CONCURRENCY AND COMPUTATION: PRACTICE AND EXPERIENCE
Concurrency Computat.: Pract. Exper. 2013; **00**:1–13
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cpe

Adaptive Method to Support Social-based Mobile Networks Using a PageRank Approach

Florin POP, Radu-Ioan CIOBANU, Ciprian DOBRE*

University POLITEHNICA of Bucharest, Faculty of Automatic Control and Computers, 313 Splaiul Independentei, Sector 6, 060042 Bucharest, Romania, Emails: florin.pop@cs.pub.ro, radu.ciobanu@cti.pub.ro, ciprian.dobre@cs.pub.ro

SUMMARY

Opportunistic networks are mobile networks that rely on the store-carry-and-forward paradigm, using contacts between nodes to opportunistically transfer data. For this reason, traditional routing mechanisms are no longer suitable. The use of additional routing criterion, such as social information about nodes, can increase the probability of successful message delivery. Popularity of a node, another important routing criterion, can be inferred using the betweenness centrality, meaning the number of times the node is on the shortest path between any other two nodes in the social graph. However, computing the betweenness centrality is impossible in practice, especially when connectivity between individuals is transient, and each node has only a local view of the entire network. We propose a fundamental rethinking, where nodes and not paths are the observation focus. In our approach, we compute the probability of a node to participate in a path formation (e.g., the probability of a node to lead to the next popular path). We present our solution, which takes inspiration from the PageRank approach, and present an algorithm to compute and update the popularity of nodes using the probability of each node to be used as carrier for random messages traversing the network. We demonstrate that this approach is highly robust, numerical insensitive to errors and converges fast, meaning it can be easily adopted in resource-constraint environments formed between highly-mobile wireless devices. Our experimental results sustain our empirical observations for various case studies. Copyright © 2013 John Wiley & Sons, Ltd.

Received . . .

KEY WORDS: mobile network; social-based mobility; PageRank; path formation; convergence

1. INTRODUCTION

Ever since mobile devices such as smartphones and tablets have become ubiquitous, there has been a steady increase in research concerning mobile networks. One particular case is represented by opportunistic networks (ONs), which are dynamically formed when mobile devices collaborate to form communication paths while users are in close proximity. In opportunistic networks, mobile devices are data providers, data receivers, as well as data transmitters; they act according to a *store-carry-and-forward* paradigm [25]. This implies that an ON node stores a message, carries it around until it encounters its destination or another node that has a higher chance of reaching the destination, and then forwards it.

The main challenge of routing and dissemination in opportunistic networks is deciding whether an encountered node is suitable for transporting a given message closer to its intended recipient. Since ONs are decentralized, each node (which is a mobile device) only has information it gathers from contacts with other nodes. Thus, a node is not aware of the topology of the network. Furthermore,

*Correspondence to: University POLITEHNICA of Bucharest, Faculty of Automatic Control and Computers, 313 Splaiul Independentei, Sector 6, 060042 Bucharest, Romania, Email: ciprian.dobre@cs.pub.ro

we cannot speak of an actual topology, since nodes are highly mobile and the shape of the network is rapidly changing.

Given that nodes in an opportunistic network are mobile devices carried by humans, the social habits and relationships of the carriers play an important role in the behavior of the network. Therefore, several of the previous solutions for opportunistic routing and dissemination, like BUBBLE Rap [13], have focused on using the popularity of nodes (with algorithms borrowed from social networks theory). In such social-driven solutions, popular nodes are considered to have a more prolific role in the network than others, because they are more prone to have encounters with other nodes, and so they are able to spread more messages to other nodes in the network. In previous work, the decision of whether a node is popular or not is given by its betweenness centrality, which is a term borrowed from the social networks theory. It signifies the number of times a node is on the shortest path between two other nodes in the social graph. However, computing the betweenness centrality according to its definition is almost impossible in practice, because it requires a global view of the entire social network formed by the nodes in the ON. Methods such as measuring the betweeness centrality in parts (on clusters of nodes) still exist, but as previous studies have shown, they can lead to non-optimum local results [26]. Various approximation algorithms have been also used, such as single-window (S-window) or cumulative-window (C-window). The S-window algorithm computes centrality as the number of encounters the current node has had in the last time window, while the C-window algorithm counts the number of individual nodes encountered for each time window and then performs an exponential smoothing on the cumulated values. Other proposed solutions [8] use the degree centrality of the node in the social graph instead, which is the number of social links the device's owner has with the other participants in the ON.

In this paper, we propose a novel method of computing and updating a node's centrality based on the PageRank algorithm [24] used by many popular search engines to compute the popularity of Web pages. In practice, this can mean that the popularity of a node is given by its chance to participate in a path formation (e.g., node to lead to the next popular path). The algorithm assigns a numerical weighting to each element of a hyperlinked set of entities (e.g., the Worldwide Web), with the purpose of "measuring" their relative importance within the set. For opportunistic networks, we analyze the link between nodes in the network in order to establish the centrality of a node. We assign a numerical weighting to each link in the network denoting the probability for a node to be considered in a virtual circuit formed for data delivery. The algorithm formed this way offers numerical stability and fast convergence in its implementation. The adaptive, stable and fast convergent approach applied to centrality computation and considering all changes in a network ensure fast decisions in routing and data delivery processes. The PageRank approach used for centrality computation is new in opportunistic networks analysis. We will present in Related Work section the existing approaches.

With this, the contribution of this paper is twofold: 1) we first present the proposed PageRank algorithm to update the popularity of nodes between nodes using the probability of each link to be traversed by messages circulating within the network; 2) we next demonstrate that our approach is highly robust, numerical stable (theoretical proved) and converges fast, which means that is can be easily adopted in resource-constraint environments formed between highly-mobile wireless devices.

The rest of the paper is structured as follows. Section 2 shows how previous methods make use of a node's centrality for routing and dissemination in opportunistic networks and how they compute it. Section 3 presents our proposed method for estimating and updating a node's centrality based on the PageRank algorithm, together with demonstrations of its stability and good convergence rate. Section 4 highlights the experimental results of our method. Finally, Section 5 presents our conclusions.

2. RELATED WORK

A thorough review of opportunistic networking is presented in [9]. The analysis, developed in the context of the EU Haggle project, highlights the properties of main networking functions, including message forwarding, security, data dissemination and mobility models. The authors also

DOI: 10.1002/cpe

propose various solutions for communication in ONs, such as PROPICMAN [22] or HIBOp [3]. Other well-known opportunistic routing and dissemination solutions include the Socio-Aware Overlay [32], Wireless Ad Hoc Podcasting [20] or ContentPlace [4].

A taxonomy for data dissemination algorithms splits them in four large categories [6]. The first category deals with the infrastructure of the network, meaning the way the network is organized into an overlay for the nodes. Then, the dissemination techniques are also split according to the characteristics of their nodes, such as node state and node interaction (which includes node discovery, content identification and data exchange). The third category of the taxonomy is represented by content characteristics, meaning the way content is organized and analyzed, and finally the last category (and the most important one) is social awareness. Social awareness is considered to be the future of opportunistic networks, because the nodes in such a network are mobile devices carried by humans, which interact with each other according to social relationships.

Several previous methods for routing and dissemination in ONs use the popularity of a node given by its centrality as a means of deciding which nodes are more likely to transport messages to their intended destinations. BUBBLE Rap [13] is such an algorithm which assumes that a mobile device carrier's role in the society is also true in the network. Thus, the first part of the algorithm is to forward data to nodes that are more popular than the current node. The second assumption made in BUBBLE Rap is that the communities people form in their social lives are also observed in the network layer, therefore the second part of the algorithm is to identify the members of the destination community and pass them the message. Therefore, a message is bubbled up the hierarchical ranking tree using a global popularity level, until in reaches a node that is in the same community as the destination. Then, the message is bubbled up using a local ranking until it reaches the actual intended recipient of the message. The popularity of a node is given by its betweenness centrality, which is the number of times a node is on the shortest path between two other nodes in the network. Community detection is done using k-CLIQUE [14], an algorithm that dynamically detects the community of a node by analyzing its encounters with other devices. There are two reasons that modularity can't be used for comunity detection in opportunistic networks. First of all, it is applied on static social networks, where the social links between nodes and their strengths are known in advance. However, in ONs, there are no social links to begin with, and nodes move around the network and interact with each other dynamically. Therefore, a distributed method of dynamically computing these communities is needed, and this is the reason why k-CLIQUE is used. Secondly, the modularity method is unable to correctly detect small communities, since it suffers a resolution limit. There are two important parameters to the k-CLIQUE algorithm: the contact threshold and the community threshold. The contact threshold specifies the amount of time that two nodes have to be in contact before being considered as part of the same community, while the community threshold is used to specify the number of community nodes two encountering devices must have in common in order for them to belong to the same community. The centralities are computed by replaying the last collected mobility trace, applying a flooding algorithm and then computing the number of times a node acts as a relay on a shortest path. However, this implementation of BUBBLE Rap proves to be unfeasible in real life. Therefore, a distributed version entitled DiBuBB was also proposed by the authors [13]. It uses distributed k-CLIQUE for community detection and a C-window or S-window algorithm for centrality computation.

Another method that uses the popularity of a node in the routing process is presented in [8], where four improvements to the original BUBBLE Rap algorithm are presented. They are entitled Social, Max, Popularity and Popularity Squared, and are based on the supposition that using information about the social relationships between the members of an opportunistic network instead of computing communities on-the-fly with k-CLIQUE can increase the effectiveness of routing. The Social method uses information about two nodes' social connections (taken from places such as Facebook or Google+) to check if two nodes are in the same community or not, in order to decide whether to use the local or the global centrality. The Max version of BUBBLE Rap computes a node's popularity as the maximum between its degree and its betweenness centralities (both normalized to the number of nodes in the network). The degree centrality is defined as the number of social connections a node has inside the ON [21]. The Popularity version uses a weighted sum of

DOI: 10.1002/cpe

the degree and the betweenness centralities as the popularity of a node, while the Popularity Squared version uses a weighted squared sum.

However, using a globally-defined node centrality may result in a bias towards the most popular nodes, which may not always be suitable candidates for forwarding messages intended for specific recipients. This may happen because the popular nodes may have low importance levels relative to some subsets of target nodes, which means that messages may reach those destinations with high delays (or not at all). For this reason, a method entitled OFPC [33] has been proposed, that performs opportunistic forwarding with partial centralities (defined as the relative importance of nodes in relation to subsets of the ON).

The efficiency of routing in ONs using centrality metrics has also been studied in recent years [23, 31]. It has been shown that routing using the classic centrality values as defined in social networks theory is not optimal because of three main reasons. First of all, the routing decisions thus made are greedy and agnostic to the destinations of the messages. Secondly, the performance of the resulting algorithms is very sensitive to the graph of encounters over which the centrality is computed, and finally, the global centrality values have to be approximated for practical reasons. This is why we propose in this paper a PageRank-based method for computing and dynamically update a node's centrality, since we show that it converges, and the method proposed here also computes the probability that, if two nodes are connected, they can deliver a message at a given point in time.

The PageRank updating algorithm was analyzed and improved in previous works. Ipsen and Kirkland show in [15] that the an iterative aggregation/disaggregation method is at least as good as that of the power method. Furthermore, by exploiting the hyperlink structure of the network it can be shown that the asymptotic convergence rate of this method applied to the Google matrix can be made strictly faster than that of the power method. So, this is a strong pillar for our proposed updating algorithm. A mathematical analysis of PageRank when α changes was given by Boldi, Santini and Vigna in [2]. They show that, for real-world graphs values of α close to 1 do not give a meaningful ranking. They also prove that the k-th iteration of the Power Method gives exactly the value obtained by truncating the PageRank power series at degree k. This is a very important results regarding method convergence. The fundamental properties of PageRank concerning stability, complexity of computational scheme, and critical role of parameters involved in the computation were given by Bianchini, Gori and Scarselli in [1]. They present an analysis of the distribution of the page score, the role of dangling pages (pages with no outlinks), and the secrets for promotion of Web pages.

3. ADAPTIVE PAGERANK METHOD FOR ESTIMATING CENTRALITY

In this section we introduce the mathematical model used for centrality computation and, more important, we prove that the proposed method is stable and has a good convergence rate. We use the model next to propose an improved algorithm over the standard iterative for computing the node centrality.

Definition 1. Let's consider an opportunistic network, $OpNet$, with maximum n nodes. The Markov model of $OpNet$ represents the graph encoded as a square matrix P whose element p_{ij} is the probability of interaction of node i with node j (interaction initiated by node i).

Definition 2. The *centrality* of a node within a graph is a measure to determine its relative importance in that graph (i.e. how influential a person is within a social network, or how well-used a road is within an urban network [10]). Borrowed from social network analysis, there are four measures of centrality that are widely used today in network analysis: *degree centrality*, *betweenness*, *closeness*, and *eigenvector centrality*. For example, the measure *betweenness centrality* [21] signifies the number of times a node is on the shortest path between two other nodes in the social graph.

We consider P to be a valid probability matrix. This means that for every node must have at least 1 outgoing transition (there is no fully isolated node that can participate in an $OpNet$), so, P should have no rows consisting of all zeros. The interaction probability matrix P is build from the

DOI: 10.1002/cpe

structure of the opportunistic network, $OpNet$, to be stochastic and primitive. A link from a node $i \in OpNet$ to another link $j \in OpNet$ can be seen as an evidence that j is an important node, and this assumption will be used for centrality computation.

Definition 3. The rank of a node n in a $OpNet$ is defined as the probability that at some particular time the node n is considered as a part of a virtual circuit formed for data delivery in $OpNet$.

After a large period of time we can consider that this probability is unique, so it can be used with success as a metric for nodes evaluation. In real-world, is it possible for a node to have no interaction for a period of time. In that case, if in the P matrix some rows contain all zeros, we will replace all zeros rows with $\frac{1}{n}e^T$, where $e = [1]_{n \times 1}$ (a column vector of all ones) and n is the order of the matrix (the number of nodes).

Definition 4. Let's consider d a $n \times 1$-dimensional column vector identifying the nodes with out-degree 0 ($deg(i)$ represents the degree of a node):

$$d(i) = \begin{cases} 1, & \texttt{if deg(i)=0} \\ 0, & \texttt{otherwise} \end{cases} \tag{1}$$

Therefore, we will compute a new non-negative and irreductible matrix $P' \in R^{n \times n}$ as follows:

$$P' = P + \frac{1}{n} e d^T. \tag{2}$$

According to the Ergodic Theorem for Markov chains [11], the chain defined by P' has a unique stationary probability distribution if P' is aperiodic and ireductible [28, 29].

Definition 5. Considering $g \in R^n$, a column vector that contains the centrality scores of all nodes in a $OpNet$, the PageRank mathematical problem [24] is defined as:

$$g^T \left(\alpha P' + \frac{1-\alpha}{n} e e^T \right) = g^T \tag{3}$$

where $0 \leq \alpha \leq 1$.

Definition 6. For an $OpNet$, we can change the uniform vector $\frac{1}{n}e^T$ with the more general vector v^T, called the *personalization vector*, where $v(i)$ represents the probability of node i to be considered as part of a specific path.

Definition 7. The connectivity matrix for an $OpNet$ is defined as $\overline{P} = \alpha P' + (1-\alpha) e v^T$, as an irreductible and aperiodic matrix: all nodes are connected.

As an extention of Definition 7, the additional normalization equation $g^T e = 1$ ensures that g^T is a probability vector, so it can be used to compute the centrality of nodes. The Definition 5 of PageRank problem can be reformulated now:

$$g^T \overline{P} = g^T. \tag{4}$$

The row vector g^T can be found by solving the eigenvector problem $g^T \overline{P} = g^T$ (i.e. $\overline{P}^T g = g$), or by solving the homogeneous linear system $g^T(\overline{P} - I_n) = 0^T$.

The Power Method used to solve this problem will use an initial approximation $g^{(0)}$ and will compute in step p: $x^{(p+1)} = \overline{P}^T x^{(p)}$. So, the multiplication operation is used for several times. Algorithm 1 presents the matrix-vector multiplication used in each step of the Power Method algorithm (introduced first by [16, 17]). A method for updating PageRank with iterative aggregation is also used to speed-up the PageRank convergence [18].

Definition 8. The upper bound for convergence, ϵ, is introduced with no dependency on nodes in a $OpNet$ as a convergence rule for PageRank Power Method algorithms (presented in Algorithm 2), for any node i:

$$\frac{||g^{(p+1)} - g^{(p)}||_1}{||g^{(p)}||_1} < \epsilon \tag{5}$$

DOI: 10.1002/cpe

Algorithm 1 Calculate $y = \overline{P}^T x$

Require: α introduced in Definition 5 ($0 \leq \alpha \leq 1$);
P matrix introduced by Definition 1 and the personalization vector v to compute $\overline{P}$
$\overline{P} = \alpha \left(P + \frac{1}{n} e d^T\right) + (1-\alpha) e v^T$;
$x \in R^n$, $x \neq 0$.
Ensure: $y = \overline{P}^T x$
$y = \alpha P^T x$.
$k = ||x||_1 - ||y||_1$.
$y = y + kv$.

The method for estimating centrality will be computed using PageRank Power Method, presented in Algorithm 2 using the convergence criteria introduced by 5.

Algorithm 2 PageRank Power Method

procedure PAGERANKPM($P, d, v, \alpha, g^{(0)}, \epsilon$)
  $\overline{P} = \alpha \left(P + \frac{1}{n} e d^T\right) + (1-\alpha) e v^T$ ▷ $0 \leq \alpha \leq 1$;
  repeat
    $g^{(p+1)} = \overline{P}^T g^{(p)}$. ▷ $g^{(0)} \in R^n$, $g^{(0)} \neq 0$ is an initial approximation;
    $\delta = \frac{||g^{(p+1)} - g^{(p)}||_1}{||g^{(p)}||_1}$.
  until $\delta < \epsilon$ ▷ $\epsilon > 0$ is the convergence rate.
end procedure

We are interested to estimate how sensitive is the centrality vector (the g - PageRank vector) to perturbations of P' matrix, to changes in damping factor α, and to changes in personalization vector v.

Let's consider that $g^T \overline{P} = g^T$ is the normal PageRank problem and $\tilde{g}^T \tilde{\overline{P}} = \tilde{g}^T$. We will compute the error $||\tilde{g} - g||_1$ and we have the following results:

- *Changes in the matrix P* [5]. The perturbed matrix will be: $\tilde{\overline{P}} = \alpha(P' + \delta P) + (1-\alpha) e v^T$ and the error is:
$$||\tilde{g} - g||_1 \leq \frac{\alpha}{1-\alpha} ||\delta P||_\infty \tag{6}$$
- *Changes in Damping Factor α* [19, 2]. The new matrix will be: $\tilde{\overline{P}} = (\alpha + \delta\alpha) P' + (1 - (\alpha + \delta\alpha)) e v^T$ and the error is:
$$||\tilde{g} - g||_1 \leq \frac{2}{1-\alpha} |\delta\alpha| \tag{7}$$
- *Changes in vector v* [5]. The perturbed matrix will be: $\tilde{\overline{P}} = \alpha P' + (1-\alpha) e (v + \delta v)^T$ and the error is:
$$||\tilde{g} - g||_1 \leq ||\delta v||_1 \tag{8}$$

Based on these results we can conclude that the PageRank problem is insensitive to perturbations for changes in vector v (the factor error estimation is 1). For the other two perturbations approaches, the sensitivity depends on dumping factor α. Table I presents results regarding sensitivity to perturbation for different values of the dumping factor. Here, the small values of the dumping factor are not convenient because the only contribution in the problem is given by the personalization vector. The ideal value for the dumping factor can be considered such that to keep the error in centrality vector under the errors presented in inputs (in all cases). So, the error factor must keep the error range (thus, we can consider ≤ 9.99), and we have the following results: $\frac{\alpha}{1-\alpha} \leq 9,99$ means that $\alpha \leq 0,91$ and $\frac{2}{1-\alpha} \leq 9,99$ means that $\alpha \leq 0,79$. So, we will keep the value $\alpha = 0,79$. Note: we will consider in all numerical computations the precision of the first two digit after decimal point.

DOI: 10.1002/cpe

α	*Changes in the matrix* P	*Changes in Damping Factor* α	*Changes in vector* v
	$\alpha/(1-\alpha)$	$2/(1-\alpha)$	1
0.10	0,11	2,22	1
0.50	1,00	4,00	1
0.79	3,76	9,53	1
0.85	5,67	13,33	1
0.99	99.00	200,00	1

Table I. Values of error factors for perturbed PageRank problem.

For the web value, $\alpha = 0.85$ the error factor is higher than 10 for the perturbation in damping factor, so it can not be a good approach for our problem. For changes in vector v we have stability for any value of the damping factor.

We consider that in Algorithm 2, the personalization vector can suffer changes in time: $v \to v + \delta v$. The $OpNet$ is a dynamic network, so it is possible that one value in this vector changes according with node interactions and feedbacks. This means that the probability of a node to be considered for a specific path in $OpNet$ can increase or decrease in time. In conclusion, the centrality of all node will be affected by this changes. We study the numerical stability of PageRankPM presented in Algorithm 2 and we study how a small change in the input $(v + \delta v)$ affects the execution of the algorithm, as well as its output. We will consider that in a numerically stable algorithm, small errors in the input (lessen in significance) will have little effect on the final output. If the algorithm is numerically unstable, errors in the input cause a considerably larger error in the output.

We will consider the following result (adapted to personalization vectors):

Theorem 1. (Convex Combinations of Personalization Vectors): For any convex combination of q personalization vectors, $v^T = \sum_{i=1}^{q} a_i v_i^T$, with $a_i \geq 0$, we have the convex combination: $g_v^T = \sum_{i=1}^{q} a_i g_{v_i}^T$.

Proof
The proof is given in [27], Theorem 2.2. □

Remark 1. We can consider that a small change in the input $v + \delta v$ is a convex combination of two personalization vectors. The second vector δv encodes the probability change for one single node for a specific route in an $OpNet$.

Remark 2. For a small change in the input, $v + \delta v$, we will consider a $g + \delta g$ change of the output. We will study the stability of a new algorithm used for centrality update: the computation of δg and the absolute error. It is important to mention that δg is not a centrality vector!

Remark 3. For a centrality vector g we have $||g||_1 = 1$. The new centrality vector $g + \delta g$ will respect the same condition, so $||g + \delta g||_1 = 1$. Considering these, we have $g^T e = 1$ and $(g + \delta g)^T e = 1$, so $\delta g^T e = 0$. This means that an increase in centrality value for one node will affect, by decreasing, the centrality for other nodes.

We will proof the following theorem:

Theorem 2. Let's consider a small changes of input for a personalization vector: $v + \delta v$. The computation method of output changes $(g + \delta g)$ is numerical stable and convergent.

Proof
The numerical stability was proofed by results (8): $||\delta g||_1 \leq ||\delta v||_1$. Now, the malformed PageRank method can be written as:

$$\left[\alpha P' + (1-\alpha)e(v + \delta v)^T\right]^T (g + \delta g) = g + \delta g,$$

$$\left[\alpha P' + (1-\alpha)ev^T + (1-\alpha)e\delta v^T\right]^T (g + \delta g) = g + \delta g,$$

DOI: 10.1002/cpe

$$\left[\overline{P} + (1-\alpha)e\delta v^T\right]^T (g + \delta g) = g + \delta g.$$

We will consider the initial PageRank problem $\overline{P}^T g = g$ and we have:

$$\overline{P}^T \delta g + (1-\alpha)\delta v e^T (g + \delta g) = \delta g.$$

Now, considering the remark 3, $(g+\delta g)^T e = e^T (g + \delta g) = 1$, we have:

$$\overline{P}^T \delta g + (1-\alpha)\delta v = \delta g,$$

$$\left(I_n - \overline{P}^T\right) \delta g = (1-\alpha)\delta v,$$

$$\delta g = \overline{P}^T \delta g + (1-\alpha)\delta v,$$

which is similar to a classical PageRank problem. The PageRank problem is convergent [12]. So, we can define an iterative method to compute δg as:

$$\delta g^{(k+1)} = \overline{P}^T \delta g^{(k)} + (1-\alpha)\delta v.$$

starting with any initial approximation $\delta g^{(0)}$.

□

Based on this result we will describe the steps for a centrality update algorithm (see Algorithm 3).

Algorithm 3 Centrality Re-Computation Algorithm

1: **procedure** CentralityUpdate($P, d, v, \alpha, g^{(0)}, \delta v, \epsilon$)
2: $\quad g$ = PageRankPM($P, d, v, \alpha, g^{(0)}, \epsilon$).
3: $\quad$ ▷ Solve the system: $[\alpha P' + (1-\alpha)e(v+\delta v)^T]^T (g + \delta g) = (g + \delta g)$.
4: $\quad P' = \left(P + \frac{1}{n} e d^T\right)$.
5: $\quad \overline{P}^T = \alpha P' + (1-\alpha) e v^T$.
6: $\quad$ **repeat** ▷ Compute δg using method proposed in Theorem 2.
7: $\quad\quad \delta g^{(k+1)} = \overline{P}^T \delta g^{(k)} + (1-\alpha)\delta v$.
8: $\quad\quad \delta = \frac{||\delta g^{(k+1)} - \delta g^{(k)}||_1}{||\delta g^{(k)}||_1}$.
9: $\quad$ **until** $\delta < \epsilon$
10: **end procedure**

Remark 4. Based on Theorem 2, δg has no dependency with g. Is is computed only using δv and the $\overline{P}$ matrix. So, we can apply Algorithm 3 for any updates in the centrality vector, moving step 2 outside procedure $CentralityUpdate$.

Based on this result we will write a new procedure that will apply multiple changes to centrality vector. The algorithm considers a set of changes in the probability for a node to be part on specific path in a $OpNet$ defined by $\{\delta v\}$.

Remark 5. We can start the $AdaptiveCentralityUpdate$ procedure with any initial approximation $g^{(0)}$ and keeping the same convergence factor (ϵ) for any updates.

Remark 6. The name $AdaptiveCentralityUpdate$ suggests that we can add the current time moment, T_i, as a last parameter for this procedure and we can replace step 2 with $g(T_i) = AdaptiveCentralityUpdate(P, d, v, \alpha, g^{(0)}, \{\delta v\}(T_i), \epsilon, T_{i-1})$, where all changes denoted by $\{\delta v\}$ set are produce at the current time T_i.

Algorithm 4 Adaptive Centrality Re-Computation Algorithm

procedure ADAPTIVECENTRALITYUPDATE($P, d, v, \alpha, g^{(0)}, \{\delta v\}, \epsilon$)
    g = PageRankPM($P, d, v, \alpha, g^{(0)}, \epsilon$).
    $\delta g = 0$;
    for any update $\delta v_i \in \{\delta v\}$ **do**
        $\delta g = \delta g$ + CentralityUpdate($P, d, v, \alpha, g^{(0)}, \delta v_i, \epsilon$).
    end for
    $g = g + \delta g$.
end procedure

4. EXPERIMENTAL RESULTS

We performed our tests using a mobility trace available on CRAWDAD[†] created at the National University of Singapore [30]. The trace contains information regarding contacts between students at the university collected based on information on class schedules and rosters in the spring semester of 2006. The contacts are created synthetically, based on the assumption that students attending the same class at the same time are in Bluetooth range of each other, thus having a contact for the duration of the class. The authors validated this assumption inside the large classrooms of the campus. Moreover, two students that attend courses in classes that are next to each other are not considered to be in contact, and the encounters are only registered during study hours. The size of the campus is of 146 ha, and the total number of students recorded in the trace is 22341. The trace recorded 4885 sessions (courses that students were attending), with an average number of 40.25 students per session, each student attending an average of 8.8 sessions a week. The diameter of the student interaction graph is 8, which means that two students are connected by at most 8 hops, with an average hop distance of 2.45. The trace files contain entries for each contact, which are composed of the IDs of the encountering nodes (from 0 to 22340), as well as the moment when the contact occurs and its duration.

In order to run the trace and analyze the contacts between nodes for our purpose, we used the MobEmu emulator [8]. This is a Java program that allows us to parse a mobility trace and replay it step by step, while being able to apply a given algorithm at any point during the trace (in our case, at every contact between two nodes).

Example 1. For our experiments, we initially used the first **500** participants in the trace (based on the order of their IDs). We ran the k-CLIQUE algorithm [14] for detecting the social communities formed between the nodes in the opportunistic network represented by the trace. The social community of a node is represented as a list of other nodes that it has social relationships with, and it can be seen depicted on the left side of Figure 1, where a star represents a k-CLIQUE connection between the node on OX and the node on OY. k-CLIQUE computes the communities by assuming that socially-connected nodes spend more time in contact with each other that with regular nodes. Therefore, two nodes belong to the same community if they have spent a certain time in contact (given by the contact threshold) and if they have a certain number of common "friends" (the community threshold). The values we chose for the contact and community thresholds were **6** hours and **8** respectively, based on obtaining suitable community sizes for all nodes in regard to the size of the opportunistic network (see the label `kclique500-6-8` in Figure 1). We also ran the emulator with different parameters, namely a community threshold of **10** hours and a contact threshold of **15** (label `kclique500-10-15` in Figure 1), in order to highlight that, when we increase the two k-CLIQUE parameters, the social communities become sparser. This may lead to less congestion when using a routing algorithm that takes advantage of nodes that are popular in several communities, but also to lower hit rates, since nodes are less likely to have large centralities in multiple communities. The right side of Figure 1 shows the values obtained after computing

[†] http://crawdad.org/

DOI: 10.1002/cpe

each node's centrality using the PageRank approach presented in Section 3. It can be seen that the nodes with the highest centrality values are also the nodes that have larger social communities, since they connect multiple communities with each other. While running k-CLIQUE on the mobility trace, we also computed a **probabilistic cost** for the link between two community nodes. This value is calculated as the total duration that the two nodes have been in contact, normalized with the maximum contact duration of the trace. This cost acts as an importance (or strength) of the social connection between two trace participants.

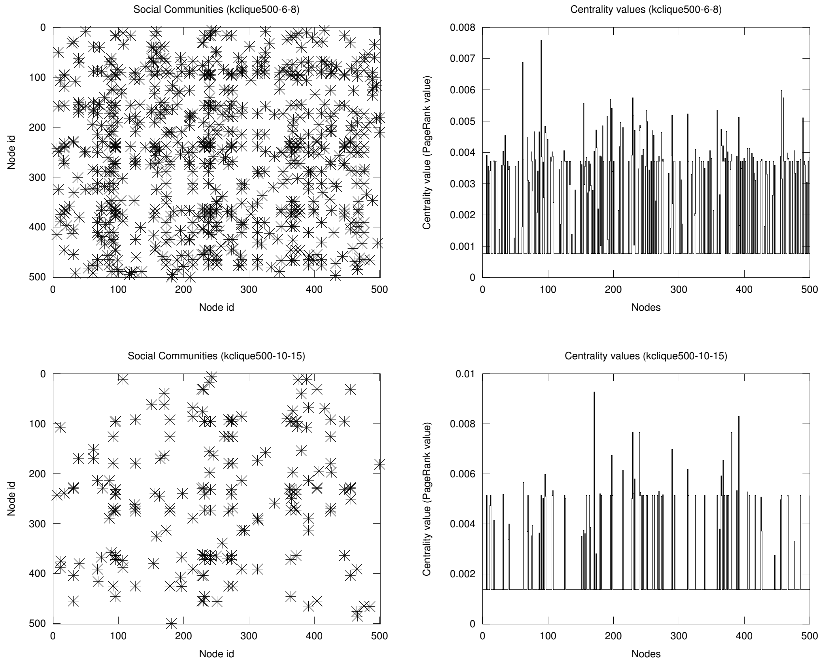

Figure 1. Experiments for 500 participants. The contact is 6 hours and community threshold is 8.

Example 2. Secondly, we also took the first **1000** participants in the trace using the same values for the contact and community thresholds: **6** hours and **8** respectively (see the label `kclique1000-6-8` in Figure 2). Similarly, we also tested with parameter values of **10** hours and **15**, as shown in the `kclique1000-10-15` charts. The conclusions are similar to the ones from the first example: larger k-CLIQUE parameters lead to sparser social community matrices, while nodes with more social connections tend to have higher PageRank centralities.

In both examples we have several nodes that are not connected in the network. The ranks of these nodes are equal and remain constant over time and are insensitive to any update. Also, Figure 1 and Figure 2 show that there are several nodes with higher ranks, which means that the popularity of these nodes are high. Based on these experimental results, it is possible to use the nodes' centrality for clustering procedure as an input in any routing protocol.

The numerical stability was proofed by results (8): $||\delta g||_1 \leq ||\delta v||_1$. We chose $\epsilon = 10^{-3}$ for all experiments and we obtained $||\delta v||_1 \leq \epsilon$. The is a theoretical results, so all the experiments obey these results. The convergence of our method was also proved theoretically in Theorem 2. The obtained results demonstrate these properties. The reader is also referred [7] to for a more extended analysis.

DOI: 10.1002/cpe

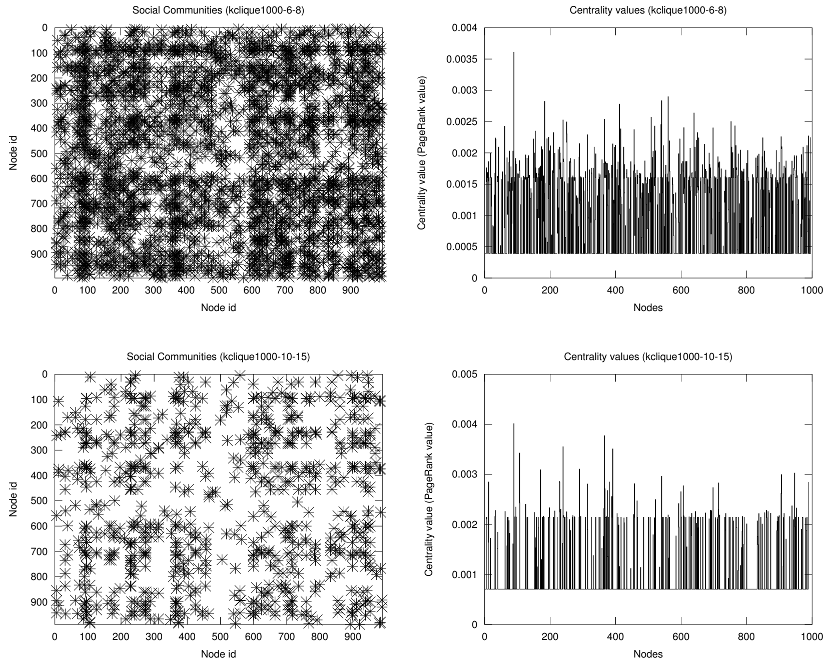

Figure 2. Experiments for 1000 participants. The contact is 6 hours and community threshold is 8.

5. CONCLUSIONS

We proposed in this paper a new model to estimate centrality of a node in a social-based Mobile Networks by adopting the *popularity of nodes* introduced as the chance to participate in a path formation (e.g., node to lead to the next popular path). We presented our solution that considers the PageRank approach used by many popular search engines to compute the popularity of Web pages. We used the PageRank algorithm because it offers numerical stability and fast convergence in implementation. Also, the model for centrality computation in opportunistic networks is similar to the Web Rank model proposed by Larry Page. The significance with regards to the current approaches is represented by the adaptive computation of node centrality, the proposed method being sensitive to changes in the network. According with this, we proposed an algorithm to update the centrality of nodes using the probability of each node to be traversed by messages circulating within the network. We proved that our algorithm converges very well, similar to the classical PageRank problem. We also showed that our approach is highly robust, numerical insensitive to errors and converges fast, which means that is can be easily adopted in resource-constraint environments formed between highly mobile wireless devices.

We performed our tests using a mobility trace available on CRAWDAD created at the National University of Singapore. We conclude that it is possible to consider the rank value as centrality of a node in order to decide the selection of it in a content distribution path.

As future work we will consider a change in focus from *popularity of nodes* to *popularity of connections* established between nodes. We will reconsider our proposed algorithm to update the popularity of links between nodes using the probability of each link, in this case, to be traversed by messages circulating within the network. As future study, also some the decomposition approach of this method will be studied for large networks. We will study the algorithm in relation with many similar ON routing algorithms, using various case studies.

DOI: 10.1002/cpe

6. ACKNOWLEDGEMENTS

This work was supported by project "*ERRIC - Empowering Romanian Research on Intelligent Information Technologies*", FP7-REGPOT-2010-1, ID: 264207. The research presented in this paper is supported by Romanian project: "*SideSTEP - Scheduling Methods for Dynamic Distributed Systems: a self-* approach*", ID: PN-II-CT-RO-FR-2012-1-0084. The work has been co-funded by the Sectorial Operational Program Human Resources Development 2007-2013 of the Romanian Ministry of Labor, Family and Social Protection through the Financial Agreement "*POSDRU/89/1.5/S/62557*".

REFERENCES

1. Monica Bianchini, Marco Gori, and Franco Scarselli. Inside pagerank. *ACM Trans. Internet Technol.*, 5(1):92–128, February 2005.
2. Paolo Boldi, Massimo Santini, and Sebastiano Vigna. Pagerank: Functional dependencies. *ACM Trans. Inf. Syst.*, 27(4):19:1–19:23, November 2009.
3. Chiara Boldrini, Marco Conti, Jacopo Jacopini, and Andrea Passarella. HiBOp: a History Based Routing Protocol for Opportunistic Networks. In *World of Wireless, Mobile and Multimedia Networks, 2007. WoWMoM 2007. IEEE International Symposium on a*, pages 1–12, June 2007.
4. Chiara Boldrini, Marco Conti, and Andrea Passarella. ContentPlace: social-aware data dissemination in opportunistic networks. In *Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems*, MSWiM '08, pages 203–210, New York, NY, USA, 2008. ACM.
5. Steve Chien, Cynthia Dwork, Ravi Kumar, and D Sivakumar. Towards exploiting link evolution. *Unpublished manuscript*, 2001.
6. Radu Ciobanu and Ciprian Dobre. Data dissemination in opportunistic networks. In *18th International Conference on Control Systems and Computer Science*, CSCS-18, pages 529–536, 2011.
7. Radu-Ioan Ciobanu, Ciprian Dobre, and Valentin Cristea. Sprint: Social prediction-based opportunistic routing. In *Proceedings of the 7th IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications (IEEE AOC 2013), IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2013)*, 2013.
8. Radu Ioan Ciobanu, Ciprian Dobre, Valentin Cristea, and Dhiya Al-Jumeily. Social Aspects for Opportunistic Communication. In *Parallel and Distributed Computing (ISPDC), 2012 11th International Symposium on*, pages 251–258. IEEE, June 2012.
9. Marco Conti, Silvia Giordano, Martin May, and Andrea Passarella. From opportunistic networks to opportunistic computing. *Comm. Mag.*, 48:126–139, September 2010.
10. Linton C Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1979.
11. Geoffrey Grimmett and David Stirzaker. *Probability and random processes*, volume 2. Clarendon press Oxford, 1992.
12. Taher Haveliwala and Sepandar Kamvar. The second eigenvalue of the google matrix. *Stanford University Technical Report*, 2003.
13. Pan Hui, Jon Crowcroft, and Eiko Yoneki. Bubble rap: social-based forwarding in delay tolerant networks. In *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, MobiHoc '08, pages 241–250, New York, NY, USA, 2008. ACM.
14. Pan Hui, Eiko Yoneki, Shu Y. Chan, and Jon Crowcroft. Distributed community detection in delay tolerant networks. In *Proceedings of 2nd ACM/IEEE international workshop on Mobility in the evolving internet architecture*, MobiArch '07, pages 1–8, New York, NY, USA, 2007. ACM.
15. Ilse C. F. Ipsen and Steve Kirkland. Convergence analysis of a pagerank updating algorithm by langville and meyer. *SIAM J. Matrix Anal. Appl.*, 27(4):952–967, December 2005.
16. Sepandar Kamvar, Taher Haveliwala, and Gene Golub. Adaptive methods for the computation of pagerank. *Linear Algebra and its Applications*, 386:51–65, 2004.
17. Sepandar D. Kamvar, Taher H. Haveliwala, Christopher D. Manning, and Gene H. Golub. Extrapolation methods for accelerating pagerank computations. In *Proceedings of the 12th international conference on World Wide Web*, WWW '03, pages 261–270, New York, NY, USA, 2003. ACM.
18. Amy Langville and Carl Meyer. Updating pagerank with iterative aggregation. In *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, WWW Alt. '04, pages 392–393, New York, NY, USA, 2004. ACM.
19. Amy N. Langville and Carl D. Meyer. Deeper inside pagerank. *Internet Mathematics*, 1(3):335–380, 2004.
20. Vincent Lenders, Martin May, Gunnar Karlsson, and Clemens Wacha. Wireless ad hoc podcasting. *SIGMOBILE Mob. Comput. Commun. Rev.*, 12:65–67, January 2008.
21. Mark E. J. Newman. The mathematics of networks. *The New Palgrave Encyclopedia of Economics*, 2(June 1996):1–12, 2007.
22. Hoang Anh Nguyen, Silvia Giordano, and Alessandro Puiatti. Probabilistic routing protocol for intermittently connected mobile ad hoc network (propicman). *2007 IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks*, pages 1–6, 2007.
23. Pavlos Nikolopoulos, Therapon Papadimitriou, Panagiotis Pantazopoulos, Merkourios Karaliopoulos, and Ioannis Stavrakakis. How much off-center are centrality metrics for routing in opportunistic networks. In *Proceedings of the 6th ACM workshop on Challenged networks*, CHANTS '11, pages 9–14, New York, NY, USA, 2011. ACM.

24. Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.
25. Luciana Pelusi, Andrea Passarella, and Marco Conti. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. *Communications Magazine, IEEE*, 44(11):134–141, November.
26. Joseph J Pfeiffer III and Jennifer Neville. Methods to determine node centrality and clustering in graphs with uncertain structure. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. http://www. cs. purdue. edu/homes/neville/papers/pfeiffer-icwsm2011. pdf*, 2011.
27. Florin Pop and Ciprian Dobre. An efficient pagerank approach for urban traffic optimization. *Mathematical Problems in Engineering*, 2012, 2012.
28. Cristina Serbanescu. Noncommutative markov processes as stochastic equations solutions. *Bull. Math. Soc. Sc. Math. Roumanie Tome 41*, 89(3):219–228, 1998.
29. Cristina Serbanescu. Stochastic differential equations and unitary processes. *Bull. Math. Soc. Sc. Math. Roumanie Tome 41*, 89(3):311–322, 1998.
30. Vikram Srinivasan, Mehul Motani, and Wei T. Ooi. Analysis and implications of student contact patterns derived from campus schedules. In *Proceedings of the 12th annual international conference on Mobile computing and networking*, MobiCom '06, pages 86–97, New York, NY, USA, 2006. ACM.
31. Kuang Xu, V.O.K. Li, and Jaewoo Chung. Exploring centrality for message forwarding in opportunistic networks. In *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*, pages 1–6, April.
32. Eiko Yoneki, Pan Hui, ShuYan Chan, and Jon Crowcroft. A socio-aware overlay for publish/subscribe communication in delay tolerant networks. In *Proceedings of the 10th ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems*, MSWiM '07, pages 225–234, New York, NY, USA, 2007. ACM.
33. Peiyan Yuan, Huadong Ma, Xiang-Yang Li, ShaoJie Tang, and XuFei Mao. Opportunistic forwarding with partial centrality. *CoRR*, abs/1208.0186, 2012.

DOI: 10.1002/cpe