New metrics for characterizing the significance of nodes in wireless networks via path-based neighborhood analysis

Leandros A. Maglaras and Dimitrios Katsaros
Dept. of Computer and Communication Engineering, University of Thessaly, Volos, Greece
maglaras@uth.gr, dkatsar@inf.uth.gr http://www.inf.uth.gr/~dkatsar

***Abstract*— This paper considers the problem of finding the most central nodes in neighborhoods of a given network with directed or undirected links taking into account only local information. An algorithm that calculates ranking, taking into account the n-hop neighborhood of each node is proposed. The method is compared to popular existing schemes for ranking, using Spearman's rank correlation coefficient and other metrics. An extension to a faster algorithm which reduces the size of the examined network is described as well.**

***Index Terms*—Centrality metrics, localized algorithms, node ranking, network analysis, wireless sensor networks.**

I. INTRODUCTION

Network topology analysis refers to the process of characterizing the physical connectivity and the relationships among entities in a communication network. Among the most significant tasks involved in topology analysis is the calculation of centrality measures [1]. Point centrality in communication is based upon the concept of betweenness, first introduced in [2]. According to betweenness centrality a node is central to the degree that it stands between others, thus playing a significant role in message passing. PageRank [3] is another very popular method for measuring centricities in social networks; the basic idea behind PageRank is that a node is significant if it is connected to other significant nodes. Various other measures of centrality and ranking have been proposed to determine the importance of a vertex within a graph [4].

These indices are of great value in the understanding of the roles played by actors in social networks, and by the vertices in networks of other types (Web, Internet, Food/Sex web) but they are not always accurate, since localized metrics are often required in order to better describe relations between nodes. Localized metrics are particularly useful in the analysis of wireless networks; they present a potential for control of communication, safety issues [5], routing protocols [6], information dissemination [7, 8].

When these centrality metrics are to be used in wireless networks, they suffer from several shortcomings. In summary, betweenness centrality suffers from the fact that it leaves many nodes unranked, since these nodes don't participate in any shortest paths computed. Moreover, the existence of *bridge edges* in the network graph, result in increasing at an excessive amount the centrality value of the *articulation* node without this node being really "important". Similarly, PageRank suffers from the fact that nodes may be ranked very high due to the fact that they are adjacent to significant nodes even though they play no specific role in packet forwarding (e.g., the *sink* nodes). Moreover, the computation of PageRank requires cumbersome calculations and knowledge of the whole network topology, which is not possible in ad hoc wireless networks that require *localized* algorithms.

The present work is motivated by the design of protocols in wireless networks that seek for nodes "central" in the network to assign to them special roles, e.g., *mediator* nodes in cooperative caching for sensor networks [7,8], *message ferrying* nodes in Delay Tolerant Networks [6], *rebroadcasting* nodes in vehicular networks [9], and so on.

In this paper a novel metric for calculating the centrality of vertices in networks is proposed. The basic idea is that the centrality of a node is to be calculated over its neighborhood. In this reduced graph, all the paths connecting the considered node with all the vertices of the neighborhood are found and a local weight is computed. Local weights are accumulated to give a global measure of centrality and a global ranking of nodes. The new metric called "*Aggregated Weight N-hop Ranking – AWeNoR*" not only rewards vertices that belong to many neighborhoods, but also rewards those ranked high in the neighborhoods they belong to. Due to this attribute no nodes (except from the isolated ones) remain unranked.

The remainder of this paper is structured as follows: in Section II, the network model, the assumptions and the AWeNoR ranking technique are described. Section III shows the results of the comparison of AWeNoR to other centrality metrics. Section IV introduces another faster technique for computing global rankings through local weight, and the article concludes with Section V.

978-0-7695-3788-7/09 $25.00 © 2009 IEEE
DOI 10.1109/PCI.2009.12

II. THE AWENOR RANKING METHOD

The basic idea behind the proposed method is to create each node's neighborhood and compute the local weights in this subgraph. All these weights are then accumulated in order to give the final rank of each vertex. In Subsection A we describe the algorithm for this method, Subsection B shows how local weights are calculated, and Subsection C demonstrates how the final rankings are computed by aggregating the local weights.

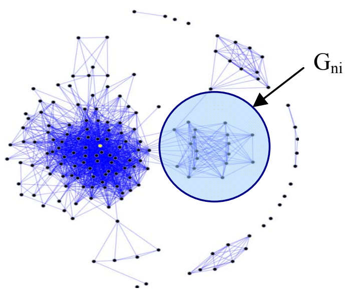

Figure 1. Graph G and a neighborhood G_{ni}.

A. The n-hop neighborhood

We consider a network G=(V,L), where V is the set of nodes (vertices) and L is the set of links (edges). Each link can be undirected or directed having weight equal to 1. Each node is given a distinct id starting (from the value one).

Definition 1. *A node j belongs to neighborhood G_{ni} of the node i, if there exists at least one path from the starting vertex i to the end vertex j, in at most n-hops away.*

In order to compute the ranking of each node, the proposed method operates as follows:

1. Find the N-hop neighborhood G_{ni} of each node i.
2. Find all the paths from node i to every other node j of the neighborhood.
3. Calculate the local weight of all the nodes in G_{ni} (except from i) according to the AWeNoR method [explained later].
4. Accumulate local weights to obtain the final (global) ranking of all the nodes.

B. Local weight

The AWeNoR algorithm aims at computing the local weights of nodes which belong to G_{ni} neighborhoods. There are two intuitions behind this algorithm. Firstly, the nodes closer to the starting node of a path are more crucial than the more distanced ones, with respect to disseminating information to the rest of the network. Secondly, all paths can be used to pass data in a neighborhood and not only the shortest path, as used by the betweenness centrality when it calculates node rankings.

The algorithm for computing local weight proceeds by deriving all paths with starting vertex i. The paths are specified as $P_i^k = (u_i^0, u_i^1, \ldots, u_i^N)$ where P_i^k is the k_{th} path from start vertex i, and u_i^j is a node at a j-hop distance from the start vertex i. For each hop, a weight is computed for each vertex (*l*) using equation (1).

$$W_l^j = \frac{a_{lj}}{K}, \quad K > 0 \quad (1)$$

where K is the total number of paths derived from the previous step of the algorithm and a_{lj} shows the number that vertex *l* appears in hop j.

The local weight for any vertex in neighborhood G_{ni} is computed using equation (2).

$$b_l^i = \sum_{\forall j} \frac{W_l^j}{j} \quad , \forall l \in G_{ni} \quad (2)$$

The size of the neighborhood is a parameter that plays a significant role. Taking N equal to the network diameter, the neighborhoods coincide with the network graph G. In that case, in order to compute the ranking of a node, all paths between vertices have to be found, thus making the algorithm inappropriate even for medium sized networks. On the other hand, giving to N a very small value, the obtained rankings may not be very representative at all.

C. Global ranking

The algorithm AWeNoR computes local weights for all nodes that belong to a neighborhood G_{ni}. Since nodes may belong to multiple neighborhoods, the local weights have to be accumulated in order to obtain the final (global) ranking of the vertex using the equation (3).

$$b_l = \sum_{\forall G_{ni}} b_l^i \quad , \forall l \in G \quad (3)$$

It must be stated that only non acyclic paths are used from AWeNoR in order to compute local weights. Also, in every neighborhood G_{ni}, the local weights are calculated for every vertex that belongs to G_{ni}, except from i itself, since its weight, using equation (1), would be equal to one.

Time complexity of the method can be expressed as O((|L|+|V|)*|V|) since every vertex and every edge will be explored in the worst case, for each neighborhood created. Parameter | L | is the cardinality of the set of edges (the number of edges), and | V | is the cardinality of the set of vertices.

III. EVALUATION OF THE PROPOSED METHOD

In order to evaluate the proposed ranking technique we used two real datasets, and since there are not available network graphs describing real wireless networks, we chose data sets coming from real social networks. The real graphs have large connectivity among nodes. Networks with both undirected and directed links were used. The visualization of the networks was performed with Pajek (http://vlado.fmf.uni-lj.si/pub/networks/pajek/) and the calculation of the betweenness and PageRank centrality values of the network nodes was done with the aid of CentiBiN (http://centibin.ipk-gatersleben.de/). The real graphs are the following:

- *Zachary's karate club*: a network of friendships between 34 members of a karate club at a US university in the 1970 [10].
- *Dolphin network*: an undirected network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand [11].

Except from the AWeNoR centrality values, the betweenness and the PageRank centrality values were also computed for every graph in order to compare them. For every graph ranking, we measure the number of ties that each ranking algorithm produces and also we compute the Spearman's rank correlation coefficient (Eq. 4) between pairs of ranking algorithms. The more ties an algorithm produces, the less useful the ranking is for use in wireless networks because it is does not discriminate among network nodes. Spearman's is a non-parametric measure of correlation widely used to describe the relationship between two variables that is used to report the difference in ranking produced by two methods[1]. In our case, this metric is used to evaluate the proposed metric in relation to PageRank and betweenness centrality values.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad ,\forall i \in G \quad (4)$$

A. Undirected experimental graphs

The first real graph, the Zachary's karate club is shown in Figure 2 and the Dolphins graph is depicted in Figure 3. The visualization is used here as a means to confirm the obtained results with the human intuition.

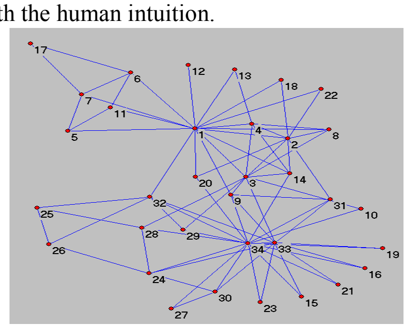

Figure 2. Zachary's karate club undirected graph.

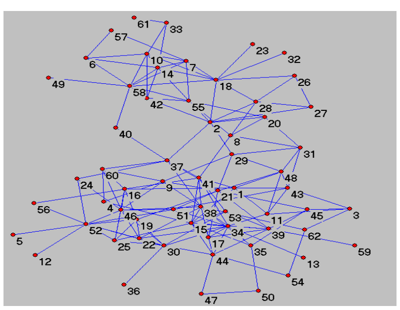

Figure 3. The Dolphins network.

TABLE 1 shows the total number of ties that each of the three methods produces for the two networks. The numbers in parentheses represent the number of vertices with zero centrality value (non-ranked). It can be seen that the betweenness centrality metric produces a significant amount of non-ranked nodes, which is a non desirable effect when the centrality metrics is used in wireless networks for characterizing the significance of nodes in the network topology.

TABLE 1. The number of ties produced by each competitor.

	Betweenness	PageRank	AWeNoR
Zachary's karate club	16 (12)	11 (0)	11 (0)
Dolphin social network	9 (9)	4 (0)	4 (0)

TABLE 2 shows the Spearman's rank correlation coefficient computed for every pair of rankings. In the Dolphins dataset, we can observe significant discrepancy in the rankings produced by AWeNoR with those produced by PageRank.

TABLE 2. Spearman's rank correlation coefficient.

	Betweenness – AWeNoR	PageRank – AWeNoR	Betweenness – PageRank
Zachary's karate club	0,8442	0,8512	0,8747
Dolphin social network	0,7712	0,9457	0,8171

TABLE 3 shows the biggest rank difference observed between the three methods. For the Dolphin network, where the number of nodes is relatively large, it is observed that the AWeNoR gives results close to PageRank. The numbers in parentheses represent the vertex where the biggest difference is observed. TABLE 4 depicts the highest ranked vertices for each graph. We observe that AWeNoR makes similar to PageRank rankings for the top-ranked nodes, even though it is a localized metric whereas PageRank requires cumbersome computations and knowledge of the whole network's topology.

TABLE 3. Biggest difference observed.

	Betweenness - AWeNoR	PageRank - AWeNoR	Betweenness - PageRank
Zachary's karate club	14(26)	12(10)	13(10)
Dolphin social network	38(40)	13(17)	41(40)

TABLE 4. Highest ranked nodes for Karate (top) and Dolphins (bottom).

RANK POSITION	PageRank	AWeNoR	AWeNoR Reduced	Betweenness
1st	34	34	34	1
2nd	1	1	3	34
3rd	33	33	33	33
4th	3	3	2	2
5th	2	2	1	32
RANK POSITION	PageRank	AWeNoR	AWeNoR Reduced	Betweenness
1st	15	15	15	37
2nd	18	38	58	2
3rd	52	46	18	41
4th	58	34	34	38
5th	38	52	44	8

B. Experimental directed graphs

The new ranking technique was also tested in directed graphs. Taking the Karate club real graph and converting each edge to a directed arc, the network of Figure 4 is created.

[1] http://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient

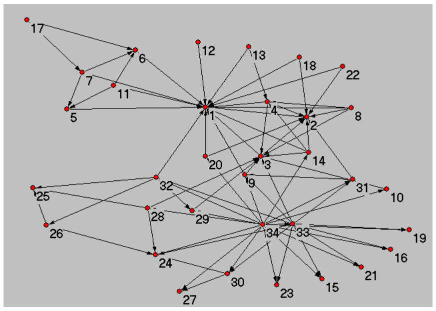

Figure 4. Zachary's karate club directed graph.

TABLE 5 shows that AWeNoR incurs significantly fewer ties than PageRank does. (Betweenness centrality is not possible to be computed for this network due to the lack of strong connectivity.)

TABLE 5. Number of ties incurred by each algorithm.

	PageRank	AWeNoR
Zachary's karate club	21	15
Dolphin social network	23	22

TABLE 6 depicts the ids of the five highest ranked (top-5) nodes for this directed network. The numbers in parentheses represent the position of the node in the ranking produced by the competitor method, in the cases where this node does not appear in the top-5 list of the competitor.

From Figure 4 and TABLE 6, we gain an insight why the AWeNoR algorithm is a "better" algorithm in determining the most significant nodes compared to PageRank: for instance, node 33 (ranked 4th) is more crucial in terms of routing than node 19, which is a sink node. The three highest ranked vertices are the same for both methods in the Karate club graph. Of course, such an observation is not a proof of the superiority of the algorithms, but it s a strong evidence that produces meaningful rankings.

TABLE 6. Highest ranked nodes for Karate (top) and Dolphins (bottom).

RANK POSITION	PageRank	AWeNoR
1st	1	1
2nd	3	3
3rd	2	2
4th	19 (24th)	33 (7th)
5th	4 (6th)	9 (8th)
RANK POSITION	PAGERANK	AWENOR
1st	1	1
2nd	15	15
3rd	16 (4th)	38 (23rd)
4th	4 (6th)	16 (3rd)
5th	19 (8th)	46 (33rd)

In the Dolphins network (Figure 3), the nodes with id 38 and 46, which are among the highest ranked by AWeNoR, have very low ranking position in PageRank metric. This is due to the fact that AWeNoR ranking rewards nodes that belong to many neighborhoods, though PageRank only those connected to significant nodes. PageRank may rank in high position those vertices that have few (even just one) neighbor, which is significant to the network, without examining if they play any role in larger neighborhoods, which is desirable by policies applied to wireless networks.

IV. THE AWENOR REDUCED RANKING

As described in Section III, in order to compute the aggregated weights, the AWeNoR algorithm has to add local weights of all neighborhoods in the network. So, for a K-hop-long network, the AWeNoR algorithm has to run K times, one for each vertex. Computing local weights for every neighborhood can be a very time consuming even for medium sized networks. In order to improve the total running time of the original AWeNoR algorithm, we further proposed here the AWeNoR-Reduced ranking method.

The AWeNoR Reduced algorithm creates neighborhoods only for some vertices according to a parameter q_i and a threshold A. Parameter q_i is used to count the times that vertex i participates in paths of all the neighborhoods created by the algorithm in every step. The AWeNoR Reduced ranking algorithm is given below:

1. Initiate algorithm. Set i=1.
2. Find the AWeNoR neighborhood G_{ni} of node i satisfying the constraint q_i<A.
3. Find all the paths from node i to every node j in the neighborhood.
4. For every path $P_i^k = (u_i^0, u_i^1, ..., u_i^N)$ update parameter q_{ui} for $\forall$ u$\in$ P_i^k except u_i^N and u_i^0.
5. Calculate the local weight of all the nodes in G_{ni} (except from vertex i) according to the AWeNoR algorithm.
6. Set i=i+1. If the last node of graph is reached, then go to step 7 else go to step 2.
7. Accumulate local weights to obtain the final ranking of all the nodes.

The AWeNoR Reduced algorithm according to TABLE 7 and TABLE 4 seem to work well in terms of finding most important vertices in a graph while fewer neighborhoods need to be created (TABLE 8).

TABLE 7. Spearman's rank correlation coefficient.

Undirected graphs	Betweenness – AWeNoR Reduced	Pagerank – AWeNoR Reduced	AWeNoR – AWeNoR Reduced
Zachary's karate club	0,8105	0,8438	0,9175
Dolphin social network	0,7925	0,9207	0,8782

The parameter A is used as a threshold in order to choose whether a vertex's neighborhood is created or not. Choosing the value of parameter A is an important issue. Giving A a rather big value AWeNoR Reduced algorithm degenerates to AWeNoR, since all neighborhoods are created. Setting A equal to zero, a risk of creating disjoint neighborhoods arises letting some nodes unranked. In the experiments conducted, a value close to zero was used in order to avoid these situations.

TABLE 8. Neighborhoods created.

Undirected graphs	AWeNoR Reduced (A=1)	AWeNoR Reduced (A=3)	AWeNoR
Zachary's karate club	3	7	34
Dolphin social network	14	17	62

Figure 5 shows the effect of parameter A to method's results compared to AweNoR, along with the number of neighborhoods created for every such choice. A strict relation between method's accuracy and cost, in terms of time consumption, is observed.

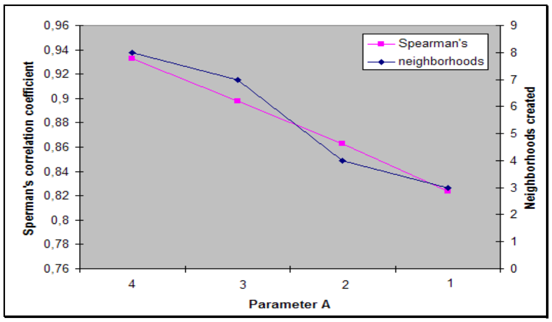

Figure 5. Sensitivity of AweNoR reduced ranking to parameter A (Zachary's karate club undirected graph)

The preferred policy is to have a value that changes according to the size or connectivity of the network, but its development is a subject of future work.

V. CONCLUSIONS AND FUTURE WORK

The issue of discovering which nodes in a wireless network are central to the topology is of fundamental importance, since it can be used as a primitive to perform routing [6], cooperative caching [9] and contamination detection.

Several metrics exist in the literature, like betweenness centrality, PageRank. Betweenness is based on shortest paths between vertices. Vertices that lie on many shortest paths between other vertices are given a high centrality value. In many cases however this metric is not realistic due to the fact that it counts only a small subset of all the paths. When PageRank is used the significance of a vertex comes from the significance of its 1-hop neighborhoods leading many times to misleading results. A sink vertex may be ranked very high just because it is adjacent to a very significant node, even though its contribution to communication is of no importance.

In this paper we proposed a new measure for determining significant nodes. For each vertex i a neighborhood is created and all paths with starting vertex i are created. For every "cluster" created a local weight is computed and a final ranking measure is created by adding local these weights. The new measure rewards vertices that belong to many neighborhoods and lie in many paths between vertices of the neighborhood. The measure was compared to Betweenness and PageRank and seems to work well for both directed and undirected networks.

The AWeNoR Reduced, a faster algorithm for finding centrality, was also presented. This refined AWeNoR algorithm creates fewer neighborhoods but the results are sensitive to some parameters. The initial neighborhood created seems to affect the final results. The threshold that defines which neighborhoods are created is also very crucial and an optimal value has to be found for every network.

Comparison of the method to other measures like [12] or [13] have to be conducted in the future where also special cases have also to be tested. The main goal though of our future work is to use this centrality metric as a primitive in the design of networking protocols, like cooperative caching for ad hoc wireless networks [7, 8].

REFERENCES

[1] P.F. Bonacich (1987). "Power and centrality: A family of measures". *American Journal of Sociology*, vol. 92, pp. 1170–1182.

[2] A. Bavelas (1948). "A mathematical model for group structure". *Human Organizations*, vol. 7, pp. 16–30.

[3] L. Page, S. Brin, R. Motwani, and T. Winograd (1998). "The PageRank citation ranking: Bringing order to the Web". *Technical Report*, Stanford Digital Library Technologies Project.

[4] L.C. Freeman (1979). "Centrality in social networks: Conceptual clarification". *Social Networks*, vol. 1, pp. 215–239.

[5] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance (2007). "Cost-effective outbreak detection in networks". *Proceedings of the* ACM *KDD Conf.*, pp. 420–429.

[6] P. Hui, J. Crowcroft, E. Yoneki (2007), "BUBBLE Rap: Social-based forwarding in Delay Tolerant Networks". *Proceedings* of the ACM MOBIHOC Conf., pp. 241–250.

[7] N. Dimokas, D. Katsaros, and Y. Manolopoulos. (2008). "Cooperative caching in wireless multimedia sensor networks". *ACM Mobile Networks and Applications*, vol. 13, no. 3-4, pp. 337–356.

[8] N. Dimokas, D. Katsaros, L. Tassiulas, and Y. Manolopoulos, (2009). "High performance, low overhead cooperative caching in wireless sensor networks". *Proceedings of the IEEE WoWMoM Conf.*

[9] M. Zhang and R.S. Wolff (2008). "Routing protocols for vehicular ad hoc networks in rural areas". *IEEE Communications magazine*, vol. 46, no. 11, pp. 126–131.

[10] W. W. Zachary (1977). "An information flow model for conflict and fission in small groups", *Journal of Anthropological Research*, vol. 33, pp. 452–473.

[11] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson (2003). The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. Can geographic isolation explain this unique trait?, *Behavioral Ecology and Sociobiology*, vol. 54, 396–405.

[12] L.C. Freeman, S.P. Borgatti and D.R. White (1991). "Centrality in valued graphs: A measure of betweenness based on network flow". *Social Networks*, vol. 13, pp. 141–154.

[13] M. Newman (2005). "A measure of betweenness centrality based on random walks". *Social Networks*, vol. 27, pp. 39–54.