Eigenvector Centrality in Highly Partitioned Mobile Networks: Principles and Applications

Iacopo Carreras[1], Daniele Miorandi[1], Geoffrey S. Canright[2] and Kenth Engø-Monsen[2]

[1] CREATE-NET
Via Solteri 38
38100 – Trento (Italy)
name.surname@create-net.org

[2] Telenor R&I
Snaroyveien 30
N-1331 Fornebu (Norway)
name.surname@telenor.com

Summary. In this chapter we introduce a model for analyzing the spread of epidemics in a disconnected mobile network. The work is based on an extension, to a dynamic setting, of the eigenvector centrality principle introduced by two of the authors for the case of static networks. The extension builds on a new definition of *connectivity matrix* for a highly partitioned mobile system, where the connectivity between a pair of nodes is defined as the number of *contacts* taking place over a finite time window. The connectivity matrix is then used to evaluate the eigenvector centrality of the various nodes. Numerical results from real-world traces are presented and discussed. The applicability of the proposed approach to select on-line message forwarders is also addressed.

1 Introduction

In this chapter, we aim at characterizing the spreading power of nodes in a mobile disconnected network. In such a scenario, nodes exploit opportunistic contacts among themselves to diffuse information, mainly in the form of messages (i.e., chunks of packets). The problem is of interest for a number of reasons, the prominent one being the research efforts towards the definition of an architecture for Delay-Tolerant Networks (DTNs) [4, 5]. DTNs are networks in which the existence of a path between any pair of nodes is not taken for granted. Differently from conventional IP networking paradigm, DTNs are able to operate in the presence of frequent network partitions. In DTNs the conventional notion of "network" itself needs to be re-thought anew: information can diffuse in the system by means of (i) node mobility: a device conveys information while moving around (ii) opportunistic forwarding: messages are passed from a node to another one when they get in *contact*. Due to the high dynamism of such scenarios, some epidemic mechanism has to be used to spread

I. Carreras et al.: *Eigenvector Centrality in Highly Partitioned Mobile Networks: Principles and Applications*, Studies in Computational Intelligence (SCI) **69**, 123–145 (2007)
www.springerlink.com

information until it reaches the intended destination [6, 7, 8]. The problem we aim at tackling is to devise distributed mechanisms able to decide whether, upon meeting a given node, it should be used as a forwarder.

The starting point for our study is the work carried out by two of the authors on the spreading power of nodes in a static setting [9, 10]. The work therein focuses on the notion of *eigenvector centrality* (EVC), shown to be a meaningful measure of the ability of the nodes to spread an epidemic in the network. The EVC is computed as the eigenvector relative to the spectral radius (i.e., the largest eigenvalue) of the adjacency matrix of the network. Such a procedure is shown to produce a smooth measure, which can be used for studying the spread of epidemics [10]. Further, it implies a natural way of defining clusters in the network. The resulting network topography can be used to define regions, each region being characterized by the fact that epidemics spread extremely fast therein.

We would like therefore to extending the EVC concept to the case of highly partitioned mobile networks. The underlying idea is that EVC can be used as a metric for deciding whether a node encountered on the way should be used as a forwarder or not. This requires two main steps: (i) extending the definition of EVC to mobile disconnected scenarios, which requires to introduce an appropriate matrix able to describe well the "interaction pattern" among nodes (ii) introducing mechanisms and techniques for allowing a distributed on-line estimation of the EVC. In the chapter, we will presents techniques and results for issue (i), while some preliminary considerations will be presented for point (ii).

The main contribution of this chapter is the extension of the eigenvector centrality principle to more dynamic and disconnected scenarios. We will indeed focus on how to extend the topographic picture, with its obvious advantages for studying epidemic spreading, to a dynamic network — one in which the links are time-dependent. Our aim is to take a mobility pattern as input, to determine the time-dependent links, and finally to produce a topographic analysis analogous to the static case — with a smooth centrality function over the nodes, regions defined so as to correspond to well connected subgraphs, and a meaningful connection to the nodes' roles in spreading. Thus we want to define some kind of time-averaged or time-integrated *connectivity matrix*, with non-negative link weights. Given a suitable definition of connectivity matrix,[3] we can apply the EVC analysis to study its topographic properties. We introduce the notion of T-tolerant connectivity matrix, whose (i, j)-th entry equals the number of contacts between nodes i and j over a time window of length T.

The remainder of this chapter is organized as follows. In Sec. 2 we review the basic EVC concept and its application to static settings. In Sec. 3 we introduce the notion of T-tolerant connectivity matrix and show how to build it. In Sec. 4 we present some numerical results, obtained from real-world mobility traces, and discuss the properties of the correspondent systems. Sec. 5 concludes the chapter pointing out some directions for future work.

[3] The term "connectivity matrix" can be somehow misleading, in that we are mostly interested in disconnected scenarios, where a randomly taken snapshot of the network returns a disconnected graph.

2 Eigenvector Centrality, Topography, and Spreading

In this section we present a brief review of earlier work involving a "topographic" view of the structure of networks with undirected links. This work is presented in detail in [9] — which presents the basic structural analysis, based on eigenvector centrality or EVC [11] — and in [10] — which shows the utility of this analysis for understanding and predicting the progress of epidemic spreading over the network.

2.1 Previous Work

A fundamental problem in network analysis is the problem of *clustering*. That is: given a network topology (possibly with weights on the links), and assuming that the network is symmetric and connected, how do we then define and identify subgraphs of the whole network which, in some well defined sense, consist of nodes which "belong together"?

One meaningful and useful notion of "belonging together" is that of "being well connected". This point of view states that the clusters (subgraphs) of a network are better connected internally than they are to other clusters. Two of us [9] have constructed a precise version of this kind of clustering criterion. That is, one can define "well-connectedness" in a number of ways; but one definition — valid for a single node — is that the node be well connected to nodes that are well connected. This is of course a circular definition, which however is readily formulated mathematically [9, 11]. The resulting single-node measure of well-connectedness is termed "eigenvector centrality" or EVC [11]. Let us consider a graph model of the network topology and denote by A the corresponding adjacency matrix.

The EVC of a node i is defined being proportional to the sum of the the EVC of i's neighbors:

$$e_i = \frac{\sum_{j=nn(i)} e_j}{\lambda}, \tag{1}$$

where $nn(i)$ denotes the set of neighbors of node i. We can rewrite (1) in a compact matrix form:

$$A \cdot e = \lambda e, \tag{2}$$

where e represents the vector of nodes' centrality scores. Otherwise stated, e is the eigenvector of A relative to the eigenvalue λ. A number of reasons [9, 10] lead to the choice of the maximal eigenvalue λ_{max}. Due to the fact that A is nonnegative, $\lambda_{max} \geq 0$ and all the components of e are nonnegative [12].

EVC can be exploited to define clusters as follows. We note that, because of the dependence of a node's EVC on that of its neighbors, the EVC may be regarded as a "smooth" height function over the network. This smoothness motivates the appeal to topographic notions. That is: local maxima of the EVC are regarded as being the most well-connected node in a region of good connectivity. In topographic terms, these local maxima are mountain peaks. Each "mountain" (region of good connectivity) is then defined by its peak, plus a rule for membership of other (non-peak) nodes [9] in the well-connected cluster. A simple membership rule is that each node belongs to the same mountain as does its highest (EVC) neighbor. In other words: a node is connected to the cluster of its best connected neighbor. With this rule, essentially all nodes are assigned uniquely to one mountain; the mountains are then the well connected clusters of the graph. These are termed *regions* in the earlier work, and we will adhere to that usage here.

In more recent work [10] it has been shown that this smooth definition of well connectedness, and the resulting notion of regions, are very useful for describing and understanding the time and space (network) progression of an epidemic on an undirected network. The basic idea is that an infection front tends to move towards neighborhoods of high connectivity (EVC), because the spreading is fastest in such neighborhoods. This says in other words that the front naturally moves "uphill" over EVC contours—which in turn implies that, in general, spreading will be faster within regions (mountains) than between them. These ideas, and related ones, have been described in detail, and confirmed in simulations, in [10].

Some important points from this work which are relevant for the present chapter are:

- The importance of a node to the process of epidemic spreading may be roughly measured by the node's eigenvector centrality.
- *Regions*, as defined by the steepest-ascent rule, are clusters of the network in which spreading is expected to be relatively rapid and predictable. (Here "relatively" simply means compared to inter-region spreading.)
- Nodes whose links ("bridging links") connect distinct regions play an important role in the (less rapid, and less predictable) spreading from one region to another.

One can readily identify two important directions for extending such work, building on the work described in the previous subsection. First, the described topographic approach is, in its present form, only applicable to a *static* network. Secondly, the analysis is thus far only applicable when some entity (researcher, manager, engineer, machine) has access to knowledge of the full topology of the network. The latter constraint stems from the fact that the EVC of any node is in fact dependent on the entire topology of the network—because it is taken from the dominant eigenvector of the network's adjacency matrix, and because the network is (assumed) connected. Thus, an interesting question is how one can define local, distributed methods for finding the EVC distribution and the region structure of a given network. This question will not however be pursued in this chapter, whereas we will rather concentrate on the first issue, i.e., how to extend the EVC analysis to a mobile disconnected system.

3 Connecting Disconnected Networks

The results presented in the previous section justify the use of the EVC principle as a suitable metric for identifying the spreading power of nodes. The results therein clearly apply to a static network topology, where it is straightforward to define the term "connectivity matrix". Things change drastically when we look at mobile networks, where links come and go. One could consider taking a snapshot of the network at a randomly chosen time instant, but this clearly does not make sense for highly partitioned mobile networks as the ones subject of our study. It is therefore imperative, in order to extend the use of EVC to such scenario, to define a suitable matrix, whose entries reflect the actual level of "interaction" among nodes in the system. In this section, we will present a simple method for building such a matrix, that we call the T-tolerant connectivity matrix of the system.

3.1 Epidemic Spreading in Highly Partitioned Mobile Networks

Let us consider first the case of a wireless network in which all nodes are static. In this case, we can build a matrix A, that we call the *connectivity matrix*, by looking at the fact that two nodes are within mutual communication distance. In other words, the (i, j)-th entry is positive, $A(i, j) > 0$ if and only if nodes i and j are within mutual communication range. For example, $A(i, j)$ could be the rate at which two nodes are able to transmit data to each other. In this way we could account for SNR variations while still being able to define a matrix A representing the "topology" of the network. Things drastically change when we consider a mobile network. In this case, indeed, there is no standard notion of network topology we can rely on for building the connectivity matrix. This task gets even more challenging if we focus on networks operated in the *subconnectivity regime* [13], in which no giant component exists.

We are actually interested in something stronger, i.e., looking at networks in which, at any given time instant, each node is isolated with high probability. We have two reasons for looking at such scenarios. The first is technology-driven, in that there is a large number of application scenarios in which such assumptions hold. Most of them fall within the category of Delay Tolerant Networks (DTNs) [5]. DTNs are networks that are able to work in the absence of continuous connectivity. The connectivity in DTNs follows a random pattern, and is not taken for granted as in standard IP-based networks. DTNs include, e.g., Vehicular Ad hoc NETworks (VANETs), where the nodes are represented by cars, which may exchange data to run distributed services, including traffic monitoring, alert messages, personalized advertising etc.[14]. On the other hand, there is also a deep scientific reason for looking with interest at networks operated in this regime. Indeed, deep results in the field of network information theory state that connected wireless networks present poor scalability properties [15]. This is mainly rooted in the fact that network operations, in the wireless domain, are interference-limited. And the need for maintaining a connected network leads to the necessity of using a quite large communication range, at the expense of network capacity. On the other hand, nodes mobility can be exploited, for a network operating in the subconnectivity regime, to achieve a scalable network model [16].[4]

[4] In particular, for a connected network of n nodes, half of which act as sources and the other half as receivers, the per-connection throughput scales as $\Theta\left(\frac{1}{\sqrt{n}}\right)$, whereas in a mobile network with the two-hop relaying scheme in [16], the same quantity scales as $\Theta(1)$.

Given the aforementioned scenario, it remains to discuss how information can be successfully delivered in such systems. Standard routing algorithms were indeed built for wired networks, in which links are static. It is therefore sufficient to discover the route to the intended destination and use it for transmitting all data packets. In a highly mobile system, on the other hand, the notion of route itself looses its significant, since the destination becomes a "mobile target" and does not have a destination address any longer. Keeping track of the movements of all nodes is clearly not feasible, due to the obvious scalability problems related to the amount of control traffic. It is therefore imperative, for such scenarios, to resort to some form of epidemic forwarding [6]. In such protocols, data packets spread in the network like a virus or some other forms of epidemics. Nodes which carry the packets may infect nodes who had not received it yet (also called "susceptible" in standard epidemiology lexicon). Also in this case, in principle, scalability problems could arise, due to the fact that the number of copies of a message in the network grows exponentially over time. Mechanisms are needed to limit the spreading of epidemic data, in the form, e.g., of limiting the lifetime of a packet [1], the number of nodes a single user can infect (also called K-copy relaying protocols [20]) or by limiting the number of hops a packet can traverse. Also mechanisms based on the use of a "vaccine" or "immunization" in the form of so-called anti-packets were devised [2]. On the other hand, we envision to enhance these schemes by assigning to each node a value (a sort of fitness level) which describes their ability to spread an epidemic on the network. A node with a high fitness level has therefore a high level of interactions with many other nodes in the network, and configures itself as a good carrier of messages to the intended destination. Forwarding decisions (i.e., decisions whether to forward a message to an encounter or not) are taken on the basis of such fitness value. In this paper, we do not aim at specifying the exact mechanism based on which such system should work. Our target is to define a suitable method for *computing* such a fitness value.

In a disconnected scenario, where nodes are isolated most of the time, the only mean of transmitting information is given by *meetings* among nodes. Nodes i and j are said to meet at time t if, given an infinitesimal amount of time δt, at time $t - \delta t$ they were not able to transmit messages to each other, while they could do so at time $t + \delta t$. Meetings form a random pattern, which depends on (i) the mobility of nodes (ii) the random channel fluctuations. Meetings are characterized by (i) the IDs of nodes which get within mutual communication range (ii) the time at which the meeting takes place (iii) the duration of the meeting. The last parameter reflects the length of the time interval during which nodes are able to communicate. We actually overlook issue (iii), and consider that the duration of a meeting is sufficient for delivering all the data which needs to be transferred. This is reasonable since, with current wireless technologies (e.g., IEEE 802.11g), nodes moving even at vehicular speeds are able, in most situations, to transfer some MBytes of data [17].

In order for the EVC analysis to produce meaningful results, the matrix A should relate to a graph which is connected with high probability. On the other hand, given our assumptions, a randomly taken snapshot of the network status would report a graph which is almost surely disconnected. We need therefore to devise a mechanism able to enable us to pass from the *subcritical* to the *supercritical* regime.

3.2 The T-Tolerant Connectivity Matrix

The basic idea which we propose for building the matrix A is to consider an integrated version of the instantaneous network connectivity. Given a constant T, we construct the T-tolerant connectivity matrix A_T by considering all the meetings taking place over a time window of length T. The (i, j)-th entry $A_T(i, j)$ equals the number of meetings taking place in the time interval $[t_0, t_0 + T)$, where t_0 is a given initial time instant.

The underlying idea is the following. If the weight of a link increases with the strength or frequency of the connection — or, in other words, with the probability of spreading an "infection" over the link — then the analysis of this connectivity matrix may be expected to give useful information about how an infection (or message) is likely to be spread in a network of mobile nodes. Subsequent analysis is then the same as done in [9] and [10]; and because of this assumption about the link weights, the results should be useful for describing the process of diffusion of messages in the network, and for getting insight into the role played by the various nodes in the spreading process.

The definition of the T-tolerant connectivity matrix is worth some comments. In particular, the choice of an adequate time window T appears non-trivial. The value of T shall be consistent with the nodes' dynamics, so that the matrix A_T gives rise to non-trivial results. If T is too small, most entries of A_T would turn out to be zero, the network would show low connectivity and, hence the EVC analysis would not carry any significant information. The parameter T shall be understood as a kind of time constant of the system. In particular, we are interested in working around the value of T for which the graph associated to the matrix A_T gets connected. Such value depends on the mobility and is therefore out of the control of system designers. On the other hand, such value determines the time horizon over which the system is able to deliver messages, placing therefore a constraint on the class of applications which can be supported by the system.

Recent results by the first two authors [18] show that, given a system with n nodes and where meetings take place at intensity λ[5], it can be shown under some independence assumptions that a *phase transition* takes place at T given by:

$$T = \Theta\left(\frac{n \log n}{\lambda}\right). \tag{3}$$

At such value, the graph corresponding to the skeleton of the matrix A_T passes from the subcritical to the supercritical result.

This time-integrated version of the connectivity matrix presents, however, also a shortcoming, and cannot be used — in general — to define the ability of the system to deliver data. Consider the example reported in Fig. 1, where we considered the graph $\mathcal{G}_T$ (called the T-tolerant connectivity graph) which we can naturally associate with the matrix A_T. The graph $\mathcal{G}_T$ has vertex set corresponding to the nodes in

[5] In this case, λ has to be understood as the intensity of the point process that can be easily constructed to reflect the dynamics of meetings taking place in the system.

the network. There is a link between i and j if and only if there is a non-zero value in the corresponding entry of the matrix A_T. The link is labelled (or weighted) with the value of the corresponding matrix entry. In the situation considered in Fig. 1 there are three nodes, A, B and C. At time t_1, nodes A and B get within mutual communication range. Thus, we set the corresponding entry in the connectivity matrix to 1, and add an edge (A, B) of unitary weight in the graph $\mathcal{G}$. At time t_2, node A meets with node C. Also in this case we add an edge of unitary weight. At time t_3, node A gets again in contact with node C. The weight of the corresponding link is thus increased to 2. In the resulting graph, there is a path connecting node B to node C. However, there is a timing (or ordering) issue that is not accounted for in this construction. Indeed, node B can successfully transmit information to node C (by using A as relay), but not vice versa.

Noting this potential problem, we proceed next to experiments. We find from our experiments that the EVC analysis, when applied to such a matrix A_T, returns useful results for understanding the spread of epidemics in disconnected mobile networks. We will give some discussion of why this is so in the final section.

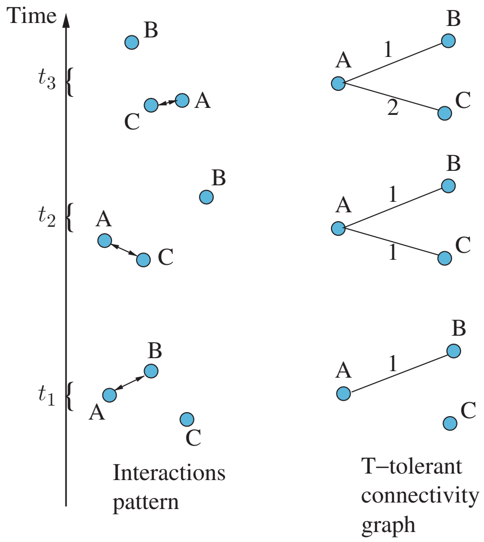

Fig. 1. Graphical representation of the timing issue in constructing the T-tolerant connectivity graph. In the resulting connectivity graph, there exists a path between B and C. Due to the dynamics, data originated from B can be delivered to C, but not vice versa

4 Experimental Results: Real-World Mobility Traces

In this section, we aim at verifying the properties, in terms of EVC distribution and spreading power of nodes, of various systems based on the deployment of disconnected mobile networks. While the analysis is carried out off-line, working on a log file describing all the contacts having taken place over the experimentation duration, we believe it to be a useful first step for understanding the ability of EVC, coupled with the T-tolerant connectivity matrix, to describe the spreading power of nodes in such framework. In particular, we considered four classes of traces, generated in different experiments and available on the CRAWDAD database at Dartmouth College [19].

The first set of experiments were performed by Intel Cambridge and reported in [20]. In this case, the devices were iMotes, equipped with a Bluetooth radio interface, and carried around by people in (i) a lab at Cambridge University (ii) people in the Intel lab at Cambridge (iii) attendees of IEEE INFOCOM05. Each iMote periodically scans the frequency range to check if other devices are present; each contact event is registered together with its time-stamp and duration.

The second set of experiments refers to the DieselNet project at University of Massachusetts at Amherst [21]. In this case, a number of buses has been equipped with IEEE 802.11b-compliant access points, used for bus-to-bus communications. Each device records the time and location of the contacts with other buses in the system. In our analysis, the trace 30122005 has been considered.

The third set of experiments comes from the Reality Mining project at MIT [22]. Data refers to a very long period (approximately one year), over which a set of students/researchers were tracked by means of mobile phones equipped with a Bluetooth interface. As for the Intel experiment, the Bluetooth device-discovery feature is exploited in order to detect the proximity of other Bluetooth-enabled devices. Each meeting is traced, and subsequently stored in a central repository for later processing.

The fourth set of experiments comes from the Student-Net project at University of Toronto [23]. In the experimentation, students were equipped with Bluetooth-enabled hand-held devices, capable of issuing inquiries and tracing any pairwise contact between users. The inquiry period was set so to preserve a 8–10 hours battery life-time. This resulted in a 16 s scan period. They performed two separate studies. The first one involved approximatively 20 graduate students only for a duration of two-and-a-half weeks, while the second one involved undergraduate students only, for a duration of 8 weeks.

All trace files have been preprocessed. We indeed observed in the traces that a single contact opportunity could lead (due to either SNR fluctuations or to the methods used to log contacts) to multiple entries in the trace file. This results in a series of "meetings" between a couple of nodes, with intermeeting times of the order of a few milliseconds, clearly not due to the devices mobility. Only the first entry in the trace file is kept for each meeting.

In Table 1, some details of the three experimentations are reported. It is worth noticing that the three experiments are referring to extremely different settings in terms of the underlying mobility pattern. As an example, Intel Exp2 refers present the results derived from nodes meeting in a relatively closed environment, such as the one performed at IEEE INFOCOM05, while the MIT Reality Mining project is considering the meetings of a nodes occurring at any time and place during the 1 year of experimentation. This results in extremely different meeting patterns, experimentation duration and number of nodes participating in the measurements.

Trace	Number of Nodes	Measurement Time (s)	Number of Records
Intel Exp1	128	359190	2766
Intel Exp2	223	522387	6732
Intel Exp3	264	255168	28216
UMass	22	83890	1504
Reality Mining	2135	3167388	25000
STUDNEWT Data 1	21	1371585	64160
STUDNEWT Data 2	23	3030076	14796

Table 1. Traces details for the four sets of data considered

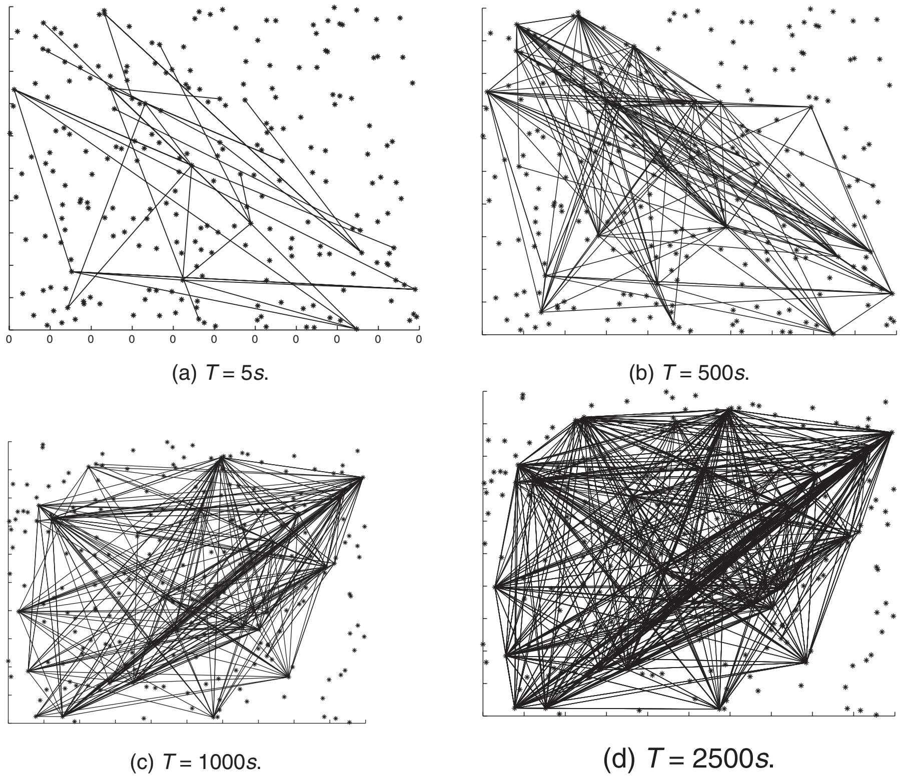

Fig. 2. T-tolerant connectivity graph for various values of T, INTEL Exp3 trace

As a first example, in Fig. 2, we plotted the resulting T-tolerant connectivity graph for the Intel Exp3 trace, with four different choices of the T value. Each node is assigned a random position in the unit square, chosen accordingly to a uniform distribution. A link between any two nodes is drawn if and only if a contact between them was observed within the specified time window T. The values considered for T are $5, 500, 1000, 2500$ seconds. Clearly, the specific connectivity graph evolution depends from the particular trace chosen. For the considered case, it can be observed that after 2500 seconds a significant number of nodes already experienced more than one meeting.

For all the datasets considered, we took T as the measurement times of the traces. It turned out, indeed, that taking long values of T does not impact the EVC analysis, at least for the traces we are considering.

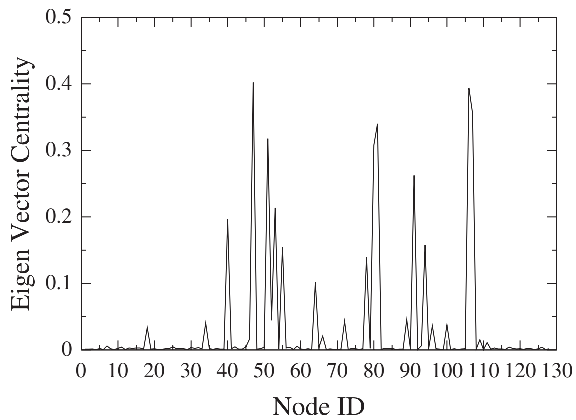

Fig. 3. Eigenvector centrality in the case of the INTEL Exp1 trace

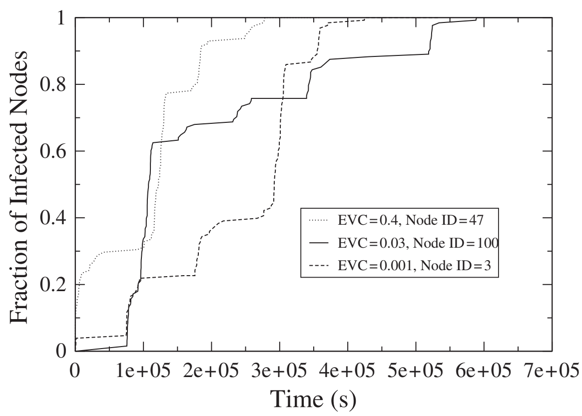

Fig. 4. Fraction of infected nodes over time, in the case of the INTEL Exp1 trace, and a variable *infecting* node

We started by considering both EVC and the spreading process as a function of the initiator (infecting) node. For all the experiments, we first computed the EVC, as detailed in Sec. 2. Then, we emulated the process of spreading (using the contacts in the corresponding trace file) starting from different "infecting" nodes. The infected process did not include any limitation in the number of copies each node could make or on the number of hops each message could traverse.

At time 0, one node is classified as "infected", and the remaining ones as "susceptible". Infected nodes spread the epidemics at any contact with susceptible ones. In order to avoid border effects due to the finiteness of the traces, traces were arranged in a cyclic fashion. The spreading process is considered concluded when all the nodes in the network are in the infected state.

In Fig. 3, we plotted the distribution of the EVC in the case of the Intel Exp1 trace, for the different nodes in the system. As it can be seen, the EVC is highly non-uniform, with nodes 1 and 9 presenting an EVC value far above the other nodes. In Fig. 4, we plotted the corresponding fraction of infected nodes vs. time for an epidemic starting from node 1 ($EVC = 0.65$), node 12 ($EVC = 0.03$) and node 7 ($EVC = 0.142$). As expected, the EVC turns out to give a significant measure of the ability of the node to initiate an epidemic. Please note, however, that the ability to spread an epidemics and the ability to initiate a spreading are, in general, different. Indeed, the choice of a given initiator node impacts mainly the first phase of the spreading process. For example, the curve corresponding to node 12 shows a sudden increase at about $0.85 \cdot 10^5$ seconds, and a detailed trace analysis showed that this corresponds to the time instant it meets with node 1, the one with the largest EVC value.

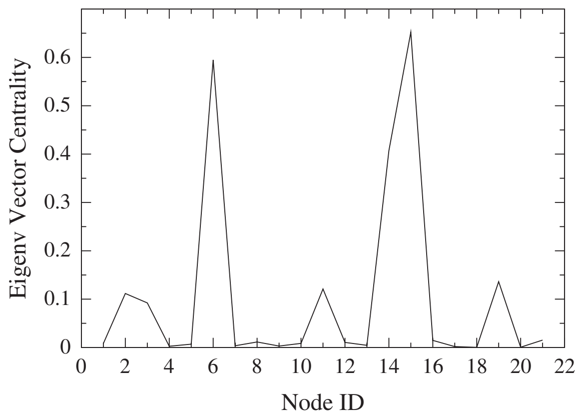

Fig. 5. Eigenvector centrality in the case of the STUDNETW Data1 trace

Similarly, Fig. 5 depicts the EVC distribution in the case of the STUDENT-Net data1 experimentation. Also in this second case, the EVC is not distributed uniformly, meaning that is reasonable to expect a different capacity of the nodes to diffuse information. This is confirmed in Fig. 6, where it is possible a observe a different infection depending from the initiator node. In this case, the difference does not seem to be as pronounced as in the INTEL Exp 1 case. This is mostly due to

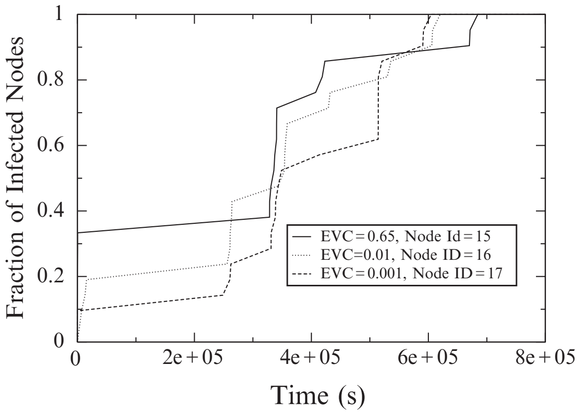

Fig. 6. Fraction of infected nodes over time, in the case of the STUDNETW Data 1 trace, and a variable *infecting* node

the limited number of nodes participating to the experimentation, and to the limited duration of the measurement.

The same analysis is applied to the Intel Exp2 trace, with the epidemics starting from nodes 3 ($EVC = 0.66$), node 19 ($EVC = 0.07$) and node 2 ($EVC = 0.002$), respectively. The results are depicted in Fig. 7 and Fig. 8. Also in this case, centrality scores are highly non-uniform, and the graph shows that the choice of the infecting node has a remarkable impact on the speed at which an epidemic spreads in the system. As for the Intel Exp3 trace, node 25, characterized by a low EVC value, at time $7 \cdot 10^4$ presents a sudden spreading increase derived from the meeting with a node with a high EVC value. It is also worth noticing that the time taken to infect the whole system seems to be not very sensitive to the EVC of the initiator node. We conjecture that such effect is due to the presence in the system of a low number of nodes which seldom meet other ones in the system, representing therefore a sort of “bottleneck” for the epidemics to spread the entire network.

We have then considered the traces from the UMASS DieselNet project experiment. Also for such case we evaluated the EVC on the whole trace duration (in this case, one day) and emulated the spreading process with different starting nodes. The results, in terms of eigenvector centrality and fraction of nodes infected with various initiator nodes, are reported in Fig. 9 and Fig. 10. It can be seen that the distribution is highly non-uniform, corresponding to a different ability of nodes to spread an epidemic in the system. Differently from other experiments, in the UMASS DieselNet trace the reported meetings occurred only among nodes participating to the experiment.[6] As a result, the meetings pattern shows a higher regularity, and the EVC distribution is more uniform, if compared with the Intel Exp1 and Intel Exp2 EVC analysis. Correspondingly, to higher values of EVC correspond a higher capability

[6] When leveraging on the Bluetooth device-discovery capabilities for reporting neighboring nodes, any Bluetooth-enabled device is reported.

of nodes both to initiate a spreading and to spread an epidemics. As an example, in Fig. 10 node 19 (EVC=0.5) not only presents the highest initial infection rate, but also infects the entire network in a short time, if compared with nodes with lower EVC values (i.e., node 2 and node 6).

In Fig. 11 and Fig. 12, the EVC and the spreading power analysis has been extended to the MIT Reality Mining experiment. Also in this case, the EVC distribution is highly non-uniform; this means that there is a large variation in spreading power among the nodes of the network. Node 3 presents the highest EVC value, and, correspondingly, a regular spreading pattern can be observed from Fig. 12. Conversely, node 22 presents an extremely low EVC score. This is reflected in an extremely low infection for the first $5 \cdot 10^5$ seconds, and a sudden and extremely fast diffusion afterwards. This behavior can be easily explained with node 22 meeting a node characterized by an extremely large value of EVC. Also in this case the fact that the various curves tend to converge can be due to a small set of student in the experiment which tend to remain isolated from the rest of the network.

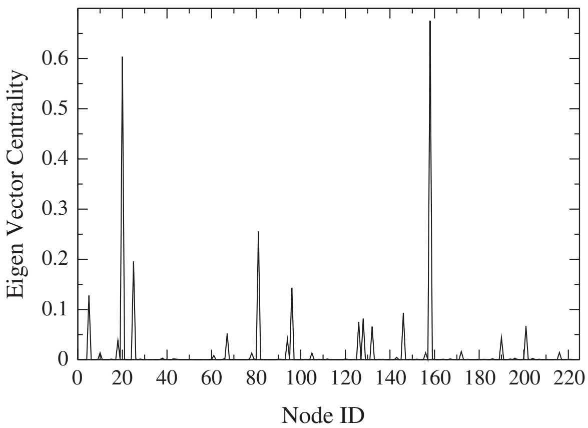

Fig. 7. Eigenvector centrality in the case of the INTEL Exp2 trace

The outcomes are different for the Intel Exp3 trace. Indeed, as it can be seen from the EVC, plotted in Fig. 13, just a fraction of the nodes in the system shows a non-negligible level of interactions. This can be due to the fact that all contacts with Bluetooth-enabled devices are recorded, not just with the other people in the experiment. A large fraction of the nodes appear for a limited number of times in the traces (in most cases, actually just once). The result is that the dependence on the EVC of the initiating node is only partial, in that it is valuable only for the nodes "truly" in the system. The other nodes appear mostly just once in the traces, so they cannot be infected until the very only meeting happens. For this reason, we decided not to report the graph of the fraction of infected nodes.

From all the considered real world deployments, we can reasonably conclude that the EVC analysis, applied to the T-tolerant connectivity matrix, returns meaningful

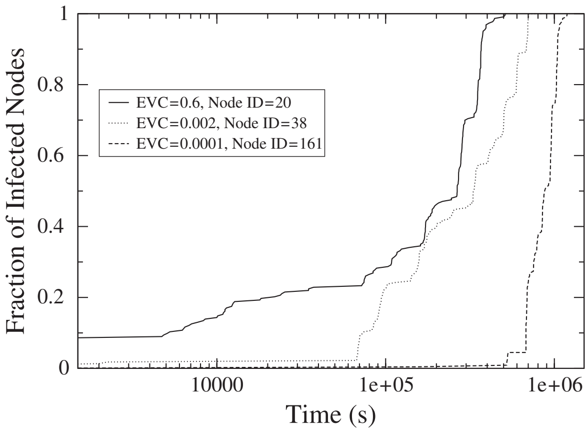

Fig. 8. Fraction of infected nodes over time, in the case of the INTEL Exp2 trace, and a variable *infecting* node

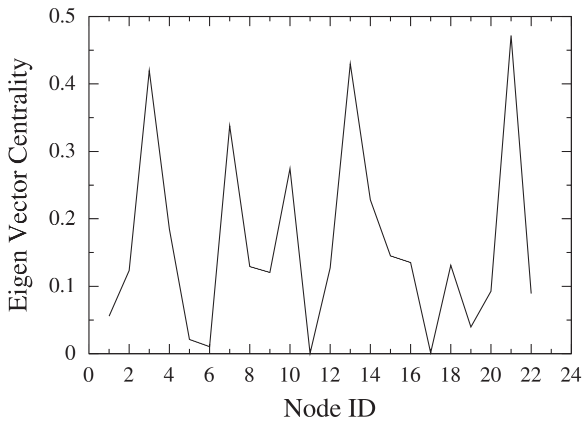

Fig. 9. EVCentrality in the case of the UMASS trace

performance metrics for assessing the ability of nodes to spread epidemics within a mobile disconnected network. The impact of such metric on the actual epidemics spreading pattern depends on the particular scenario considered. There might be cases in which nodes are able to infect the entire system in very short time, while in other circumstances a node with a large EVC may initiate a spreading process which, after a fast initial ramp-up, takes very long times to reach all nodes. The rationale underpinning this behavior resides in the level of partition of the network. Nodes meeting on a more regular basis contribute to the initial boost of the spreading process, while nodes meeting only rarely (or only once) are responsible for the heavy tail of the epidemic spreading process. The EVC value of the initiator node impacts

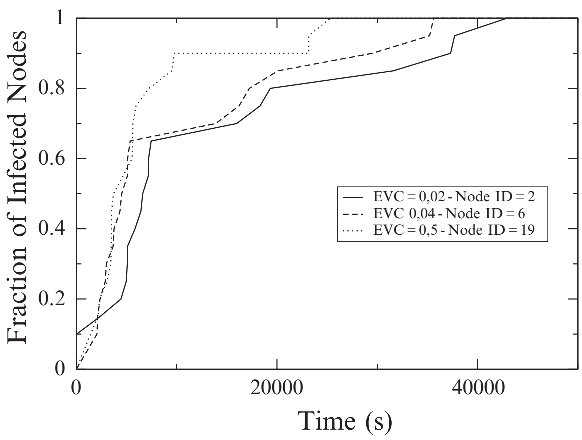

Fig. 10. Percentage of infected nodes over time, in the case of the UMASS trace, and a variable *infecting* node

mostly the initial part of the spreading, and only partially the subsequent spreading of messages in the system.

Region Analysis

In the following, a simplified region analysis is performed for the data set corresponding to the case of the UMass DieselNet project trace, with a time window of 8000 seconds. By following the steepest ascent rule [9] an ancestor is iteratively assigned to each node, until a local maximum is reached. The local maximum corresponds to a *centre* of the region [9]. From the conducted analysis, the network results to be partitioned in 2 distinct regions, node 1 being the centre of the first region with a EVC of 0.3822, and node 10 of the second one with an EVC of 0.5537. The first region consists of a total of two nodes, whereas the latter region consists of 13 nodes.

In Fig. 14, the resulting graph is depicted, with the corresponding EVC values. Region centres are depicted with larger circles; the nodes belonging to the region of node 10 are gridded, while the ones belonging to the region of node 1 are plotted in black. As can be seen from the figure, centre nodes are characterized by a high connectivity degree. Further, note that the two centres are not directly connected. (They cannot be, since each one is a local maximum of the EVC.) Thus, any path between the two centres has to go through another node which bridges the two regions.

Finally, in Fig. 15 and Fig. 16, the effect of the regions centre is highlighted in the case the UMass DieselNet project trace, with a time window of 8000 seconds. In Fig. 15, node 3 is starting the epidemics spreading. When node 10 becomes infected, it is possible to observe a significant boost in the spreading process. This effect is even more evident when the epidemics spreading reaches nodes 1: in less than 5000 seconds the remaining nodes of the network are infected. This is due to the well-connectedness of the region centre nodes. It is possible to observe a similar behavior

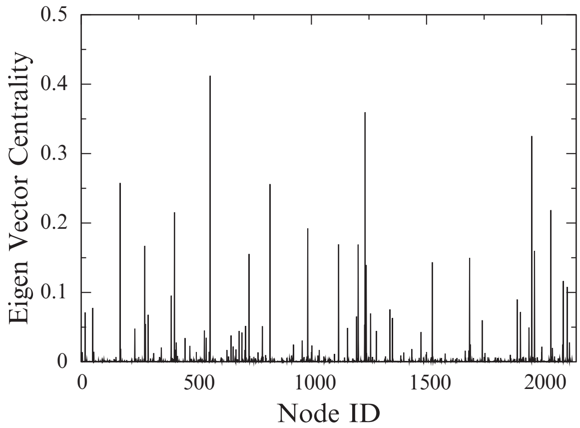

Fig. 11. Eigenvector centrality in the case of the MIT trace

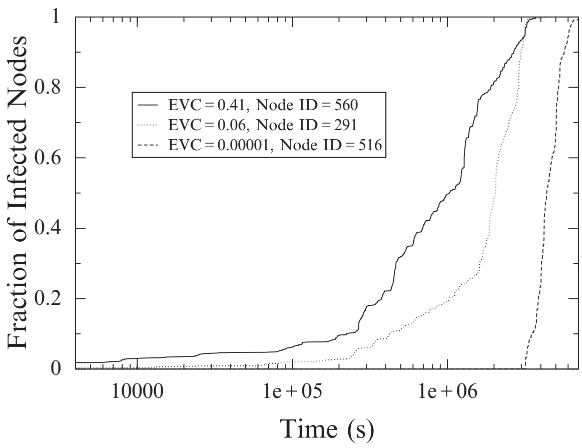

Fig. 12. Percentage of infected nodes over time, in the case of the MIT trace, and a variable *infecting* node

in Fig. 16, where node 2 is the node initiating the epidemics spreading. Node 2 belongs to the region of node 1, and, therefore, node 1 is infected before node 10 (as opposed to the previous case).

The region analysis has been applied to a variety of traces. We took a set of 6 traces, and considered for each of them two cases. In the first one, we used the whole trace, whereas in the second one we limited the analysis to only the first half of the records. This corresponds to choosing two different values of T. The results are summarized in Table 2. Some interesting remarks can be inferred from the results therein. First, all traces whose analysis was limited to the first half of the records returned a disconnected (highly partitioned) graph. On the other hand, all full trace files lead to a connected system. This emphasizes the importance of a correct choice

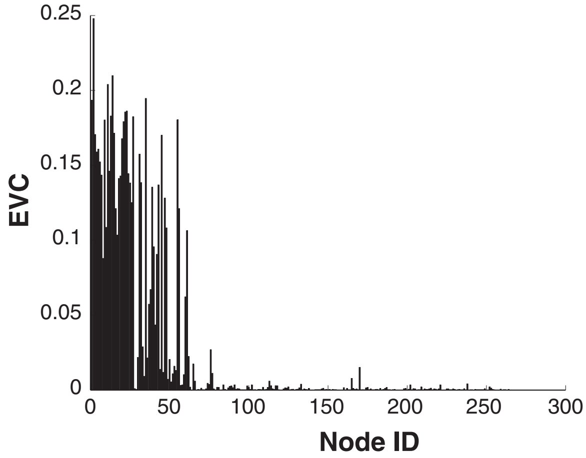

Fig. 13. Eigenvector Centrality in the case of the INTEL Exp3 trace

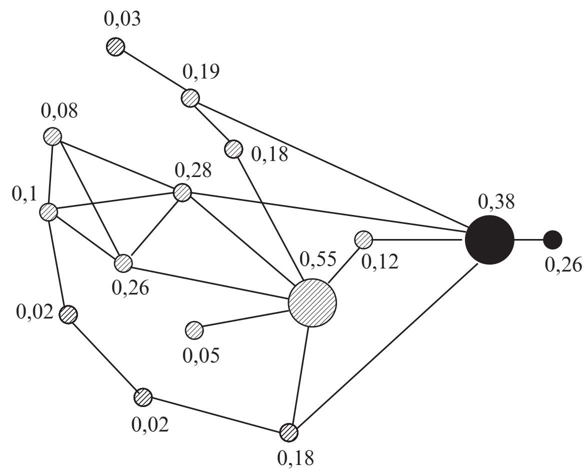

Fig. 14. Region analysis in the case of the UMASS trace, and time window of 8000 seconds

of the parameter T, which plays a key role in ensuring that the EVC analysis returns meaningful results. Second, all full traces returned a single region. We conjecture that such result is due to the fact that all experiments targeted rather homogeneous scenarios (people in a lab, attendees at a conference, student in one class, buses in a town etc.), where nodes tend naturally to form one single cluster. It remains therefore still an open issue to understand what the results of such analysis could be when applied to a real-world deployment, not limited to the simple experiments we considered in this chapter.

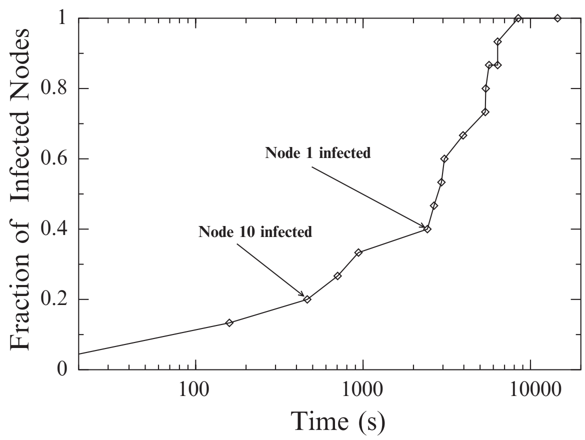

Fig. 15. Fraction of infected nodes, with a time window of 8000 seconds, and node 3 starting the epidemics spreading

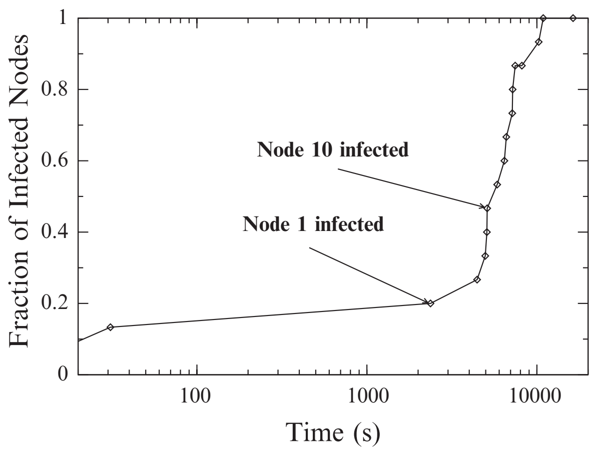

Fig. 16. Fraction of infected nodes, with a time window of 8000 seconds, and node 2 starting the epidemics spreading

5 Conclusions & Open Issues

In this chapter we have addressed the problem of enhancing forwarding mechanisms in disconnected wireless networks with information concerning the "fitness" of the nodes, understood as a measure of their ability to spread epidemics. The reason for such a study comes from the observation that, in the case of highly partitioned mobile networks, information can be diffused system-wide only by opportunistically exploiting the contacts among nodes, and some form of epidemic forwarding is necessary to achieve such purpose.

Trace	Number of Regions	Connected	Connected Components
INTEL Exp1 (full)	1	1	1
INTEL Exp1 (half)	25	0	25
INTEL Exp2 (full)	1	1	1
INTEL Exp2 (half)	117	0	117
INTEL Exp3 (full)	1	1	1
INTEL Exp3 (half)	92	0	92
MIT (full)	1	1	1
MIT (half)	389	0	389
STUDNETW Data1 (full)	1	1	1
STUDNETW Data1 (half)	1	1	1
STUDNETW Data2 (full)	1	1	1
STUDNETW Data2 (half)	1	1	1
UMASS 2072005 (full)	1	1	1
UMASS 2072005 (half)	3	0	3

Table 2. Results of the region analysis for various trace files

Based on earlier work addressing the problem of understanding the spread of epidemics in static wired networks, we identified the EVC as a suitable metric. The T-tolerant connectivity matrix concept was introduced as a mean to enable the application of the EVC analysis to the disconnected wireless scenario under study.

The resulting method has been applied to a wide range of trace files, coming from real deployments of delay-tolerant networks. The results reported show that the EVC, coupled with the T-tolerant connectivity matrix, is able to provide a good understanding of the ability of nodes in such systems to spread information on a network-wide scale.

It is not obvious that the EVC approach—which works well for networks with a fixed topology—should also work well when applied, via use of the T-tolerant connectivity matrix, to a dynamic network. We recall the discussion around Figure 1, where we showed that the use of the connectivity matrix can imply the existence of transmission possibilities which in fact do not exist in the true, dynamic network. So we ask, why do we find such a good correspondence between the predictions of the static EVC picture and the true epidemic simulations?

We believe that the answer lies in the fact that, for these traces, many pairs of nodes meet repeatedly. We note, for example, that, if A and B meet repeatedly in the course of the experiment depicted in Figure 1, then the communication paths shown on the right hand side (from the connectivity matrix) will also in fact exist in the true, dynamic, network. That is, repeated meetings eliminate most or all of the artifactual errors introduced by forming the connectivity matrix.

Furthermore, we note that we give most weight in the connectivity matrix to links which involve repeated meetings. Conversely, links involving only a single meeting—which is thus never repeated in the course of the trace—receive the lowest possible nonzero weight in the connectivity matrix. Hence, our approach gives most weight to node pairs which meet repeatedly; and our EVC analysis works best when such pairs dominate the analysis. Thus we get a consistent picture of why the EVC analysis can work well for such dynamic networks.

While the results presented in this chapter show that EVC has the potential for enhancing existing forwarding protocols proposed for disconnected wireless networks, many open issues are still present. The first one is concerned with the possibility of computing on-line the EVC in a distributed fashion. While this problem has been tackled in [3], the application to the present scenario is not straightforward. Each node could indeed well count the meetings having taken place (over a time interval T) with other nodes; this would give each node its own row (= column) of the T-tolerant connectivity matrix. With this information, each node could also exchange its current EVC value with those nodes it meets. We recall from [3] that a distributed EVC calculation needs two things: (i) iterated weight passing (taking also into account link weights), and (ii) periodic normalization of the weight vector (which otherwise grows to infinity). Now we see that operation (i) is possible for such a dynamic net. The problem is then (ii).

Normally, one wishes to rescale the weight vector by dividing by a local estimate of the eigenvalue. In [3], a method is given for finding this estimate. This method assumes however constant connectivity, so that it can converge much faster than the slower, weight passing operation. Here we study networks which do not have such constant connectivity. Hence we consider other approaches.

- Each node can simply use the *sum* of its link weights as an estimate of its centrality. This is equivalent to a single iteration of the EVC power method (starting from a uniform start vector). For cases where the power method converges rapidly, this may be a good estimate of the true EVC; and in any case, it represents useful and easily accessible information.
- Two iterations are also possible. That is, each node gets (after enough meetings) the one-iteration estimates (column sums) for all of its neighbors, multiplies these by the corresponding link weights, and takes the sum of these, over all neighbors, as its two-iteration estimate.
- We see from considering one and two iterations that, in principle, any finite number n of iterations may be managed (with a corresponding cost in storage, signalling, and clock time needed). Each estimate, when finally obtained by a node, must simply be stored, labelled (by its iteration number), and passed on (when the opportunity arises). Thus, in real-world networks, we can imagine that $n > 2$ can be practical, giving a good, or even very good, approximation to the true EVC—*without* the need to ever estimate the global eigenvalue λ_{max}.

We see from the above that a reasonable estimate of the EVC may be obtained by dynamic networks of the type we study here. An implicit precondition for this idea to work is that the nodes' pattern of movements and meetings must be reasonably stable and repetitive. The reason for this is that (a) the nodes need a 'startup' time T to obtain estimates of the link weights; (b) the nodes then need some time (depending on the desired iteration number n) to compute the centrality estimates based on the stored connectivity matrix; and (c) if the pattern of meetings has *changed* significantly since the startup time, then the answer will cease to be valid. In other words, the mobility pattern must be, in some statistical sense, stable over a time scale which is considerably longer than the time needed to compute the EVC estimates—which is, roughly, $O(n \times T)$.

If this precondition is not even met for $n = 1$, we see no way for such a network to compute, using its own meetings, any reasonable EVC estimate. However, any network with this degree of dynamism will in fact render invalid the use of the T-tolerant connectivity matrix. We note that we get some indication of repetitive movement from the good results we have found in this paper; but of course this question must be studied more carefully, by examining the traces in detail. We also leave for future work the problem of how the network itself can discover which n (if any) is appropriate to its own mobility patterns.

Finally, it remains also to devise how such a metric — whatever way it gets computed — can be effectively used for enhancing the performance of epidemics-style forwarding mechanisms for disconnected wireless networks. The knowledge of the fitness value of the encounters represents indeed additional information which could be used to optimize the performance of the system. While it is easy to introduce simple techniques for making use of such knowledge, it is nonetheless unclear at the moment what could be the optimal way of exploiting it for optimizing the system performance.

Acknowledgments

The work of I. Carreras and D. Miorandi has been partially supported by the European Commission within the framework of the BIONETS project IST-FET-SAC-FP6-027748, `www.bionets.eu`. The work of G. Canright and K. Engø-Monsen was partially supported by the Future and Emerging Technologies unit of the European Commission through Project DELIS (IST-2002-001907).

References

1. A. A. Hanbali, P. Nain, and E. Altman, “Performance of two-hop relay routing protocol with limited packet lifetime,” in *Proc. of ValueTools*, Pisa, Italy, 2006.
2. X. Zhang, G. Neglia, J. Kurose, and D. Towsley, “Performance modeling of epidemic routing,” in *Proc. of Networking*, 2006.
3. G. Canright, K. Engø-Monsen, , and M. Jelasity, “Efficient and robust fully distributed power method with an application to link analysis,” Department of Computer Science, University of Bologna, Tech. Rep. UBLCS-2005-17, 2005. [Online]. Available: `http://www.cs.unibo.it/bison/publications/2005-17.pdf`
4. V. Cerf, S. Burleigh, A. Hooke, L. Torgerson, R. Durst, K. Scott, K. Fall, and H. Weiss, “Delay-tolerant network architecture,” 2005, iETF Internet Draft. [Online]. Available: `http://www.dtnrg.org/wiki`
5. K. Fall, “A delay-tolerant network architecture for challenged Internets,” in *Proc. of ACM SIGCOMM*, Karlsruhe, DE, 2003.
6. A. Khelil, C. Becker, J. Tian, and K. Rothermel, “An epidemic model for information diffusion in manets,” in *Proc. of ACM MSWiM*, 2002.

7. I. Carreras, I. Chlamtac, F. De Pellegrini, and D. Miorandi, "Bionets: Bio-inspired networking for pervasive communication environments," *IEEE Trans. Veh. Tech.*, 2006, in press. [Online]. Available: http://www.create-net.org/$\sim$dmiorandi
8. T. Small and Z. Haas, "The shared wireless infostation model Ű a new ad hoc networking paradigm (or where there is a whale, there is a way)," in *Proc. of ACM MobiHoc*, 2003, pp. 233–244.
9. G. Canright and K. Engø-Monsen, "Roles in networks," *Science of Computer Programming*, vol. 53, pp. 195–214, 2004.
10. G. Canright and K. Engo-Monsen, "Spreading on networks: a topographic view," in *Proc. of ECCS*, Paris, 2005.
11. P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *Journal of Mathematical Sociology*, vol. 2, pp. 113–120, 1972.
12. H. Minc, *Nonnegative matrices*. New York: J. Wiley and Sons, 1988.
13. R. Meester and R. Roy, *Continuum Percolation*. New York: Cambridge Univ. Press, 1996.
14. U. Lee, E. Magistretti, B. Zhou, M. Gerla, P. Bellavista, and A. Corradi, "MobEyes: smart mobs for urban monitoring with vehicular sensor networks," UCLA CSD, Tech. Rep. 060015, 2006. [Online]. Available: http://netlab.cs.ucla.edu/wiki/files/mobeyestr06.pdf
15. P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. on Inf. Th.*, vol. 46, no. 2, pp. 388–404, Mar. 2000.
16. M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. on Netw.*, vol. 10, no. 4, pp. 477–486, Aug. 2002.
17. J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine, "MaxProp: routing for vehicle-based disruption-tolerant networks," in *Proc. of IEEE INFOCOM*, Barcelona, ES, 2006.
18. F. D. Pellegrini, D. Miorandi, I. Carreras, and I. Chlamtac, "A graph-based model for disconnected ad hoc networks," in *Proc. of IEEE INFOCOM*, 2007.
19. CRAWDAD, the community resource for archiving wireless data at Dartmouth. [Online]. Available: http://crawdad.cs.dartmouth.edu/
20. A. Chaintreau, P. Jui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on the design of opportunistic forwarding algorithms," in *Proc. of IEEE INFOCOM*, Barcelona, ES, 2006.
21. The disruption tolerant networking project at UMass. [Online]. Available: http://prisms.cs.umass.edu/diesel/
22. Machine perception and learning of complex social systems. [Online]. Available: reality.media.mit.edu/
23. J. Su, A. Goel, and E. de Lara, "An empirical evaluation of the student-net delay tolerant network," in *Proc. of MOBIQUITOUS*, San Jose, US, July 2006.