Anthropologist Based on Social Network Analysis and Data Mining Approaches

M. Florence Diana
Head Department of Computer Application
Bon Secours College for Women
Thanjavur, Tamil Nadu, India

Abstract: An anthropologist shares with the “SNA and data mining community” his own anthropological perspective framed during more than five decades of network thinking about a broad range of anthropological problems. For 50 years he has viewed all people, things, and ideas in dynamic relationships. That perspective is a network perspective and at the same time anthropological, combining ethnographic, historical, holistic, and comparative views.

It is valuable and beneficial to the community of scholars who use network analysis to try to understand what is going on, what went on before, and what the future prospects are. As an anthropologist, his interest is more in the wholes generated by network linkages—systems of households, bands, lineages, communities, corporations, governments—than in the individual persons linked. Even now, when personal network “communities” are getting so much attention network analysis can clarify the more complex wholes such as multinational corporations and supranational systems. Those important entities and the problems they represent should not be left to economists and politicians. Concepts considered include system evolution and increasing complexity; anthropological views of transactions, relations, modes of transactions, spheres of transactions in multicentric economies, complexity across the full range of embedded networks—material, biological, and sociocultural. Social network analysis can help to define systems at various levels of integration, both within communities and in the widest conceivable supranational level. Techniques such as regular equivalence and block-modeling are useful in sorting the subsystems of complex systems. He concludes with hope that as improved methods of data mining and network analysis are developed, other anthropologists and social scientists will be able to measure the evolution of supranational sociocultural systems that involve both states and multinational corporations. Digital Anthropology reveals how only the intense scrutiny of ethnography can overturn assumptions about the impact of digital culture and reveal its profound consequences for everyday life.

Keywords: Data mining, Network thinking, Network analysis, Supranational, sociocultural systems.
Email communication Graph

Nodes = People
Links = Emails

Applications of Social Network DM
Typical applications of social network analysis and data mining:
- Detection of criminal activity, Counter terrorism, "homeland security" and intelligence.
- Analysis of relationships within companies.
- Sociological and anthropological studies.
- Reciprocal trust schemes such as e-bay ratings.
- Recommended friends on Facebook.
- Filter or recommend social media content. Etc….

How do we Analyse Networks

Degree Centrality
- Number of connections to other nodes.
- High values mean many connections.
- Can measure links in and out separately.
- Applications
  - Who is most listened to on Twitter?
  - Who has most contacts within a company?
  - Which user’s reviews influence others the most?

Closeness Centrality
- The average number of steps required to reach any other node. Communications are easier if you don't have to go through too many people.
- Applications
  - Is this person central to the group?
  - Is your message likely to reach the audience?
Betweenness Centrality
- How much of a link between other nodes is this node?
  - Applications
- Someone who has a high betweenness centrality is often a broker between others.
- What happens if this person leaves the network?

Structural holes
- Gaps in linkage between groups.
  - Applications
- Bridges across this access information from both, suggesting influence and understanding of an organisation.
  - Can we create a bridge?
  - Is there an opportunity to control or influence communications between groups?

Degree of centralisation
- is the network held together by just a few nodes?
  - Or is it more cohesive?
- Measures include average and variance of degree centrality
  - Applications
- Is a crime network vulnerable to disruption?
  - What happens to a company if a few key people leave?

Data Mining Approaches to Networks
- Structural Equivalence
  - Find nodes with similar roles in the network
- Cluster Analysis
  - Identify groups of nodes which are closely connected - and characterise them
- Identifying the Most Influential People
- Predicting Node Types (e.g. Fraudster)
- Profiling Sub-networks (e.g. terrorist cell)

Twitter - Clustered Network
To reduce clutter, we can cluster people who reference each other, and only show links within clusters.
Data Mining Social Networks – Challenges

- **Standard problems**
  - Incompleteness – We don’t know everything
  - Incorrectness – What we think we know is wrong
  - Inconsistency – We have contradictions in our data

- **Data transformation** - Getting data into a form acceptable by your tools

- **Fuzzy Boundaries** - Networks do not normally have distinct boundaries

- **Network Dynamics** - Relationships change over time
Example Application Viral Marketing
"In our experiments with the Epinions knowledge-sharing Web site, the most valuable customer had a network value of over 20,000, meaning that marketing to that customer was as effective as marketing to over 20,000 others in the absence of network effects, but the customer's number of direct links to others in the network (i.e., people who read his reviews) was much smaller."

Impact Computers on SNA of
The rise in the power and use of computers has had two main impacts.
1. New data is available from logs of email conversations, phone calls, chat and website usage, facebook friends, tweets etc...
2. Computers can be employed for analysis and data mining.

Role of computer analysis and conclusion
Data collected about social networks can be complex and large. Imagine a network documenting each purchase you've made using a credit/debit card, every phone call and SMS, each email etc. When these kinds of data are collected over large populations, the resulting graphs are much too large to be understood by eye. Computers have provided the ideal infrastructure for Fostering social interaction, Capture it at a very fine granularity, Practically no reporting bias. The emerging of computational social science has the potential to revolutionize social science.

REFERENCES