

Department of Mathematics and Statistics

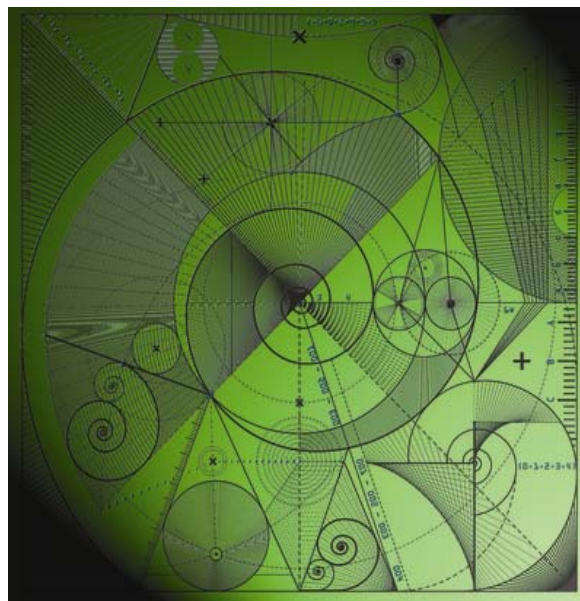
Preprint MPS-2013-09

10 April 2013

Centrality and spectral radius in dynamic communication networks

by

Danica Vukadinović Greetham, Zhivko Stoyanov
and Peter Grindrod



Centrality and spectral radius in dynamic communication networks

Danica Vukadinović Greetham, Zhivko Stoyanov, and Peter Grindrod

Centre for the Mathematics of Human Behaviour
Department of Mathematics and Statistics
University of Reading, UK
{d.v.greetham, z.v.stoyanov,p.grindrod}@reading.ac.uk

Abstract. We explore the influence of the choice of attenuation factor on Katz centrality indices for evolving communication networks. For given snapshots of a network observed over a period of time, recently developed communicability indices aim to identify best broadcasters and listeners in the network. In this article, we looked into the sensitivity of communicability indices on the attenuation factor constraint, in relation to spectral radius (the largest eigenvalue) of the network at any point in time and its computation in the case of large networks. We proposed relaxed communicability measures where the spectral radius bound on attenuation factor is relaxed and the adjacency matrix is normalised in order to maintain the convergence of the measure. Using a vitality based measure of both standard and relaxed communicability indices we looked at the ways of establishing the most important individuals for broadcasting and receiving of messages related to community bridging roles. We illustrated our findings with two examples of real-life networks, MIT reality mining data set of daily communications between 106 individuals during one year and UK Twitter mentions network, direct messages on Twitter between 12.4k individuals during one week.

Keywords: evolving networks, spectral radius, centrality ranking

1 Introduction

Today's interconnected world with millions of users of mobile devices, computers and sensors leaving digital traces provides social scientists with previously unseen opportunities to create and validate their theories on a large scale. These social networks, captured in digital world, present us with research challenges: they are large, multi-layered and dynamic, i.e. they evolve from moment to moment. Thus, there is a need for the methods developed for regular and arbitrary static networks to be extended and adapted to dynamic, evolving networks.

One of the very important and well researched characteristics of an individual (a node) in a social network is its centrality score. Centrality measures the relative importance of a node and determines its involvement in a network. Although different centrality measures were proposed, tested and compared on undirected,

directed and weighted networks (for reviews see [3,16]), only relatively recently research focused on centrality in evolving networks [9,13].

For static networks, Katz centrality [15] computes the relative influence of a node within a network by measuring the number of the immediate neighbors and all the other nodes in the network that connect to the node under consideration through the immediate neighbors. Walks made to distant neighbors are penalised by an attenuation factor α . This concept was recently revisited in [9,13]. Communicability across time-steps is based on the extension of Katz centrality to evolving networks. The concept is already successfully implemented on a small scale mobile phone and email communication networks [13,12] and C. Elegans brain networks [6], however scaling it up to very large data-sets involves handling large matrices.

Another measure of centrality in static networks, Bonacich centrality [1], [2] introduces another parameter, similar to Katz centrality, but penalising direct and indirect links. Recently, this measure was revisited in [11] where the authors investigated whether the measure converges, and proposed a normalised variant. Although their motivation came from the claim that computing a spectral radius ρ_A (the largest eigenvalue of an adjacency matrix of a graph) is difficult, “especially for large networks” (see pp. 2 of [11]), we note that in social network analysis settings where networks are mostly sparse, power iteration or similar methods could be used, that efficiently compute an approximated value of ρ even for very large sparse matrices, or else Perron-Frobenius theorem (see e.g. [10]) provides simple but useful bounds. However, we argue that a constraint such as $\alpha < \frac{1}{\rho(A)}$ is limiting and should be relaxed for different reasons, particularly as it might penalise too heavily not-so-long paths, and thus lower significantly the centrality ranking of nodes that connect different communities which might have implications in large social network analysis. The “structural holes” theory [4] refers to the absence of links between two parts of a network. Brokerage exploits structural holes – an individual is connected to two other individuals or communities not mutually connected. This position could be beneficial for such an individual (a broker) as she/he could control a flow of information between two communities, profit from two different sources of information and mediate trade between them. Also bridges between different communities in a social network are important when trying to identify communities in the large unknown network and to run a network-based intervention which depends on community structure to change behaviour of individuals in the network [17].

In the empirical analysis of several real-world and artificial generic models of networks, Jamaković et al. [14] looked at the different upper bounds of spectral radius from the simple bound given by the graph’s maximal degree to more complex bounds featuring local information (average neighbours degree) or global information such as diameter. They found that for three real-world networks they investigated, a bound given by [7] was the closest to the observed values, while for the Internet autonomous systems topology the same bound was over-estimating the real value significantly. When using artificial generic networks (random, small-world, preferential attachment networks) with the same number

of nodes and edges as in the real-world networks, the spectral radii of all three types of networks were much smaller than the real-world one in the Internet AS topology case. This is important because the spectral radius is also found to be connected to epidemic spreading in networks (see [18,5]).

In the following section we discuss how communicability indices are related to spectral radius and propose a new centrality measure which relaxes convergence constraints previously imposed by the spectral radius. We then create vitality measure based on centrality indices and show how to detect the individuals whose lack of existence would result in the biggest changes in centrality in evolving networks. We apply our findings to two real-life networks, and conclude with the discussion.

2 Relaxed Communicability

An evolving network is a family of graphs $G_i = (V, E_i)$, where the vertex set V is given in advance and is fixed throughout time, and an edge set E_i is a set of edges on V in the time i . We assume that the time is discrete and finite, i.e. $i = 1, \dots, n$. The corresponding adjacency matrices are denoted with A_i .

2.1 Communicability

For a static matrix A , the Estrada-Hatano communicability indices for a matrix A can be obtained from a communicability matrix as the row/column sums, given by

$$Q = e^A \quad (1)$$

where the matrix exponential of a matrix A is defined as

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k \quad (2)$$

This can be extended to evolving networks by $Q = \prod_{i=0}^n e^{A_i}$, which can be computed directly for small-scale networks. Another version of a communicability matrix, closer to the Katz definition, from [13] is given by:

$$Q = \prod_{i=0}^n (I - \alpha A_i)^{-1} \quad (3)$$

where I is identity matrix, $\alpha < \frac{1}{\max(\rho(A_i))}$, $i = 0, \dots, n$ denotes consecutive time-steps and $\rho(A_i)$ is the spectral radius of A_i . Henceforth, we refer to (3) as the “*standard communicability*”. Broadcast and receive indices are equal to the row, resp. column sums of Q .

2.2 Spectral Radius Bound

In the case where existing communicability indices are used (3), Katz centrality for each A_i can be written as

$$(I - \alpha A_i)^{-1} = \sum_{k=0}^{\infty} \alpha^k A_i^k \quad (4)$$

and in order that (4) converges in standard matrix norm, one has that the attenuation factor $\alpha < \frac{1}{\rho(A_i)}$ and similarly $\alpha < \frac{1}{\max(\rho(A_i))}$, $i = 0, \dots, M$ for (3). On the other hand, looking at each individual A_i , if α is interpreted as a probability that, once sent, a message will be successfully transmitted by any receiving node to any of its contacts, then the expected length of a single transmission sent from nodes in the network corresponding to A_i is

$$\sum_{k=0}^{\infty} k \alpha^k (1 - \alpha) = \frac{\alpha}{(1 - \alpha)} \quad (5)$$

This implies for matrices with a spectral radius of more than 3, we must choose $\alpha < \frac{1}{\max(\rho(A_i))}$, the expected value of transmission length will be less than $\frac{1}{2}$, and if a spectral radius is greater than 2, the expected transmission length is less than 1.

As it was shown in [14] some of real-world networks have radius greater than 2, resulting in expected path lengths between 1 and 2, which means that especially paths between two communities are too heavily penalised. In order to mitigate the attenuation, we propose to normalise A and relax the condition on the attenuation which allows for longer paths and so rewards individuals that act as bridges between different communities appropriately.

2.3 Relaxed Communicability

For a large data-set where the size of the matrix Q is prohibitive, and computing inverse of such a large matrix represents a challenge, an approximation of Q can be computed using a Taylor series approximation ignoring summands of order higher than some n , depending on the application. In order to compute $(I - \alpha A)^{-1} \mathbf{1}$ without storing Q , the following method can be used where \mathbf{b} is initialised to the all ones vector of length n .

$$(I - \alpha A)^{-1} \mathbf{b} = \mathbf{b} + \alpha A \mathbf{b} + \alpha^2 A^2 \mathbf{b} + \dots \quad (6)$$

We will use this representation to define new relaxed communicability indices. Instead of having $\alpha = \frac{1}{2 \max(\rho(A_i))}$, for the expression to converge, it is enough for α to be less than 1, and that A is normalised. From the expression for the expected path length (5), ensuring that

$$\alpha < \frac{l}{1 + l} \quad (7)$$

where $l \in \mathbb{N}$ is the expected path length, we have that α will always be less than 1 and we can set parameter l on a desired path length depending on a context, i.e. what kind of centrality we are interested in. Thus, to obtain relaxed communicability indices, one should choose a length of path l depending on the application, calculate α from (7) for given l , initialise b to be all-ones vector and multiply it with α and matrix A_i normalised with 2-norm of A_i iteratively. Summing up all iterative factors up to the order n , which depends on how small one's approximation error needs to be, gives the result for A_i . Results need to be multiplied for all consecutive A_i s. In the case of a small graph, Q can be obtained directly from (3) using computed α and replacing A with $\frac{A}{\|A\|}$.

2.4 Vitality Measure

In order to rank the nodes by importance during a time period we formulated vitality-based measure by computing the corresponding centrality indices in the absence of one node at time. For a series of adjacency matrices A_{i_1}, \dots, A_{i_2} we compute communicability indices using both standard and relaxed communicability indices. Furthermore, for each vertex k , we compute Q_k , which is obtained deleting exactly the k th row and the k th column from A_{i_1}, \dots, A_{i_2} , and then calculating both versions of communicability. Then we calculate the difference between Q_k indices and Q for each k , as a sum of least squares to check which nodes are responsible for the biggest changes in indices' values. We give a pseudo-code for vitality measure on Fig 1, which is independent of the version of communicability used (standard or relaxed). We will discuss in the next section how results depend on the type of communicability used.

3 Applications

We used two real-world data sets. The first one is the mutual mobile phone communications over a year for 106 individuals which was captured as a part of the MIT reality mining data set [8]. The second is the data-set obtained from Twitter UK mentions network collected on our behalf by Datasift, Twitters certified partner. The network was created from public messages that users located in UK sent to each other on Twitter using @ sign during 1 week in Dec 2011. In both cases we aggregated data on daily basis.

3.1 Case-study 1: MIT Reality Mining Data

Data is aggregated on a daily basis, and contains 365 binary adjacency matrices, from the 20th July 2004 onwards, denoted with A_1 to A_{365} . An entry (i,j) of A_k is equal to 1 if there was at least one phone call between i and j on day k . On Fig 2 given are the spectral radii of all 365 matrices. One can observe how the communication structure changes through the year.

On Fig 3 we show an example of a daily communication network. The vertices with labels 10, 45, 59 and 71 (highlighted on Fig 3) have relatively small degree,

VITALITY MEASURE

```
compute indices (column and row sums of Q for A_i_1..A_i_2)
ls=0
for j=1:N
    remove A_k(j, .) and A_k(., j) for k=i_1,..,i_2
    compute indices_j (column and row sums of Q_j)
end
for i =1:j-1
    ls(j,1)=ls(j,1)+(indices_j(i,1)-indices(i,1))^2;
    ls(j,2)=ls(j,2)+(indices_j(i,2)-indices(i,2))^2;
end;
for i=(j+1):n
    ls(j,1)=ls(j,1)+(indices_j(i-1,1)-indices(i,1))^2;
    ls(j,2)=ls(j,2)+(indices_j(i-1,2)-indices(i,2))^2;
end;
ls=sqrt(ls)
```

Fig. 1. Computing vitality measure. Indices are 2xn array - the first column is Q's column sum, and the second is Q's row sum. Q_j is obtained from all A_i s removing j-th column and j-th row from each adjacency matrix, and $indices_j$ are then column and row sum of Q_j .

but they connect different communities and therefore are important. We picked a sample of seven daily networks on the days 32, 55, 96, 135, 158, 220 and 315, looking at the different values of spectral radii. We computed communicability indices using the standard and relaxed versions. The Table 1 presents results of rankings in descending order (1 top to 106 bottom) in both cases, showing much higher rankings when the relaxed version with the length of path 3 was used. On Fig 4 we show scatter plot of standard vs. relaxed broadcast indices (left panel) and standard vs. Estrada-Hatano communicability indices on the right panel. The upper left diagonal of the figure represent nodes that have higher rankings in relaxed than in standard indices. Note that in both standard and relaxed indices more weight for broadcast indices lies on the first matrix in the sequence, while for receive indices it is the last matrix that carries most of weight. While Estrada-Hatano indices do not correlate with standard or relaxed indices, they still rank higher most of community bridges. Thus Estrada-Hatano indices could be used when the expected transmission length is not known, but if the length of transmission is important, our parametric approach will highlight more relevant nodes.

3.2 Case-study 2: Twitter Mentions Network Data-set

The data-set comprised of around a million of tweets between UK users that contained mention of another UK user (sign @). The nodes represented the

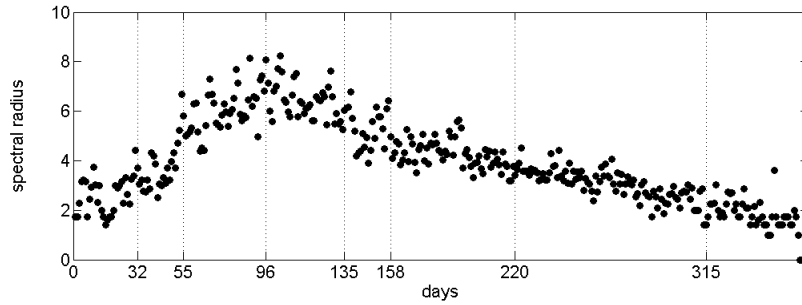


Fig. 2. MIT data: Spectral radii of A_1 to A_{365} matrices.

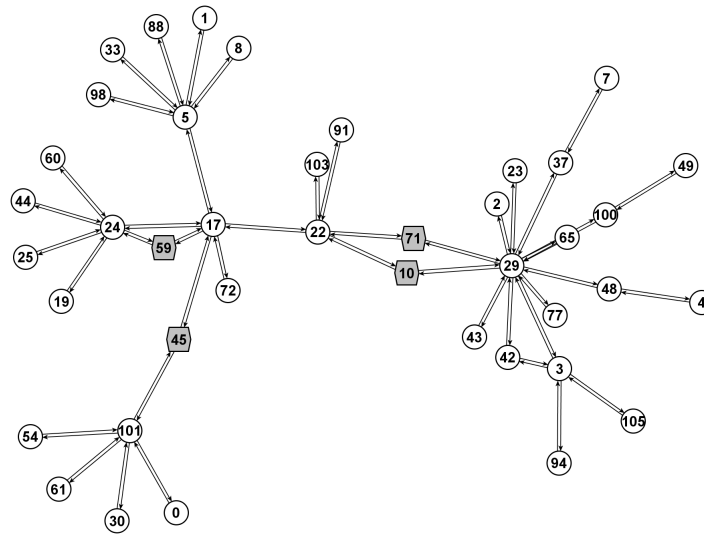


Fig. 3. MIT data: An example of a daily network (its largest connected component), on the day 32.

Table 1. Ranking (in descending order, top 1 to bottom 106) of broadcast vs. relaxed broadcast

Vertex	Rank (broadcast)	Rank (relaxed broadcast)
10	71	21
45	35	25
59	85	28
71	71	21

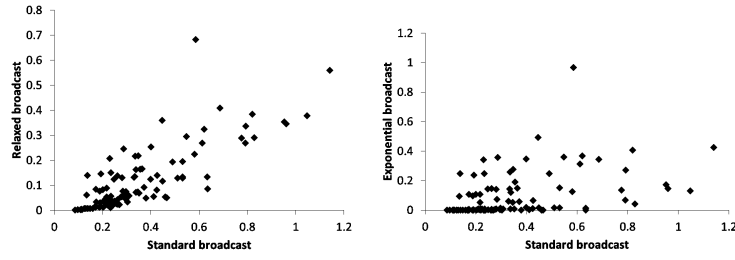


Fig. 4. Standard vs. relaxed broadcast indices, left, and standard vs. exponential broadcast indices, right.

users, and if user A’s tweet contained ”@B”, an edge between A and B was created. Only reciprocated edges were kept and multi-edges were ignored. All daily tweets were aggregated into a daily network, so we finished with 7 daily undirected graphs with 12408 nodes and around 2.7k edges in average. We computed both communicability and relaxed communicability indices, both using rank obtained from communicability, and rank obtained from vitality based measure (deleting each node and computing the sum of differences for all the other nodes as described earlier).

3.3 Results

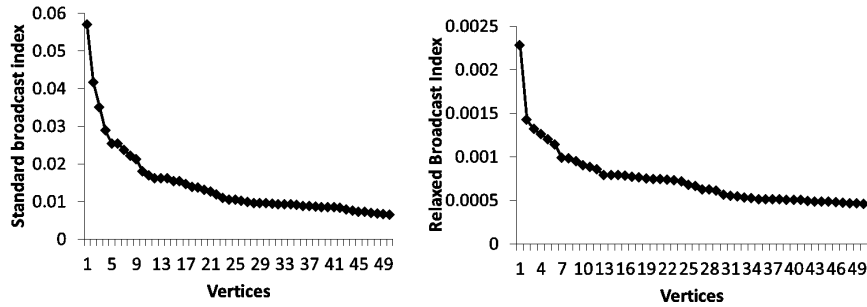


Fig. 5. Top 50 vertices according to the ranking based on standard (left) and relaxed (right) broadcast.

Although the computation of vitality measure is quite demanding (one needs to recompute communicability matrices for each node once) this is feasible as the daily networks are quite sparse. At 12408 vertices and 7 time-steps, this collection contains relatively big, but not large networks. Their broadcast indices decrease quickly so we ranked the indices from largest to smallest with respect

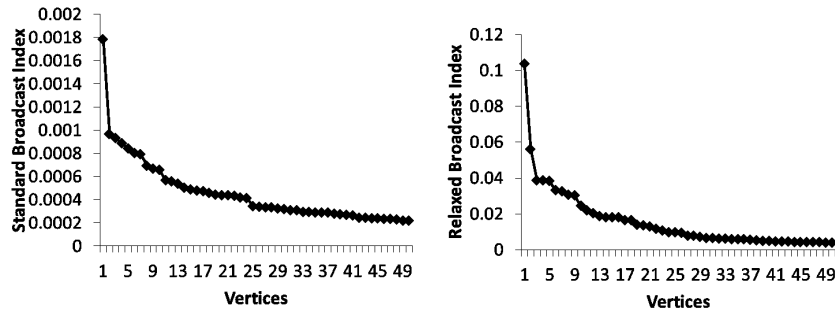


Fig. 6. Top 50 vertices according to the ranking based on standard (left) and relaxed (right) vitality.

to broadcast and looked into more details at the first fifty indices. On Fig 6 one can see the difference in ranking between the two methods. Several vertices that are ranked much higher in relaxed than in standard broadcast index correspond again to vertices with relatively small degrees and were picked up as they connect different communities (e.g. vertex ranked 39 in relaxed is ranked 278 in normalised and has a degrees equal to $(2, 0, 0, 3, 1, 0, 0)$ respectively in 7 daily networks.)

4 Conclusions

We used communicability indices to rank the nodes in evolving communication networks. While the computation of communicability for small-data sets is relatively simple and fast, for the large data-sets it means handling of large matrices, so one can use a Taylor approximation. We introduced a parameter (transmission length) that allows for targeting specifically brokers or bridges between communities. We have applied this approach on two real-life evolving networks obtained from mobile phone communications and Twitter. Using the vitality based measure, we proposed a way to rank vertices depending on the amount of change their communication abstinence would bring to the rest of the evolving network. We hope that a parametric approach that can be optimised according to a particular application will be a useful addition to a standard evolving social network analysis toolbox, especially when the expected length of message/communication transmission plays an important role, i.e. it is either given or can be approximated.

Acknowledgments. This work is funded by the RCUK Digital Economy programme via EPSRC grant EP/G065802/1 'The Horizon Hub' and EPSRC MOLTEEN EP/I016031/1. We would like to thank Datasift for providing us with the Twitter dataset.

References

1. P. Bonacich. Power and centrality: A family of measures. *American Journal of Sociology*, 92:1170–1182, 1987.
2. Phillip Bonacich and Paulette Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23(3):191 – 201, 2001.
3. Stephen P. Borgatti and Martin G. Everett. A graph-theoretic perspective on centrality. *Social Networks*, 28(4):466 – 484, 2006.
4. Ronald S. Burt. Brokerage and closure: An introduction to social capital. *Eur Sociol Rev*, 23(5):666–667, 2007.
5. C. Castellano and R. Pastor-Satorras. Thresholds for epidemic spreading in networks. *Physical Review Letters*, 105:218701, 2010.
6. J. J. Crofts and D. J. Higham. Googling the brain: Discovering hierarchical and asymmetric network structures, with applications in neuroscience. *Internet Mathematics (Special Issue on Biological Networks)*, 2011.
7. Kinkar Ch. Das and Pawan Kumar. Some new bounds on the spectral radius of graphs. *Discrete Mathematics*, 281(1-3):149 – 161, 2004.
8. N. Eagle, A. S. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106:15274–15278, 2009.
9. E. Estrada and N. Hatano. Communicability in complex networks. *Physical Review E*, 77, 2008.
10. Felix Gantmacher. *The Theory of Matrices, Volume 2*. AMS Chelsea Publishing, 2000.
11. Rumi Ghosh and Kristina Lerman. Parameterized centrality metric for network analysis. *Physical Review E*, 83(6):066118+, June 2011.
12. P. Grindrod and D. J. Higham. Models for evolving networks: with applications in telecommunication and online activities. *IMA Journal of Management Mathematics*, to appear.
13. P. Grindrod, D. J. Higham, M. C. Parsons, and E. Estrada. Communicability across evolving networks. *Physical Review E*, 83, 2011.
14. A. Jamaković, R.E. Kooij, P. Van Mieghem, and E.R. van Dam. Robustness of networks against viruses: the role of the spectral radius. In *Symposium on Communications and Vehicular Technology, 2006*, pages 35 –38, November 2006.
15. L. Katz. A new index derived from sociometric data analysis. *Psychometrika*, 18:39–43, 1953.
16. Tore Opsahl, Filip Agneessens, and John Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245, 2010.
17. Th. Valente. Network interventions. *Science*, 337(6090), 2012.
18. Yang Wang, Deepayan Chakrabarti, Chenxi Wang, and Christos Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *In SRDS*, pages 25–34, 2003.