# Social Network Analysis Concepts in the Design of Wireless Ad Hoc Network Protocols

**Dimitrios Katsaros, University of Thessaly**
**Nikos Dimokas, Aristotle University**
**Leandros Tassiulas, University of Thessaly**

**Abstract**

This article presents a survey of ad hoc networking protocols that have used concepts such as centrality metrics and community formation from the area of social network analysis, which is seen as a network measurement task that deals with structural properties of the network graph. We recognize the synergy among social network analysis and ad hoc networking as a fertile research area that can provide significant advances for the design of network protocols, especially in environments where the communication is opportunistic in nature and therefore cannot be easily or efficiently described as an optimization problem, and other systematic approaches like cross-layer optimization are more difficult to apply.

The recent advances in device miniaturization, and progress in wireless communications and the respective system/application software have made the presence of ad hoc networks ubiquitous. A wealth of ad hoc networks is encountered today, such as mobile ad hoc networks (MANETs), wireless sensor networks (WSNs), and wireless mesh networks (WMNs). They have potential applications in disaster relief, conference and battlefield environments, wireless Internet connectivity, and smart vehicles. Ad hoc networks consist of wireless hosts that communicate with each other in the absence of a fixed infrastructure; each host acts as a relay that forwards messages toward their destination.

The absence of fixed infrastructure and the frequent changes in network topology (due to mobility and/or intermittent operation of the hosts) makes self-organization of these wireless networks a necessity. Self-organization touches on many aspects of nodes' networking; for instance, calls for solutions to the problem of creating a hierarchy (clustering) over a flat ad hoc network or creating a network spanner to reduce redundant rebroadcasts, pertain to the design of routing (unicasting, multicasting) protocols, which touch on problems related to the design of information dissemination (e.g., caching, message ferrying) protocols and also to the modeling of the network itself (e.g., network topology description, establishment of mobility models that closely represent true human behavior).

The plain fact that the nodes of an ad hoc network are strongly interdependent rather than independent and autonomous agents, which communicate via wireless channels that transfer resources, helped researchers to realize the significance of borrowing concepts from the field of social network analysis (SNA) [1] to the design of more efficient protocols. This borrowing was further enforced by the fact that many of the ad hoc networks are basically human-centered and follow the way humans come into contact. Moreover, because of the lack of infrastructure, it is rather challenging to develop more systematic design optimization approaches as in, for instance, cellular networks. Greedy best effort techniques are used primarily for opportunistic ad hoc networks and may benefit significantly from the social networking perspective.

Informally, a *social network* is a collection of *actors* (i.e., network nodes), a set of relational information on pairs of actors (i.e., wireless links), and possible attributes of the actors and/or the links. The notion of a social network and the methods of SNA constitute a very old discipline, and attracted significant interest initially from the social and behavioral communities, later in data mining, and only recently from the networking community. This interest stems from the focus of SNA on relationships among entities, and the patterns and implications of these relationships. SNA could be viewed as another network measurement task, while the traditional tasks of network measurement deal with issues such as traffic monitoring, latency, bandwidth, and congestion. The analysis of the *social* aspects of a network is the study and exploitation of the structural information present in the network, such as existence and strength of communities, node centralities, network robustness to node removal, and topology evolution over time, among others.

The purpose of this article is to provide an overview of the most important concepts of SNA that have been used in the design of wireless networks protocols; additionally, it aims at pinpointing the shortcoming of the traditional SNA concepts, and at proposing some possible roads for further research concerning the synergy between SNA and protocol design. The rest of this article is organized as follows. The next sec-

*Research supported by the project "Control for Coordination of Distributed Systems," funded by the EU.ICT program, Challenge ICT-2007.3.7.*

  

tion describes the most popular SNA concepts that have been used so far in protocol design. We then provide a categorization of the application of SNA concepts in various networking areas. We then demonstrate the inefficiency of traditional SNA concepts when applied to protocol design and propose possible roads for future investigation, and the final section concludes the article.

## Social Network Analysis Concepts

The area of SNA is a broad, diverse, and theoretically varied field with a long and rich history; in the following two subsections we present only its most popular concepts that have also found significant applications in protocol design.

### Centrality Measures

One of the primary questions of SNA is the identification of the most important actors in a social network using graph-theoretic terms. The importance or prominence of an actor in this context is synonymous with the *strategic* location of the actor within the network. This prominence is described by numerous *centrality* measures. Currently, the most noteworthy and substantively interesting centrality metrics relevant to the present study belong to two generic categories: a) metrics based on the degree information of an actor and b) those based on the geodesic (i.e., shortest path) distances of actors. The former category includes the degree and spectral centrality, whereas the latter includes the closeness, shortest path, and bridging centrality. The members of each family are described in the sequel.

*Degree-Based Centrality Metrics* — The degree-based family of centrality measures consider an actor prominent if the *ties* of the actor make the actor *visible* to the other actors in the network. Intuitively, according to these definitions an actor is prominent if s/he is adjacent to many other (highly prominent) actors. Degree centrality takes into account only the number of ties, whereas spectral centrality takes into account both the number and quality of the neighboring actors. Specifically, these metrics are defined as follows.

*Degree Centrality* — It is loosely defined as the number of one-hop neighbors of an actor [1]. For a network consisting of $n$ actors, the degree centrality of an actor $a_i$ is

$$DegC_{a_i} = \frac{degree(a_i)}{n-1}. \tag{1}$$

*Spectral Centrality* — There are various definitions of *spectral* centrality metrics, which are referred to as spectral because they are based on the spectral properties of the matrix that represents the relationships among the actors. These metrics define the prominence of an actor recursively (i.e., an actor is prominent if it is pointed to by other prominent actors). The most popular of the spectral centrality measures is the PageRank metric [2], which is one of the methods used by Google to rank web pages. The PageRank of an actor $a_i$ in a network consisting of $n$ actors is recursively defined as follows:

$$PR_{a_i} = \alpha\frac{1}{n} + (1-\alpha) \sum_{a_j : a_j \rightarrow a_i} \frac{PR_{a_j}}{k_{out}(a_j)}, \tag{2}$$

where $\alpha$ is a scalar quantity in the range (0, 1), and $k_{out}(a_j)$ is the *outdegree* of actor $a_j$. Solving the set of the above equations is equivalent to finding the principal eigenvector of matrix $A$ with elements $A_{ij}$:

$$A_{a_i,a_j} = \frac{\alpha}{n} + \frac{(1-\alpha)}{k_{out}(a_j)} ADJACENCY\_MAT_{a_j,a_i}. \tag{3}$$

The aforementioned metrics are also interrelated in the sense that the number of incoming links of an actor is a gross measure of its PageRank centrality index. The accuracy of this approximation depends on the topology of the underlying graph; in the Web, for instance, due to its weak degree correlations, both theoretical and empirical analyses conclude that this approximation can be relatively accurate.

*Geodesic Distance-Based Centrality Metrics* — This family of metrics exploits the distance (i.e., shortest path) between actors in order to define centrality measures. Since shortest path computations are frequently used for various networking tasks (e.g., routing, monitoring), a lot of variations on these metrics can be found in the literature, but in the next paragraphs we present only the most popular ones.

*Closeness Centrality* — This describes the efficiency of information propagation from one actor to all the others, and it is defined as the inverse of the sum of the distances between a given actor and all other actors in the network [1]. The closeness centrality of an actor $a_i$ is

$$C_{a_i} = \frac{1}{\sum_{j \neq i} distance(a_i, a_j)}. \tag{4}$$

The distance can be measured in number of hops, delays, and so on. Closeness centrality gives an estimate of how long it will take information to spread from a given actor to the rest of the network actors. Evidently, this measure could be used in applications where, for instance, we need to elect a single leader actor to propagate alert messages.

*Shortest-Path Betweenness Centrality* — It describes the frequencies of actors in the shortest paths between indirectly connected actors and is formally defined as the fraction of the shortest paths between any pair of actors that pass through an actor [1]. The shortest-path betweenness centrality of an actor $a_i$ is

$$SPBC_{a_i} = \sum_{j \neq i} \frac{sp_{j,k}(a_i)}{sp_{j,k}}, \tag{5}$$

where $sp_{j,k}$ is the number of shortest paths linking actors $j$ and $k$, and $sp_{j,k}(a_i)$ is the number of shortest paths linking actors $j$ and $k$ that pass through $a_i$. Shortest-path betweenness centrality is a measure of the extent to which an actor has control over information flowing between others. This centrality metric can be used, for instance, in message-carrying applications where we need to forward a packet to a node that is promising to deliver it with success and/or faster to its final destination.

*Bridging Centrality* — It is calculated by multiplying the shortest-path betweenness centrality by a bridging coefficient [3]. The bridging coefficient is the ratio of the inverse of a node degree to the sum of the inverses of all its neighbor degrees. The bridging coefficient of an actor describes how well the actor is located between high-degree actors. The bridging centrality of an actor $a_i$ is

$$BrC_{a_i} = SPBC_{a_i} \times \beta(a_i). \tag{6}$$

Bridging centrality could be used in applications where the dense areas of a network should be identified for purposes of, say, placing proxies or preventing congestion.

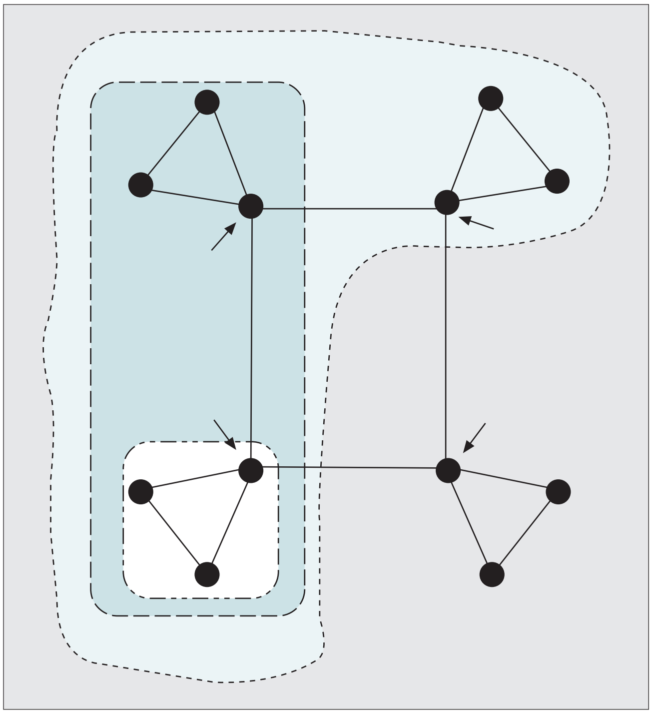

Figure 1. *A sample graph where the hard community definition fails to identify communities.*

Apart from the individual actor's versions for closeness and shortest-path betweenness centrality, there are also the *group* versions of these metrics, which calculate the importance of groups of actors. For instance, the *group betweenness centrality* (GSPBC) of a set of actors is roughly defined as the total fraction of shortest paths that traverse at least one member of the set.

*A Brief Critique of the Centrality Metrics* — All of the above centrality metrics have been defined in a centralized fashion (i.e., taking into account all the network actors). Such centralized computations are prohibitive for ad hoc networks due to the communication complexity of learning the whole network topology, and thus localized versions of them have been used in the literature of protocol design. Inspired by these metrics, localized centrality metrics have been proposed in the literature, such as a truly distributed centrality called μ-Power Community Index [4] and cumulative contact probability [5] based on Poisson modeling of social contacts.

There are also cases (e.g., leader election) where it is useful to have a relatively accurate ranking of the nodes considering the whole network topology, but we do not wish to pay the high computational cost of precise calculations (e.g., exact closeness centrality values). In these cases, most of the aforementioned measures are not appropriate. It would be very useful if we had in our arsenal low-cost methods for the estimation of approximations for these centrality metrics; currently the literature is lacking such techniques with sound theoretical background.

## Community Definitions

Another feature of complex self-organized networks is the formation of compartments that have their own role and/or function. In the network representation, such compartments appear as sets of nodes with a high density of internal links, whereas links between compartments have comparatively lower density; these subgraphs are called *communities*. These partitions can potentially lead to reduced, more manageable representation of the original network. The identification of communities in wireless networks can be used, for instance, in packet-switched networks to improve delivery of information by selecting appropriate forwarders instead of performing naive *oblivious* flooding.

The definition of communities as sets of nodes with high density among these nodes is vague and qualitative; moreover, there is a lack of theory or consensus on measures to quantify the goodness of a community structure. For example, if we consider as such a measure the *cut size* — the number (or the sum of weights) of edges that lie at the boundaries of communities — then, despite the fact that this criterion is intuitive, it has negligible applicability since the partition with minimum cut size is often trivial. Therefore, the topic of community detection in graphs has a long history, and multiple methods and heuristics have been proposed to partition networks into communities.

An early definition is based on the *conductance* [6]; the conductance of a community is the ratio between the cut size of the community and the minimum between the total degree of the community and that of the rest of the graph. The problem of finding a cut with minimal conductance is NP-hard. A similar measure is the *normalized cut*. Another popular way of defining communities is based on *modularity* [7], which is a global criterion to define a community, a quality function, and the key ingredient for many popular methods for graph clustering. In the standard formulation of modularity, a subgraph is a community if the number of edges inside the subgraph exceeds the expected number of internal edges the same subgraph would have in a random graph.

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - P_{ij}\right)\delta\left(C_i, C_j\right), \tag{7}$$

where the sum runs over all pairs of vertices, $A$ is the adjacency matrix, $m$ the total number of edges of the graph, and $P_{ij}$ represents the expected number of edges between vertices $i$ and $j$ in the random graph. The δ-function yields one if vertices $i$ and $j$ are in the same community ($C_i = C_j$), zero otherwise.

Despite the appealing definition of modularity, it has been shown that optimizing modularity can over- or underpartition the network, failing to find the most natural community structure. To compensate for this, we can add an ad hoc resolution parameter that can be tuned to bias toward small or large communities, at the expense of requiring administrative (human) intervention.

Conductance and modularity are network-wide metrics and thus are cumbersome when used for ad hoc network protocols. In this area, localized community definitions are more appropriate (i.e., definitions that look at individual nodes) [8]. One such definition is that of a *hard community*, which implies that a node belongs to a specific community if the number of its links toward the nodes of this community is strictly larger than the number of its links toward other nodes not belonging to this specific community.

This definition is quite restrictive, since it allows a node to be in at most one community or in no community at all; for instance, no community will be discovered for the graph shown in Fig. 1, although at first glance we could recognize four communities based on our human perception (i.e., the four triangles); a careful look would leave us puzzled as to whether the four nodes (pointed to by arrows) at the corners of the square really belong to any community. Actually, we could consider as a community any set of nodes enclosed by the dashed lines.

Thus, a new and more flexible definition of a community is needed, the *generalized community*, which also allows for overlapping communities and requires no administratively tuned parameters. This definition describes a set of nodes to be a community if the number of their links (collectively) toward the

| Article | Centrality | Community | Other feature |
|---|---|---|---|
| [10] | SPBC | — | Clustering coefficient |
| [11] | SPBC | *k*-clique percolation, overlapping | — |
| [12] | — | — | Interaction strength |
| [5] | Cumulative contact prob. | *k*-clique percolation, overlapping | — |
| [13] | SPBC | — | — |
| [14] | Variant of *SPBC* | — | — |
| [4] | μ-power community index | — | Localized clustering coeff. |
| [15] | | *k*-clique, modularity optimized | — |
| [16] | *SPBC* for edges | Modularity optimized, non-overlapping | — |
| [17] | Localized BrC | — | — |
| [8] | *Deg*, *C*, *SPBC*, *BrC* | Generalized, overlapping | Localized clustering coeff. |

Table 1. *Literature categorization with respect to the SNA concepts used.*

nodes of the community is larger than the number of their links toward nodes not belonging to the community. The definition implies that a node may belong to a community even if its number of links toward the nodes of the community is smaller than the number of its links toward nodes not belonging to the community — this node actually works as a connector node.

Apart from the above definitions, a very popular graph-theoretic way of defining communities is based on the clique percolation method, which builds up communities from $k$-cliques. A $k$-clique corresponds to a complete (fully connected) subgraph of $k$ nodes (e.g., a 3-clique is equivalent to a triangle). Two $k$-cliques are considered adjacent if they share $k - 1$ nodes. A *$k$-clique community* is defined as the maximal union of $k$-cliques that can be reached from each other through a series of adjacent $k$-cliques.

Finally, another relevant SNA concept that measures the cliquishness of a network is the *clustering coefficient* [1], defined as follows:

$$CC = \frac{3 \times \text{number of triangles in the network}}{\text{number of connected triples of vertices}}, \qquad (8)$$

where a *connected triple* means a single vertex with edges running to an unordered pair of others.

*A Brief Critique of Community Definitions* — Community definition based on NP-hard metrics, although mathematically sound, are practically difficult to handle and maintain, especially for mobile networks. Moreover, these metrics define non-overlapping communities, and when used in ad hoc network protocols, they require an additional task of finding which nodes will be the relays. On the other hand, overlapping communities seem more appropriate for message forwarding applications, because the nodes that belong to more than one community can immediately be selected as gateways to forward the message across communities. Also, as said earlier, modularity-optimized communities can overpartition or underpartition the network, failing to detect the *true* community structure, and the remedy to this situation can simply add more administratively tuned parameters to the algorithm for which there is no effective way to determine the best parameter values. Finally, clique-based community definitions are not expected to work well in wireless ad hoc networks, since these networks are sparse; thus, these algorithms will end up with a large number of small communities.

The main drawback though of all the presented approaches concerns their stability across timescales; apparently a community with members that change rapidly over time may not be truly useful for protocol design, since it requires frequent runs of the community-finding algorithm. At this point, the literature needs definitions of community stability as well as algorithms that will run incrementally to address additions of nodes and/or edges. In the next section we survey the most significant work using SNA notions in designing efficient protocols for wireless networks, broadly grouped into three categories: work that mainly addresses routing problems, work that concerns information dissemination tasks, and finally, work used for network *entities* modeling.

## SNA in Network Protocol Design

As already explained, the human-based nature of opportunistic networking has been one of the driving forces for the exploitation of social theories and tools in protocol design. For instance, it is observed that nodes of such networks tend to meet a certain group of nodes with higher probability than other nodes outside the group [9]. Such behaviors have been exploited in routing protocols and also in sensornet protocols — despite the fact that no human intervention is apparent there. Soon SNA theories found their way as a network measurement task for both semi-static and highly mobile ad hoc networks. Table 1 illustrates the synergy between SNA and ad hoc networking.

### SNA in Routing

Routing protocols for ad hoc networks can be broadly divided into proactive table-driven and reactive on-demand schemes. The former category employs routing tables, which contain approximated shortest paths between nodes and direct packets accordingly, but they suffer from huge communication overhead to update the tables when mobility occurs. The latter category protocols initiate the path discovery process in response to receiving message delivery requests. A special case of on-demand protocols are those for delay-tolerant networks that are based on best next-hop hill-climbing heuristics. Such routing protocols address the common problem of partitioning sparse mobile ad hoc networks by physically carrying messages between disconnected parts of the network. These schemes are sometimes referred to as store-carry-and-forward protocols. This type of routing consists of each node independently making forwarding decisions that take place when two nodes meet. A message gets forwarded to encountered nodes until it reaches its destination.

The SimBet protocol was the first one of this category that exploited the concept of betweenness centrality in order to make forwarding decisions. SimBet exploits the exchange of pre-estimated betweenness centrality metrics and locally determined *social similarity* (based on the number of common neighbors) to the destination node. When the destination node is unknown to the sending node or its contacts, the message is routed to a structurally more central node where the potential of finding a suitable carrier is dramatically increased. This idea was further enhanced in [11] where the concepts of centrality and communities were combined to assist in making forwarding decisions in the Bubble protocol. Nodes are grouped into overlapping communities, and each node has a global ranking and local ranking (inside its community). When a node has to deliver a message to another node, it searches the hierarchical global ranking tree until it reaches a node in the same community as the destination node. Then, using a local ranking system, it searches for a *significant* node to forward the message until the destination is reached or the message expires. A similar methodology using centralities and communities was also followed by [5], but solved a multicasting problem instead of unicasting.

An inherent drawback of both routing protocols is that they keep selecting the same nodes as forwarders because these nodes have the largest centrality values, thus consuming their energy quite fast. For instance, in the SimBet protocol 10 percent of the nodes carry out 54 percent of all forwards. Instead of this *greedy* approach, we can perform routing utilizing community information [18]. Suppose that we can come up with a partitioning of the nodes into communities and that each node is aware of the community to which it belongs as well as the community of every other network node. Then, when a node wishes to forward a packet to a destination, it is sufficient to forward it to a node that belongs to the same community as the destination. Such a basic technique could be the inspiration of a routing scheme, after resolving significant details such as avoiding cycles, reducing latency, and so on. On the other hand, such techniques are able to utilize more network paths, thus leading to a greater network stability region, similar in spirit to the backpressure principle.

Motivated by the unfairness of the SimBet-like protocols, the FairRoute protocol [12] exploited the social process of *perceived interaction strength* and *assortativity*. The former metric suggests forwarding the messages to nodes that have stronger relationships with the destination node; the interaction strength increases with increased number of contacts and reduces exponentially over time. Still, using this metric does not avoid bias in selected forwarders (i.e., it creates hot spots in communication). To remedy this situation, the FairRoute protocol allows a node to forward a packet to another node only if this second node has a queue length equal to or shorter than the first node's queue. This is the notion of assortativity — the tendency of nodes with similar number of edges to connect. This policy has a significant positive effect on hot spots but achieves slightly lower throughput.

*Complications in SNA-Based Routing and Some Possible Solutions* — It is obvious that the utilization of *SPBC*-based metrics for packet forwarding will inevitably create hot spots. On the other hand, shortest path communication is appealing due to its relation to latency minimization. Thus, how can we bridge these seemingly contradictory goals? We can come up with a couple of solutions to address this trade-off. The first one is not to create global communities, but allow each node to define its own communities around it. This is like having multiple overlays over a single network. Of course, these node-dependent communities can span only the node's neighborhood and need not span the whole network. Then, depending on the source and destination nodes, the routing of a packet will follow different paths, thus effectively reducing energy depletion of the same nodes. The second solution is to integrate power control and routing. With this technique, the data will be routed on the shortest paths, and the nodes that are expected to relay more packets for others are skipped or jumped over. Finally, a solution with a different flavor would be to define routing-specific centralities instead of routing schemes based on specific centralities (as done so far). This means that the centrality of a node will be estimated based on the number of packets forwarded via it, not on the physical connectivity of this node.

## SNA in Information Dissemination

As content provisioning becomes the driving application of modern ad hoc networks, placing information in the nodes of such a networking structure becomes significant and challenging due to the lack of infrastructure and the volatile network topology. Cooperative caching is a technique where multiple nodes share and coordinate cache data to cut communication cost and exploit the aggregate cache space of cooperating nodes. The benefit of cooperative caching is reduced energy consumption and smaller data retrieval latency. The first approaches to cooperative caching did not exploit the link structure of topology, but mainly worked as en route caching algorithms, where each node made practically independent decisions about whether to cache or not a *passing-by* information item (or the path to it). Dimokas *et al*. in [13] introduced the notion of *mediator* nodes in the *NICoCa* protocol for sensornets that coordinate the caching decisions (e.g., cache admission, eviction, routing). The mediators are practically responsible for the implementation and operation of the cooperation, and for their selection the authors exploited the idea of shortest-path betweenness centrality in a distributed setting. Nodes with large SPBC values are selected as mediators to strive for latency reduction and effective control of communication among neighboring nodes. Later, the original scheme was further improved [4] by replacing the SPBC metric with a new one that required less communication for its calculation and avoided some drawbacks of SPBC.

The issue of information dissemination in ad hoc networks — in a less dynamic setting than online caching — was investigated in [14]. The problem addressed was the optimal placement of content in opportunistic networks, which can be expressed as the $k$-median problem when the global network topology and information demand stream are known. The authors developed a scalable near-optimal placement algorithm for the 1-median problem; the algorithm identifies nodes with high *conditional betweenness centrality* (a special case of SPBC) and shrinks the network into a smaller one that contains only those nodes, thus solving a smaller optimization problem.

*Challenges for SNA-Based Information Dissemination* — The main issue with SNA-based information dissemination is that the protocols proposed so far work only for static (or semi-

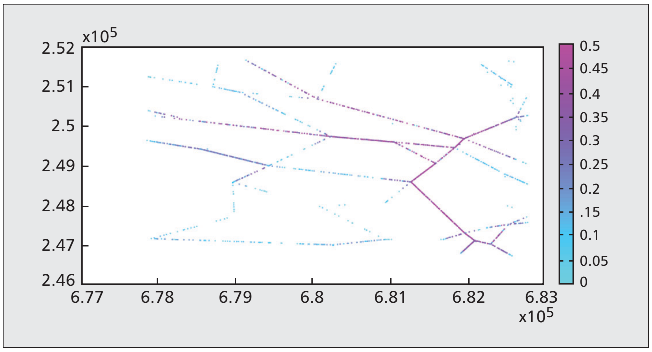


Figure 2. *SPBC of vehicles over geographic location for a specific time-instance (from [8]).*

static) wireless ad hoc networks. Specifically, the *NICoCa* protocol can be seen as an implementation of a dynamic connected dominating set over the underlying sensornet. But when mobility comes into play, the problem becomes harder to address. None of the aforementioned approaches can work efficiently. This leads to the issue of developing solutions for the problem of finding dominating sets for mobile networks that must satisfy some additional constraints (e.g., latency of information dissemination, queue length of the nodes). Moreover, these solutions for the problem of mobile dominating sets must admit distributed solutions (i.e., utilizing only local network connectivity information).

Also, none of the SNA concepts described so far can be used efficiently in the context of a vehicular ad hoc network (VANET). It seems impossible to use the traditional notion of SPBC over a network whose link duration might be a few seconds [8]. Variations of the *DegC* seem more appropriate, but they are not able to capture the rich topology information present in a VANET. Therefore, at this point we also need further research (i.e., new SNA concepts) — not extensions of the traditional ones — to exploit in human-centered VANETs. For instance, under such a centrality metric, in order for one node to influence another over some period of time, there must be a path that connects the source and destination nodes through intermediaries at different times.

## SNA in Network Modeling: Mobility and Topology Characterization

In the previous subsections we have surveyed a number of ways in which wireless networks protocols can make direct use of SNA concepts (i.e., implementing those notions as components of the protocol). SNA concepts have also been used for descriptive purposes, to investigate the connectivity properties and nodes' ranking of a wireless network. The study in [17] focused on wireless mesh networks, and the main question concerned the differentiation between seemingly similar nodes. For instance, if the system administrator of a wireless mesh network (WMN) had to update and reboot a subset of nodes, which of the nodes should s/he update/reboot first? This is clearly a ranking problem for the network nodes, and the authors suggested the concept of bridging centrality for performing that ranking, since bridging nodes are more important from a robustness perspective (as they help to bridge connected components together), and a possible failure of them would increase the risk of a network partition.

For static or semi-static wireless networks it is relatively easy to use off-the-shelf concepts, but when dealing with node mobility, the properties of a network graph are harder to analyze because the connectivity varies rapidly. Therefore, such studies can either examine a series of snapshots of the graph or develop novel methodologies, for example, propose new definitions for the localized coefficient taking into account the time-varying network topology. Hue *et al*. in [15] addressed the problem of distributed detection of communities in delay-tolerant networks (DTNs) without introducing new community concepts, but developing distributed algorithms for identifying *k*-clique and modularity-optimized communities.

Communities formed in the topology of wireless networks as a consequence of human activity were also exploited in the study reported in [16] which contributed a new mobility model for mobile ad hoc networks. Input to the mobility model is the social network of humans carrying the mobile devices. This model allows sets of devices to be grouped together following the pattern of social relationships among humans, which may vary over time. The validation of this mobility model against real traces showed that the synthetic mobility traces were a very good approximation of human movement patterns.

The work [8] examined a series of snapshots of the evolving network graph and investigated the topological properties over time of a simulated VANET. The aim of the study was to provide answers to the general question of what a VANET communication graph looks like over time and space. For this purpose the authors examined various centrality measures and network clustering features to assess the suitability of each metric/feature, and subsequently proposed the use of the appropriate metrics in the design of protocols. In particular, a significant question investigated by that study was whether the examined centrality metrics reveal different patterns of the graph. It became clear that the betweenness, closeness, and bridging centrality indices follow more or less similar distributions. One significant observation was that the road topology alone (e.g., position of junctions) does not determine the positions of possible central nodes. In Fig. 2 we see that nodes with high SPBC (purple) appear at any position in the VANET area, not only at junction positions; therefore, SNA concepts are truly useful for dynamic networks.

*The Road Ahead for SNA-Based Network Modeling* — The mining of time-varying networks cannot be done effectively or efficiently with currently used graph-theoretic tools, since they fail to model the topology changes. Probably other established tools are more appropriate, which have not been exploited so far. For instance, a *tensor* could be used to represent a continuously changing adjacency matrix. With the aid of the tensor, we could answer reachability queries for pairs of nodes, which is useful in the design of message-ferrying protocols for DTNs. Additionally, we could define a centrality metric such as that mentioned earlier. Moreover, the clustering of a tensor could give rise to new notions of communities.

# Further Research

With a retrospective look at Table 1, we see that the networking community has used SNA concepts without contributing significant novel ideas. Its main effort is to develop distributed variants of existing algorithms/concepts and use them subsequently in the design of protocols. From that perspective the shortest-path betweenness centrality seems the most easily understood and handy tool. The interesting question is whether the networking community can offer significantly new concepts and algorithms in the area of SNA, and if the question receives an affirmative answer, what are the possible roads for future research?

In the following we provide evidence that the special problems encountered during the design of communication networks can be a source of inspiration for the researchers working in the synergy between social network analysis and protocol development.

First of all, let us examine the concepts of DegC and SPBC,

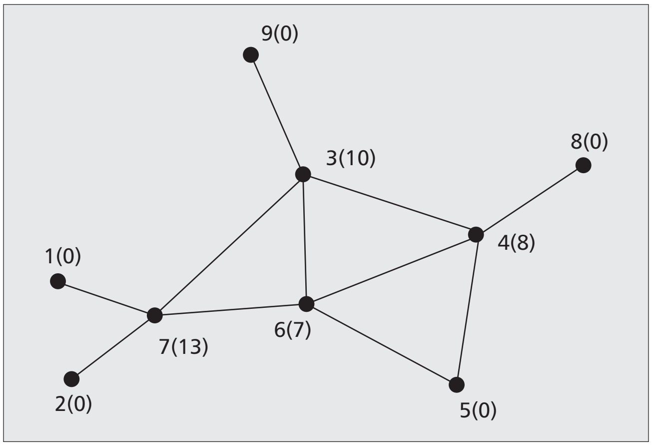


Figure 3. *SPBC values (in parentheses) for a sample graph.*

and compute their values for the nodes of the graph in Fig. 3. Looking at Fig. 3, we see that the nodes 3, 4, 7, and 6 are equally central with respect to their degree; they all have a degree equal to 4. In addition, if we compute the shortest-path betweenness centrality for each node in the whole graph, then node 7 is the most central, followed by nodes 3, 4, and 6. This is somehow counterintuitive, since node 6 *all network nodes* at its vicinity (at a distance two hops away).

This is because SPBC is affected by the number of isolated network nodes, such as node 8 and 9, which increase significantly the SPBC value of the node at the other end of the edge. Starting from this observation, Dimokas *et al.* [4] proposed a new concept of centrality, which can be seen as a distributed approximation of SPBC, called the μ-*Power Community Index* (μ-PCI), defined as follows.

*Definition 1*: The μ-PCI of a node $v$ is equal to $k$, such that there are up to $\mu \times k$ nodes in the μ-hop neighborhood of $v$ with degree greater than or equal to $k$, and the rest of the nodes in that neighborhood have a degree less than or equal to $k$.

Under this new centrality measure and for $\mu = 1$, we see that $PCI(7) = PCI(4) = 2$, whereas $PCI(6) = PCI(3) = 3$, which agrees more with our intuition. This metric was subsequently used to develop a cooperative caching protocol which proved to be better than that based on a distributed version of SPBC [13].

Another shortcoming of most centrality metrics is that they are all *deterministic* (applied in the current network topology snapshot) instead of probabilistic (estimating the probability for a node to contact other nodes in the future). Starting from this observation, the work in [5] developed the *cumulative contact probability* as a new centrality metric. The task of routing in wireless/wired networks can inspire novel ideas for centrality metrics; for instance, generalizations of the shortest-path and flow betweenness can be devised embedding the actual routing schemes. Work should also be done in developing approximation algorithms with proven bounds for calculating centralized centrality metrics using only local information.

Apart from the work in devising centrality or community discovery algorithms, we argue that the majority of effort should be concentrated on the investigation of the time-varying properties of a network and subsequently the development of appropriate concepts. This research area is still in its infancy, and the results are sparse in the networking community. In any case the synergy between complex network science and communication networks will benefit both disciplines.

## Conclusions

This article has surveyed the most important concepts from social network theory that have found significant application in protocol design for routing, information dissemination, and network modeling in the environment of ad hoc networks. The article by no means serves as an exhaustive survey, but as a vehicle to promote understanding and proliferation of ideas, and suggests some roads for possible future work. We envision the synergy between social network theory and ad hoc networking as a fertile research area.

## Acknowledgments

The authors wish to thank the anonymous reviewers for their valuable comments and suggestions.

## References

[1] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Structural Analysis in the Social Sciences Series, Cambridge Univ. Press, 1994.
[2] S. Brin *et al.*, "PageRank Citation Ranking: Bringing Order to the Web," tech. rep. 1999-66, Stanford Univ., Comp. Sci. Dept., 1999.
[3] W. Hwang *et al.*, "Bridging Centrality: Graph Mining from Element Level to Group Level," *Proc. ACM SIGKDD*, 2008, pp. 336–44.
[4] N. Dimokas *et al.*, "High Performance, Low Complexity Cooperative Caching for Wireless Sensor Networks," *Proc. IEEE WOWMOM*, 2009.
[5] W. Gao *et al.*, "Multicasting in Delay Tolerant Networks: A Social Network Perspective," *Proc. ACM MobiHoc*, 2009, pp. 299–308.
[6] B. Bollobas, *Modern Graph Theory*, Springer Verlag, 1998.
[7] N. E. J. Newman, "Modularity and Community Structure in Networks," *Proc. Nat'l. Acad. Sci.*, vol. 103, no. 23, 2006, pp. 8577–82.
[8] G. Pallis *et al.*, "On the Structure and Evolution of Vehicular Networks," *Proc. IEEE/ACM MASCOTS*, 2009, pp. 502–11.
[9] A. Chaintreau *et al.*, "Impact of Human Mobility on Opportunistic Forwarding Algorithms," *IEEE Trans. Mobile Comp.*, vol. 6, no. 6, 2007, pp. 606–20.
[10] E. Daly and M. Haahr, "Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETs," *Proc. ACM MobiHoc*, 2007, pp. 32–40.
[11] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble Rap: Social-Based Forwarding in Delay Tolerant Networks," *Proc. ACM MobiHoc*, 2008, pp. 241–50.
[12] J. M. Pujol, A. L. Toledo, and P. Rodriguez, "Fair Routing in Delay Tolerant Networks," *Proc. IEEE INFOCOM*, 2009, pp. 837–45.
[13] N. Dimokas, D. Katsaros, and Y. Manolopoulos, "Cooperative Caching in Wireless Multimedia Sensor Networks," *Proc. ACM MobiMedia*, 2007, pp. 377–82.
[14] P. Pantazopoulos *et al.*, "Efficient Social-Aware Content Placement in Opportunistic Networks," *Proc. IEEE WONS*, 2010, pp. 17–24.
[15] P. Hui *et al.*, "Distributed Community Detection in Delay Tolerant Networks," *Proc. ACM MobiArch*, 2007.
[16] M. Musolesi and C. Mascolo, "Designing Mobility Models Based on Social Network Theory," *ACM Mobile Comp. Commun. Rev.*, vol. 11, no. 3, 2007, pp. 59–70.
[17] S. Nanda and D. Kotz, "Localized Bridging Centrality for Distributed Network Analysis," *Proc. IEEE ICCCN*, 2008.
[18] P. Hui and J. Crowcroft, "How Small Labels Create Big Improvements," *Proc. IEEE PERCOM Wksps.*, 2007.

## Biographies

DIMITRIOS KATSAROS (dkatsar@inf.uth.gr) is currently a lecturer at the University of Thessaly, Greece. His research interests include distributed systems, such as the Web and Internet, social networks, mobile and pervasive computing, and wireless ad hoc and wireless sensor networks. He has organized special issues on cloud computing in *IEEE Internet Computing* and *IEEE Network*. He has a Ph.D. in informatics from Aristotle University of Thessaloniki, Greece.

NIKOS DIMOKAS (dimokas@delab.csd.auth.gr) received his B.Sc. and M.Sc. in computer science from the University of Crete, Greece, in 2001 and 2004, respectively. He received his Ph.D. from Aristotle University of Thessaloniki in 2010. Currently, he is a postdoctoral researcher at the Department of Informatics of Aristotle University of Thessaloniki. His research interests include wireless sensor networks and vehicular ad hoc networks.

LEANDROS TASSIULAS [F] (leandros@uth.gr) obtained his B.Eng. from Aristotle University of Thessaloniki in 1987, and his M.Sc. and Ph.D. degrees from the University of Maryland, College Park in 1989 and 1991, respectively. He is a professor in the Department of Computer and Telecommunications Engineering, University of Thessaly. His research interests are in the field of computer and communication networks with emphasis on fundamental mathematical models of communication networks, architectures and protocols of wireless systems, sensor networks, high-speed Internet, and satellite communications. He received an NSF Research Initiation Award in 1992, an NSF CAREER Award in 1995, an ONR Young Investigator Award in 1997, a Bodosaki Foundation award in 1999, and the IEEE INFOCOM '94 best paper award.