in a network (Scellato et al, 2010). An underlying geographic tenet is that near things are more related to each other than far things (Tobler, 1970). Unfortunately, social network analyses are seldom spatially enabled, so the potential of spatial relationship to add information about strength of relationship is unexplored. Part of the challenge is finding social network datasets that are spatially tagged. However, even with appropriate datasets, a number of complex challenges exist in adding spatial information to social networks. One of the problems is scale (Olson and Carley, 2010, Goodchild and Gopal, 2005). Selecting an appropriate geographic scale for a network may involve smaller aggregate areas for some nodes and relationships and larger for others. An exploratory spatial data analysis methodology for accurate selection of spatial scale for network analysis needs to be developed. Once spatial scale is selected, it is possible to add existing topographic or other spatial data to the node and link network representation to begin to explore aspects of the network. It becomes possible to ask such questions as whether networks with equal numbers of members and large spatial extent function the same as networks with smaller spatial extent. It also becomes possible to understand whether there is a relationship between type/effectiveness of network and place or type of member and place.

Although network analysis looks at all links between nodes, spatial analysis might look extensively at the spatial pattern of each node's closest links. Early work examining individual spatial distributions suggest that some nodes have circular spatial network distributions while others are linear. In some cases these distributions conform to elevation, road, and telecommunications features, but in other cases there appear to be additional variables determining individual node network patterns. This type of effect is only apparent when social networks are spatially enabled and are enhanced with additional geospatial data. Other challenges emerge when considering the display of networks with extensive near and far nodes since the scale of the nodes may be different than the spatial extent that is spanned. Additional challenges include understanding how to evaluate networks when some percentage of the attached nodes are spatially enabled and others are not. When is a form of interpolation permissible and when should data be treated as spatially unknown? How should spatially enabled networks with untagged nodes be displayed to allow accurate analytic interpretation? It is possible that some of the existing social media datasets using IP addresses as spatial analogues or actual declared locations might be used to create test sets to resolve some of these issues. For example, removing a percentage of the spatial information from a tagged network might provide additional understanding of the potential error involved in varied strategies of network display and analysis when only a proportion of the network is spatially enabled. It is also possible that additional spatial information such as

Census data might be used to increase insight into network relationships, network formation, and network distribution/development.

Temporal Challenges: What dynamic patterns are unique to particular networks and what patterns are repeated in all social networks?

Social networks are temporally dynamic, but most often treated as though they are static. Members join, participate, and leave and affect the shape and strength of the network and its relationships. Adding temporal analysis to social network analysis is likely to provide additional information about the strength of relationships and the importance of nodes to a network. Questions about the growth of networks, length of time to become a significant network node, lifespan of varied network types, and other network dynamics might be framed and answered differently if temporal aggregation into appropriate discrete units accompanies network analysis. However the same issue of scale is present with temporal tagging. Issues of how often a network should be sampled and how this is determined are currently unresolved. Understanding the temporal patterns of a network involves temporal exploratory data analysis and selection of statistically significant patterns for analysis. Jointly applying temporal exploratory data analysis and spatial exploratory data analysis may have statistical interactions and implications not yet articulated in the literature.

Integration of Network, Spatial, and Temporal Visualization:

Each network domain is likely to have an optimal method of data display and exploration. Providing a view of each element of a network, its relationships, its spatial extent and characteristics of place, and its temporal pattern is likely to be essential if the interaction of network, place, and time are to be understood.

References:

Bender-deMoll, S., and McFarland, D., **The Art and Science of Dynamic Network Visualization**. *Journal of Social Structure* 7(2), 2006.

Goodchild, M., and Gopal, S., **The Accuracy of Spatial DataBases**. London: Taylor and Francis, 2005.

Olson, J. and Carley, K.M., **Visualizing Spatial Dependencies in Network Topology.** CASOS Technical Report, July 12, 2010.

Scellato, S., Mascolo, C., Musolesi, M., Latora, V., **Distance Matters: Geo-social Metrics for Online Social Networks.** Proceedings of the 3rd Workshop on Online Social Networks (WOSN'10). Colocated with USENIX'10. Boston, MA, USA. June 2010.

Tobler W., (1970) **A computer movie simulating urban growth in the Detroit region.** *Economic Geography* 46(2): 234–240.

Wilson, C., Boe, B., Sala, A., Puttaswamy, K., and Zhao, B., **User Interactions in Social Networks and their Implications**. Computer Science Department, University of California at Santa Barbara, EuroSys2009, Nuremburg, Germany.