Why SNA? A Network Engineer's Perspective

Srinivasan V

Based on a survey by Katsaros et al. [1]

December 22, 2010

Ubiquitous Ad-hoc networks

- Ad-hoc networks have become ubiquitous thanks to:
  - device miniaturization
  - improvements in Wireless Communication
- Eg: MANETs, WSNs, WMNs
- Deployable for
  - disaster relief
  - conference and battlefield environments
  - smart vehicles
  - wireless Internet connectivity

Need for Self-organization

- Lack of fixed infrastructure
- Frequent changes in Network topology (due to mobility and/or intermittent operation of hosts)

Modes of Self organization

- Forming a hierarchy (clustering)
- Designing network spanner for efficient routing (information dissemination)
- Modeling of the network (topology description and mobility models)

SNA vs Traditional Approaches

Communication is opportunistic in nature

- Cannot be easily/efficiently described as an optimization problem
- Systematic approaches like cross layer optimization are difficult to apply
- Can benefit from SNA [2], since most adhoc networks are human-centered.
  - follow the way humans come into contact

Social Networks

- Collection of actors (nodes in the network) and set of relation information (edges) between them.
- Has been of interest in sociology, data mining, and recently in networking communities.
- SNA as a network measurement task dealing with structural properties of the network graph
  - existence of communities
  - node centrality
  - topology evolution
  - network robustness

Centrality measures

- Most important actors in the network, using graph-theoretic techniques
- Most "strategically located"
- Based on degree information of the actors
  - degree centrality
  - spectral centrality
- Based on geodesics i.e., shortest path between actors
  - closeness centrality
  - betweenness centrality
  - bridging centrality
- Can also be defined for groups of nodes by looking at them as a supernode

Degree based centralities

- **Degree centrality:** $C_{deg}(i) = \frac{\texttt{degree}(a_i)}{n-1}$
- **Spectral centrality:**
  - based on spectral properties of the adjacency matrix
  - can define prominence recursively (a node is prominent if it is pointed to by prominent nodes)
  - e.g. PageRank metric [3], used by Google for ranking webpages

$$PR(i) = \frac{\alpha}{n} + (1-\alpha) * \sum_j \frac{PR(j)}{k_{out}(j)}$$

    where $\alpha \in [0,1]$, k_{out} = outdegree of node i

  - solving the above equation is equivalent to finding the principal eigen vector of matrix B obtained from adjacency matrix A.

$$B_{i,j} = \frac{\alpha}{n} + \frac{1-\alpha}{k_{out}(a_j)} * A_{j,i}$$

  - In case of weak degree correlation (as in Web), indegree is a gross measure of Pagerank index.

Geodesic based centralities

- Based on the shortest path distances between nodes
- Shortest paths are mostly used for various networking tasks such as routing
- **Closeness centrality:**
  - describes efficiency of information propagation from a given node
  - $C_(i) = \frac{1}{\sum_j \mathtt{distance}(i,j)}$
  - distance may be measured in number of hops, delays etc.
- **Shortest path Betweenness centrality (SPBC)**
  - describes frequency of a node in the shortest paths between other nodes
  - measure of a node's control on information flow between other nodes
- Bridging centrality [4] is an extension of SPBC taking into account a node's connection to high degree nodes.

Critical analysis of Centrality techniques

- Defined in a centralized fashion
- Such network-wide centralized computations are prohibitive in large scale ad-hoc networks
- Localized centrality measures have been defined
  - μ power community index [5]
  - Cumulative contact probability [6] (for poisson model of social contacts)
- There is still a need for easily computable, but relatively accurate ranking of nodes across the entire network

Community identification

- Set of nodes which have high density of internal links [7]
- Much lower density of links across different groups
- Complex self-organized networks tend to exhibit presence of communities.
- In adhoc networks, community identification can help in efficient delivery of information, against naive flooding.
- Several different approaches have been proposed to quantify the goodness of a community structure

Approaches to Community Identification

- Most of the community measures are based on **Cut size**
  - Cut size = the number of edges that lie at the border of the communities
  - Eg. Minimum cut size, Conductance and Normalized cut
- **Modularity**
  - A subgraph is a community if the number of edges within the subgraph exceeds the expected number of edges the subgraph would have in a random graph
- **Hard community**
  - $i \in C_k$, if $\sum_{j \in C_k} A(i,j) > \sum_{j \notin C_k} A(i,j)$
  - this definition is quite restrictive, and allows a node to be part of at most one community

Approaches to Community Identification

- **Generalized community**
  - Allows for overlapping communities
  - A set of nodes is a community if number of links (collectively) to the nodes in the community exceeds that to the nodes not in the community
- **k-Cliques**
  - A k-clique is a subgraph of k nodes which is completely connected.
  - Two k-cliques are adjacent if they share k-1 nodes.
  - *k-clique communities* are defined as the maximal union of adjacent k-cliques
- **Clustering coefficient**
  - Defines the cliquishness of the network
  - $CC = \frac{3 * number\ of\ triangles\ in\ the\ network}{number\ of\ connected\ triples\ of\ vertices}$

Critical analysis of Community Identification Techniques

- Community definitions are based on metrics which are NP-hard to compute
- Difficult to handle and maintain especially for mobile networks
- Most of them provide non-overlapping communities
  - Overlapping communities are better for forwarding related applications in adhoc nets
- Modularity-optimization might overpartition or underpartition networks failing to detect the true community structure.
- Clique based techniques may not work in adhoc networks which are sparse.
- Need community finding algorithms that are stable across the timescales.
  - Must be able to run incrementally subject to addition/removal/mobility of nodes

SNA in Network Protocol Design

- Motivated by Human-based nature of opportunistic networks
- SNA techniques have been used in
  - Routing
  - Information Dissemination tasks
  - modeling the entities of a network

Need for SNA in Routing

- Routing can be table-driven or on-demand
- Approaches involving routing tables are difficult to maintain in the presence of mobility
- For DTNs, on-demand protocols follow a next-hop hill-climbing approach
  - Also called the *store-carry-and-forward* technique
  - Each node independently makes a forwarding decision when two nodes meet.

SNA in Routing: Examples

SimBet

- Uses betweenness centrality to make forwarding decisions
- It also involves exchange of social similarity (no. of common neighbours) to the destination
- In case of no common nodes to the destination, the packet is routed to structurally central node
    - from where there is high likelihood of finding a path to the destination

SNA in Routing: Examples

Bubble protocol [8]

- Uses centrality and community measures to make forwarding decisions
- Each node has a local ranking (within community) and global ranking
- Forwarding nodes use both these ranks alternately to reach the destination node

SNA in Routing: Examples

FairRoute [9]

- Uses perceived interaction strength to the destination
- But this might also cause creation of hotspots in the network
- *Assortativity* - a node forwards a packet only if the receiver queue size is less than or equal to sender queue size.
- By combining both these, hotspots are reduced, but with a slightly lower throughput

Issues

- Centrality based routing protocols tend to select same nodes as forwarders
- As a result the central nodes spend a lot of energy, and links to them become congested
- But SPBC based routes are attractive, for latency minimization
- Routing using local community structure might help solve some of these issues
- Also we can integrate power control and routing
- We can also define routing specific centralities, instead routing based on pre-defined centrality notion

SNA in Information Dissemination

- Content provisioning is a prime application of adhoc networks
- Placing information in nodes becomes challenging due to volatility of network topology
- Cooperative caching: managing an aggregate cache across multiple nodes to reduce communication costs [10]
- Optimal placement of information in nodes [11] - equivalent to the k-median problem
  - A scalable near-optimal placement algorithm was proposed based on Betweenness Centrality
- The problem of cost-effective outbreak detection in sensor networks can be reformulated as the problem of selecting the most influential nodes in a social network. [12]

Issues

- Current list of protocols work only for static or semi-static scenarios.
- Developing solutions for finding dominating sets for mobile networks
  - Must also take into account latency of information dissemination
  - Solution must be computable in a disrtibuted fashion
- SNA techniques are not applicable in Vehicular Adhoc Networks [13] (VANETs) where the link duration is a few seconds

SNA in Network Modeling

- SNA techniques have been used to identify robustness of network to a targeted node failure.
- Distributed community detection algorithms have been developed for DTNs.
- Social network data can also be used to define mobility models for human-centered adhoc networks.
- Studies have been carried out in VANETs to look at the evolution of network graph across snapshots in time and space.

Issues

- Mining of time-varying network data cannot be done efficiently with graph-theoretic tools.
- A tensor could be used to represent a continuously changing adjacency matrix.
- Centrality and Community detection over such tensors could lead to new notions

Further Research

- Networking community has used SNA tools without contributing much novelty to the field of SNA
- Devising centrality and community detection algorithms that are quick to compute and capture the ground truth
- Investigation of time varying network topology to develop appropriate concepts
- Synergy between Complex network science and Communication networks will benefit both the disciplines.

References I

Dimitrios Katsaros, Nikos Dimokas, and Leandros Tassiulas.
Social network analysis concepts in the design of wireless ad hoc network protocols.
IEEE Network.

S. Wasserman and K. Faust.
Social Network Analysis.
Cambridge University Press, Cambridge, 1994.

L. Page, S. Brin, R. Motwani, and T. Winograd.
The pagerank citation ranking: Bringing order to the web.
Technical report, Stanford University, Stanford, CA, 1998.

Woochang Hwang, Taehyong Kim, Murali Ramanathan, and Aidong Zhang.
Bridging centrality: graph mining from element level to group level.
In *Proceedings of the 14th ACM SIGKDD*, pages 336–344. ACM, 2008.

References II

Nikos Dimokas, Dimitrios Katsaros, and Yannis Manolopoulos.
Cooperative caching in wireless multimedia sensor networks.
MONET, 13(3-4):337–356, 2008.

Wei Gao, Qinghua Li, Bo Zhao, and Guohong Cao.
Multicasting in delay tolerant networks: a social network perspective.
In *Proceedings of the 10th ACM MobiHoc, May 18-21, 2009*, pages 299–308. ACM, 2009.

M. E. J. Newman.
Modularity and community structure in networks.

Pan Hui, Jon Crowcroft, and Eiko Yoneki.
Bubble rap: social-based forwarding in delay tolerant networks.
In *Proceedings of the 9th ACM MobiHoc, May 26-30, 2008*, pages 241–250. ACM, 2008.

References III

Josep M. Pujol, Alberto Lopez Toledo, and Pablo Rodriguez.
Fair routing in delay tolerant networks.
In *INFOCOM*, pages 837–845. IEEE, 2009.

Nikos Dimokas, Dimitrios Katsaros, Leandros Tassiulas, and Yannis Manolopoulos.
High performance, low complexity cooperative caching for wireless sensor networks.
In *WOWMOM*, pages 1–9. IEEE, 2009.

Panagiotis Pantazopoulos, Ioannis Stavrakakis, Andrea Passarella, and Marco Conti.
Efficient social-aware content placement in opportunistic networks.
In *Proceedings of WONS'10*, WONS'10, pages 17–24, Piscataway, NJ, USA, 2010. IEEE Press.

References IV

Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne M. VanBriesen, and Natalie S. Glance.
Cost-effective outbreak detection in networks.
In *Proceedings of the 13th ACM SIGKDD August 12-15, 2007*, pages 420–429. ACM, 2007.

George Pallis, Dimitrios Katsaros, Marios D. Dikaiakos, Nicholas Loulloudes, and Leandros Tassiulas.
On the structure and evolution of vehicular networks.
In *MASCOTS*, pages 1–10. IEEE, 2009.