Social Networks xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Resistance distance, closeness, and betweenness

Enrico Bozzo, Massimo Franceschet*

Department of Mathematics and Computer Science, University of Udine, Via delle Scienze 206, 33100 Udine, Italy

ARTICLE INFO

Keywords:
Information
Resistance distance
Geodesic distance
Closeness centrality
Betweenness centrality

ABSTRACT

In a seminal paper Stephenson and Zelen (1989) rethought centrality in networks proposing an information-theoretic distance measure among nodes in a network. The suggested information distance diverges from the classical geodesic metric since it is sensible to all paths (not just to the shortest ones) and it diminishes as soon as there are more routes between a pair of nodes. Interestingly, information distance has a clear interpretation in electrical network theory that was missed by the proposing authors. When a fixed resistor is imagined on each edge of the graph, information distance, known as resistance distance in this context, corresponds to the effective resistance between two nodes when a battery is connected across them. Here, we review resistance distance, showing once again, with a simple proof, that it matches information distance. Hence, we interpret both current-flow closeness and current-flow betweenness centrality in terms of resistance distance. We show that this interpretation has semantic, theoretical, and computational benefits.

1. Introduction

A large volume of research on networks has been devoted to the concept of centrality, in particular closeness and betweenness centrality measures. Historical references are Shimbel (1953), Beauchamp (1965), Sabidussi (1966) and Freeman (1977), which are all based on early intuitions of Bavelas (1948). An up-to-date account on centrality measures on networks is Chapter 7 of Newman (2010). Typically, geodesic (shortest) paths are considered in the definition of both closeness and betweenness centrality. These are optimal paths with the lowest number of edges or, if the graph is weighted, paths with the smallest sum of edge weights.

The use of shortest path has, however, some drawbacks (Newman, 2005). In many cases, shortest paths form a small subset of all paths between two nodes; it follows that paths even slightly longer than the shortest one are not considered at all in the definition of centrality. Furthermore, the geodesic distance between two nodes – the length of the shortest path between the nodes – does not consider the actual number of (shortest) paths that lie among the two vertices: two nodes that are separated by a single path have the same distance of two nodes that are separated by many paths of the same length. In many applications, however, paths longer than geodesic ones are also relevant, since information or whatever flows on the networks does not necessarily choose an optimal path. In a social network, for instance, a fad does not know the optimal route to move among actors, but it simply wanders around more or less randomly. Moreover, nodes separated by many pathways are often perceived closer than nodes separated by few pathways, even if the paths have all the same length. This because communication between nodes is typically enhanced as soon as more routes are possible (Stephenson and Zelen, 1989; Klein and Randić, 1993).

To overcome these limitations, alternative notions of closeness and betweenness that are not based on the notion of shortest path, or on that of optimal path is some other sense, have been proposed, independently, by Newman (2005) and Brandes and Fleischer (2005). These alternative centrality measures can be explained both in terms of electric current flowing trough a resistor network or using the notion of random walk through the graph. In this work we use the current-flow analogy and hence we refer to these measures as *current-flow closeness* and *current-flow betweenness*. The idea is to view a network as a resistor network in which the edges are resistors and the nodes are junctions between resistors. Each edge is possibly assigned with a positive weight indicating the conductance (the reciprocal of the resistance) of the edge. Hence, the distance between two nodes i and j is defined as the potential difference of nodes i and j when a unit of current is injected in source i and removed from target j; since the current is equal to unity, the potential difference is also the effective resistance between nodes i and j. A high resistance (potential difference) between nodes indicates that the two nodes are far away, while low resistance between nodes means that the nodes are close points. Current-flow closeness centrality of a node is defined as the reciprocal of the mean distance of the node from

* Corresponding author.
E-mail addresses: enrico.bozzo@uniud.it (E. Bozzo), massimo.franceschet@uniud.it (M. Franceschet).

0378-8733/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.socnet.2013.05.003

the other nodes of the network: a low average distance means high centrality, while nodes with a high average distance have low centrality (Brandes and Fleischer, 2005). Both Newman (2005) and Brandes and Fleischer (2005) propose also a current-flow version of betweenness centrality. For a given node, current-flow betweenness measures the current flow that passes through the vertex when a unit of current is injected in a source node and removed from a target node, averaged over all source–target pairs. Furthermore, they provide effective algorithms to compute current-flow centrality measures.

Current-flow closeness and betweenness are strongly related to an interesting, but, to our assessment, underestimated, notion of distance on graphs, known as *resistance distance*. It is defined, independently, by Stephenson and Zelen (1989), following an information-theoretic approach, and by Klein and Randić (1993), following an electrical-theoretic approach. In this paper we extensively review, with the aid of many examples, the notion of resistance distance, comparing it with the alternative notion of shortest-path distance. We then interpret both current-flow closeness and betweenness centrality in terms of resistance distance, that is, we rewrite the formulas that define these centrality measures in terms of the resistance distance matrix. This interpretation has some advantages. Semantically, it provides an alternative interpretation to current-flow centrality measures, in particular to betweenness centrality. Computationally, it allows to compute, or to approximate, both centrality measures in terms of the same resistance distance matrix. Mathematically, given the deep theoretical understanding of resistance distance, it allows to prove new statements, or to demonstrate old theorems in a shorter and simpler way, about current-flow closeness and betweenness.

The outline of the paper is as follows. In Section 2 we review the related literature. In Section 3 we recall the notions of graph Laplacian and of its generalized inverse, which are the building blocks to define resistance distance. The notions of resistor networks and resistance distance are reviewed in Section 4. In Sections 5 and 6 we interpret current-flow measures in terms of resistance distance, highlighting the semantic, computational, and mathematical benefits of this approach. We draw our conclusions in Section 7.

2. Related work

The notion of information centrality has been originally defined by Stephenson and Zelen (1989) using an information-theoretic approach, while resistance distance was proposed by Klein and Randić (1993) in an electrical-theoretic context. As shown by Brandes and Fleischer (2005) and also later in this paper, these two notions are intimately connected.

The notion of information centrality has been reviewed in Borgatti (2005), Brandes and Erlebach (2005) and Borgatti and Everett (2006). The most thorough mathematical study about resistance distance is provided by Ghosh et al. (2008). The authors show that the resistance distance notion satisfies many interesting mathematical properties and it has different intriguing interpretations. For instance, resistance distance is a metric on the graph and in particular the resistance distance matrix is an Euclidean distance matrix. Furthermore, the resistance distance is a monotone decreasing function as well as a convex function of the edge weight (conductance) vector of the graph. Resistance distance has an easy interpretation in terms of random walks on graphs. The resistance distance between nodes i and j of a graph is proportional to the average commute time of nodes i and j of the Markov chain defined by the graph, which is the average number of steps it takes to return to node i for the first time after starting from i and passing through j. Curiously, resistance distance has been also used in the context of bibliometrics to express a rational version of the popular Erdős number, a measure of collaboration distance from mathematician Paul Erdős[1](Erdős, 1972; Balaban and Klein, 2002).

Brandes and Fleischer (2005) define the notions of current-flow closeness and betweenness centrality using resistor networks and show that current-flow closeness centrality corresponds to information centrality. Independently, Newman (2005) defines current-flow betweenness centrality using both random walks and resistor networks and shows that the two approaches correspond. Both papers propose effective methods to compute the exact value of current-flow centrality measures. Bozzo and Franceschet (2012) devise methods for finding approximations of the generalized inverse of the graph Laplacian matrix, which arises in many graph-theoretic applications. In particular, they apply the devised methods to the problem of approximating current-flow betweenness centrality on a graph and experimentally demonstrate that the approximations are both efficient and effective. In fact, the notion of (edge) current-flow betweenness centrality can be found already in Newman and Girvan (2004), where the authors propose and study a set of algorithms for discovering community structure in networks based on the idea of interactively removing edges with high betweenness scores.

Noh and Rieger (2004) investigate random walks on complex networks and derive an exact expression for the mean first passage time between two nodes. Moreover, they introduce the notion of random walk centrality, which determines the relative speed by which a node can receive information from elsewhere in the network. The defined measure bears some similarity to current-flow closeness centrality (and hence to information centrality)[2].

Finally, it is worth noticing that current-flow betweenness (Brandes and Fleischer, 2005; Newman, 2005) is intrinsically different from the notion of flow betweenness defined by Freeman et al. (1991). Indeed, current-flow betweenness considers all paths between nodes, while flow betweenness counts only edge-independent paths between nodes.

3. The graph Laplacian and its generalized inverse

Let $\mathcal{G} = (V, E, w)$ be an undirected weighted graph with V the set of nodes, E the set of edges, and w a vector such that $w_i > 0$ is the positive weight of edge i, for $i = 1, \ldots, |E|$. We denote by n the number of nodes and m the number of edges of the graph. We will assume throughout the paper that the graph $\mathcal{G}$ is connected.

The weighted Laplacian of graph $\mathcal{G}$ is the symmetric matrix

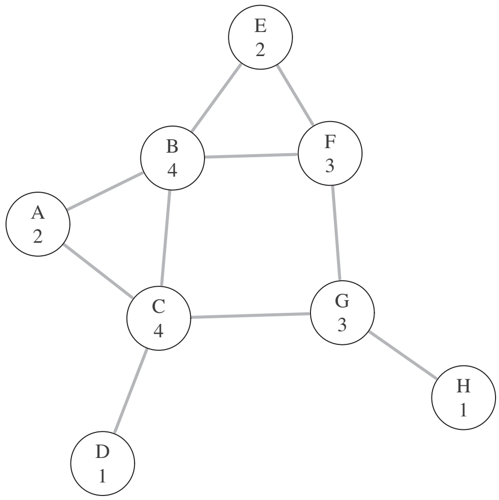
$$G = D - A,$$

where A is the weighted adjacency matrix of the graph and where D is a diagonal matrix such that the i-th element of the diagonal is equal to $\sum_j A_{i,j}$, that is, the (generalized) degree of node i. An example is shown in Fig. 1.

Since G is symmetric it has real eigenvalues. Moreover, all rows of G sum to 0, that is $Ge = 0$, where e is a vector of all ones. Hence 0 is an eigenvalue of G, having the vector e as associated eigenvector. It can be shown, see for example (Ghosh et al., 2008), that the other

[1] Paul Erdős was an eccentric Hungarian mathematician who is currently the most prolific and the most collaborative among mathematicians. He wrote more than 1400 papers cooperating with more than 500 co-authors.

[2] We conjecture that random walk centrality and current-flow centrality are, in general, different measures, and they coincide for the class of graphs such that all nodes have the same (generalized) degree (in the unweighted case, these graphs are called regular).



	A	B	C	D	E	F	G	H
A	2	−1	−1	0	0	0	0	0
B	−1	4	−1	0	−1	−1	0	0
C	−1	−1	4	−1	0	0	−1	0
D	0	0	−1	1	0	0	0	0
E	0	−1	0	0	2	−1	0	0
F	0	−1	0	0	−1	3	−1	0
G	0	0	−1	0	0	−1	3	−1
H	0	0	0	0	0	0	−1	1

Fig. 1. A network and its Laplacian matrix. Edge weights are all equal to 1. Nodes are labelled with their degree (the number of incident edges).

eigenvalues are positive so that, if we denote them with λ_i, for $i = 1, \ldots, n$, they can be sorted as follows:

$$0 = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_n.$$

Having 0 as an eigenvalue, G is singular and cannot be inverted. As a substitute for the inverse we use the Moore–Penrose generalized inverse of G, that we simply call generalized inverse of G (Ben-Israel and Greville, 2003). As customary, we denote this kind of generalized inverse with G^+. It is convenient to define G^+ starting from the spectral decomposition of G. Given a vector v, the square diagonal matrix whose diagonal entries are the elements of v is denoted with Diag(v). Actually, since G is symmetric it admits the spectral decomposition

$$G = V \Lambda V^T,$$

where $\Lambda = \mathrm{Diag}(0, \lambda_2, \ldots, \lambda_n)$ and $VV^T = I = V^T V$. The columns of V are eigenvectors of G normalized in a such a way their length is one. In particular the first column of V is $e/\sqrt{n}$. By using the spectral decomposition of G, its generalized inverse can be defined as follows

$$G^+ = V \mathrm{Diag}(0, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_n}) V^T. \quad (1)$$

Thus G^+ is also symmetric and it has 0 as eigenvalue associated with the eigenvector e. The eigenvalues of G^+ different from the smallest one are the reciprocal of the eigenvalues of G. Since $Ge = 0 = G^+ e$ by using the symmetry of the two matrices it follows that $e^T G = 0^T = e^T G^+$. If we set $J = ee^T$, that is J is a matrix of all ones then

$$GJ = JG = G^+ J = JG^+ = O, \quad (2)$$

where O is a matrix of all zeros. Moreover, from the definitions it follows that

$$GG^+ = G^+ G = I - \frac{1}{n} J. \quad (3)$$

and that

$$G^+ = \left(G + \frac{1}{n} J \right)^{-1} - \frac{1}{n} J, \quad (4)$$

a formula that can be found in Ghosh et al. (2008) and is used implicitly in Brandes and Fleischer (2005). Hence, the generalized inverse G^+ of G can be obtained by inverting a suitable perturbed version of G and then subtracting the perturbation.

The generalized inverse of the graph Laplacian G is particularly useful to represent the solution of a linear system of the form $Gv = b$ for some known vector b. It can be shown that the system has solutions if b sums up to zero, and if this is the case $v^* = G^+ b$ is a solution. Any other solution of the system can be obtained by adding to v^* a multiple of e. Since $Jv^* = JG^+ b = 0$, the entries of v^* sum up to zero as well, and this characterizes v^* in the set of solutions. For the sake of completeness we mention that in the case where b does not sum up to zero then the system $Gv = b$ has no solution. However in this case $v^* = G^+ b$ can be shown to minimize the length of the residual $b - Gv^*$.

4. Resistance distance

In this section we first introduce resistor networks, which are functional to the definition of resistance distance. We then survey the notion of resistance distance, illustrating some of its interesting mathematical properties with the aid of some examples. In particular, we introduce and discuss the notion of mean resistance distance between nodes in a network, a measure of the largeness of a network in terms of resistance distance.

4.1. Resistor networks

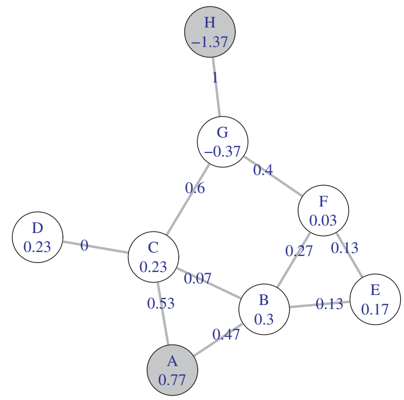
Consider a network in which the edges are resistors and the nodes are junctions between resistors. Each edge is possibly assigned with a positive weight indicating the conductance of the edge. The resistance of an edge is the inverse of its conductance. Outlets are particular nodes where current enters and leaves the network. A vector u called supply defines them: a node i such that $u_i \neq 0$ is an outlet; in particular if $u_i > 0$ then node i is a source and current enters the network through it, while if $u_i < 0$ then node i is a target and current leaves the network through it. Since there should be as much current entering the network as leaving it, we have that $\sum_i u_i = 0$. We consider the case where a unit of current enters the network at a single source s and leaves it at a single target t. That is, $u_i^{(s,t)} = 0$ for $i \neq s, t$, $u_s^{(s,t)} = 1$, and $u_t^{(s,t)} = -1$. We are interested in how current flows through the network, for an arbitrary choice of source and target outlets.

Let $v_i^{(s,t)}$ be the potential of node i, measured relative to any convenient reference potential, for source s and target t outlets. Kirchhoff's law of current conservation states that the current that enters in a node is equal to the current that flows out. This implies that the node potentials satisfy the following equation for every node i:

$$\sum_j A_{i,j} (v_i^{(s,t)} - v_j^{(s,t)}) = u_i^{(s,t)}, \quad (5)$$

where A is the weighted adjacency matrix of the network. Actually, the current flow through edge (i, j) is the quantity $A_{i,j}(v_i^{(s,t)} - v_j^{(s,t)})$, that is, the difference of potentials between the involved nodes multiplied by the conductance of the edge: a positive value indicates that the current flows in a direction (say from i to j), and

Please cite this article in press as: Bozzo, E., Franceschet, M., Resistance distance, closeness, and betweenness. Soc. Netw. (2013), http://dx.doi.org/10.1016/j.socnet.2013.05.003



	A	B	C	D	E	F	G	H
A	0.50	0.07	0.06	−0.06	−0.07	−0.09	−0.14	−0.26
B	0.07	0.27	−0.01	−0.13	0.09	0.04	−0.11	−0.23
C	0.06	−0.01	0.25	0.13	−0.11	−0.10	−0.05	−0.17
D	−0.06	−0.13	0.13	1.00	−0.24	−0.22	−0.17	−0.30
E	−0.07	0.09	−0.11	−0.24	0.55	0.14	−0.11	−0.24
F	−0.09	0.04	−0.10	−0.22	0.14	0.35	0.00	−0.12
G	−0.14	−0.11	−0.05	−0.17	−0.11	0.00	0.35	0.23
H	−0.26	−0.23	−0.17	−0.30	−0.24	−0.12	0.23	1.10

Fig. 2. A resistor network with all resistances equal to unity. Each node is identified with a letter and is labelled with the value of its potential when a unit current is injected in node A and removed from node H. Each edge is labelled with the absolute current flowing on it. Each node potential can be obtained as difference between entries of columns A and H of the generalized inverse matrix of the Laplacian of the network, which is shown below the network. Notice that all values are approximated with precision of two digits.

negative value means that the current flows in the opposite direction. It is interesting to observe that, if $i \neq s, t$ then Eq. (5) can be rewritten as

$$v_i^{(s,t)} = \frac{\sum_j A_{i,j} v_j^{(s,t)}}{\sum_j A_{i,j}},$$

so that the potential of a non-outlet node is the weighted mean of the potential of its neighbors. This implies that the potential of a non-outlet lies between the minimum and the maximum of the potential of its neighbors.

In matrix form, Eq. (5) reads:

$$(D - A)v^{(s,t)} = Gv^{(s,t)} = u^{(s,t)}. \tag{6}$$

Recall that $G = D - A$ is the graph Laplacian matrix. As noticed in Section 3, if G^+ is the Moore–Penrose generalized inverse of the Laplacian matrix G, then a solution of Eq. (6) is given by:

$$v^{(s,t)} = G^+ u^{(s,t)}. \tag{7}$$

This means that the potential of node i with respect to source s and target t outlets is given by:

$$v_i^{(s,t)} = G^+_{i,s} - G^+_{i,t}. \tag{8}$$

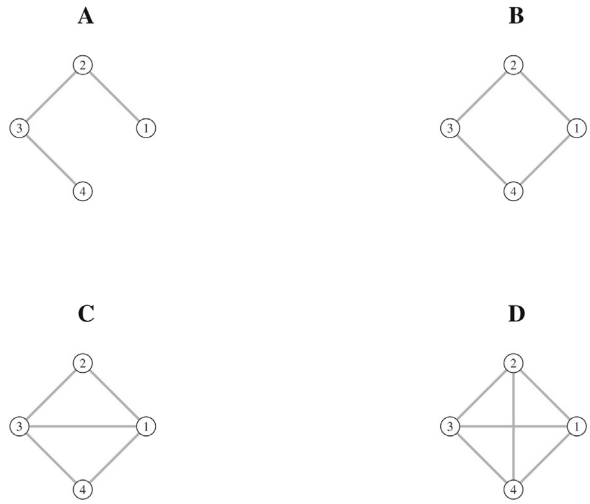
Therefore, the generalized inverse matrix G^+ contains information to compute all node potentials for any pair of source–target nodes.

An example of resistor network with node potential solution is provided in Fig. 2. Notice that Kirchhoff's law is satisfied for each node. For instance, the current entering in node B is 0.47 (from node A) which equals the current leaving node B, which is again 0.47 (0.13 to E, 0.27 to F, and 0.07 to C). Moreover, the current leaving the source node A is 1, and the current entering the target node H is also 1. Notice that there is no current on the edge from C to D, since both nodes have the same potential. Moreover, observe that the potential of non-outlets is the mean of the potential of their neighbors. The effective resistance between A and H is the potential difference between A and H, which is 2.14. We will see briefly that this can be considered as a (resistance) distance between nodes A and H. Any other potential vector obtained from the given solution by adding a constant is also a solution, since the potential differences remain the same, and hence Kirchhoff's law is satisfied. The given potential vector is, however, such that its entries sum up to zero.

4.2. Resistance distance

Typically, distance on graphs is defined in terms of the length (or weight) of the geodesic – the shortest path – between two nodes. This distance, which is in fact a metric on the graph, is called geodesic or shortest-path distance. Geodesic distance, however, has a couple of drawbacks: (i) paths longer than the shortest one give no contribution to the measure, and (ii) the number of paths lying between two nodes is irrelevant. An alternative notion of distance that takes account of these issues is defined, independently, by Stephenson and Zelen (1989) and by Klein and Randić (1993). The new distance, which is called *resistance distance* in Klein and Randić (1993), has the following characteristics:

- Multiple paths. The existence of multiple paths between two nodes reduces the distance: two nodes separated by many paths are closer than two nodes separated by fewer paths of the same length. Using an information-theoretic perspective (Stephenson and Zelen, 1989), the information contained in many paths is higher of the information contained in fewer ones, and communication is enhanced among more informative channels. Using a current flow analogy (Klein and Randić, 1993), many resistors in parallel offer less effective resistance than fewer resistors in parallel;



$$R_A = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix} \qquad R_B = \begin{bmatrix} 0.00 & 0.75 & 1.00 & 0.75 \\ 0.75 & 0.00 & 0.75 & 1.00 \\ 1.00 & 0.75 & 0.00 & 0.75 \\ 0.75 & 1.00 & 0.75 & 0.00 \end{bmatrix}$$

$$R_C = \begin{bmatrix} 0.000 & 0.625 & 0.500 & 0.625 \\ 0.625 & 0.000 & 0.625 & 1.000 \\ 0.500 & 0.625 & 0.000 & 0.625 \\ 0.625 & 1.000 & 0.625 & 0.000 \end{bmatrix} \qquad R_D = \begin{bmatrix} 0.0 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.0 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.0 \end{bmatrix}$$

Fig. 3. Four graphs with increasing densities and the corresponding distance matrices.

- Path redundancy. Two nodes separated by a set of edge-independent paths – paths that taken in pairs do not share edges – are closer than two nodes divided by redundant paths – paths that taken in pairs share some edges. In a way, highly redundant paths are closer to a single route, while edge-independent paths are actually different routes;
- Path length. Two nodes separated by a shorter path – a path with less edges – are closer than two nodes set apart by a longer path. In terms of information, the noise along a longer route is higher, hence information is lower, and communication is hindered (Stephenson and Zelen, 1989). Similarly, the resistance of a long series of resistors is higher than that of a short chain of resistors (Klein and Randić, 1993).

These characteristics are effectively illustrated in Fig. 3, where four graphs A, B, C, and D, with increasing densities, are depicted. Notice that the first graph A is a acyclic (it is a tree), hence each pair of nodes is connected by exactly one path, while the last graph D is complete, thus each pair of nodes is connected by an edge. The corresponding resistance distance matrices are given below the graphs. In the acyclic case, graph A, geodesic distance and resistance distance correspond. In graph B, nodes 1 and 3 are separated by two paths of length 2. The geodesic distance between them remains 2 as in graph A, but the resistance distance reduces to 1. Similarly, nodes 1 and 2 are separated by two paths of lengths 1 and 3: their geodesic distance is 1, and their resistance distance is 0.75. In graph C, nodes 1 and 3 are reachable via three edge-disjoint paths of lengths 1, 2, and 2. Their resistance distance still decreases to 0.5, while the geodesic distance is 1. Nodes 1 and 2 are now reachable via three paths of lengths 1, 2 and 3, two of them share an edge. Their resistance distance is 0.625, a bit larger than the equivalent distance for nodes 1 and 3, and still smaller than their geodesic distance of 1. Finally, in the complete graph D, all pairs of nodes are reachable via five paths of length 1, 2, 2, 3, and 3, which however share some edges. The resistance distance is equal to 0.5 and the geodesic distance is equal to 1 for each pair of nodes. Notice that this distance is exactly that of nodes 1 and 3 in graph C, which are reachable via three edge-disjoint paths of lengths 1, 2, and 2: the larger number of paths among two nodes in graph D is balanced by the larger redundancy of these paths.

An interesting property of the resistance distance, which is useful to illuminate its meaning, is that, given any undirected, unweighted, and connected graph with n nodes, the sum of resistance distances between pairs of nodes connected by an edge is $n-1$, independently of the number of edges of the graph (we assume here that each pair defining an undirected edge is considered only once in the sum) (Klein and Randić, 1993). Notice that, the geodesic distance of two nodes connected by an edge is 1, and hence the sum of geodesic distances on a graph is the number of edges of the graph. If, for instance, the graph is acyclic, hence it is a tree, then resistance and geodesic distances are the same. This because in a tree there is a unique path between any two nodes, and the resistance distance for two nodes is the length of this path, that is, the geodesic distance. Since a tree of n nodes has $n-1$ edges, we have that the sum of distances on edges is the number of edges, that is $n-1$. However, the presence of cycles in the graph reduces the resistance distances in comparison with the geodesic counterparts, since, in general, more paths are available between pairs of nodes when loops are introduced in the graph. For a graph with the maximum number of cycles, a complete graph in which any pair of nodes is linked by an edge, each of the $n(n-1)/2$ pairs of nodes linked by an edge are distant $2/n$ in the resistance case, which is

Please cite this article in press as: Bozzo, E., Franceschet, M., Resistance distance, closeness, and betweenness. Soc. Netw. (2013), http://dx.doi.org/10.1016/j.socnet.2013.05.003

less than their geodesic distance of 1 as soon as $n > 2$. Notice that the sum of resistance distances on edges is again $n - 1$. In summary, the denser the graph, the more paths there are between nodes, and the smaller is resistance distance compared to geodesic distance. This can be seen on the sequence of graphs of increasing density depicted in Fig. 3.

Finally, we provide the formal definition of resistance distance. We view a network as a resistor network as described in Section 4.1. Given nodes i and j, the resistance distance $R_{i,j}$ between i and j is the effective resistance between i and j, that is, the potential difference of nodes i and j when a unit of current is injected from source i and removed from target j:

$$R_{i,j} = v_i^{(i,j)} - v_j^{(i,j)}. \tag{9}$$

The matrix R whose entries are $R_{i,j}$ is the resistance distance matrix. It is useful to express resistance distance in terms of the elements of the generalized inverse matrix of the Laplacian of the graph. Recall that, from Eq. (8), the potential of node i with respect to source s and target t is given by $v_i^{(s,t)} = G^+_{i,s} - G^+_{i,t}$. Therefore, we have that:

$$R_{i,j} = v_i^{(i,j)} - v_j^{(i,j)} = (G^+_{i,i} - G^+_{i,j}) - (G^+_{j,i} - G^+_{j,j}) = G^+_{i,i} + G^+_{j,j} - 2G^+_{i,j}. \tag{10}$$

In matrix form, we have that:

$$R = e\,\mathrm{diag}(G^+)^T + \mathrm{diag}(G^+)e^T - 2G^+, \tag{11}$$

where e is a vector with all components equal to 1, and $\mathrm{diag}(G^+)$ is the diagonal of matrix G^+.

4.3. Average resistance distance of a network

One property of large-scale networks that is frequently investigated is the average geodesic (shortest-path) distance between pairs of nodes in the network. One of the most discussed network phenomena is the *small-world effect*: in most real network the typical geodesic distance is surprisingly short, in particular when compared with the number of nodes of the network. This result, in the context of social networks, is the origin of the idea of the six degrees of separation, the popular belief that there are only about six steps between any two people in the world (Milgram, 1967), and even less (four) on online social networks such as Facebook (Backstrom et al., 2012). One might find interesting to investigate the average distance between nodes in a network in terms of resistance distance:

$$\delta = \frac{1}{n^2}\sum_{i,j} R_{i,j}. \tag{12}$$

Notice that, by Eq. (2), the elements of G^+ sum to 0. Exploiting this observation and Eq. (10) we have:

$$\delta = \frac{1}{n^2}\sum_{i,j} R_{i,j} = \frac{1}{n^2}\sum_{i,j}(G^+_{i,i} + G^+_{j,j} - 2G^+_{i,j})$$
$$= \frac{1}{n^2} 2\,n\,Tr(G^+) = \frac{2}{n} Tr(G^+).$$

But the trace of a matrix is the sum of its eigenvalues, and the eigenvalues of the generalized inverse G^+ are 0 and $1/\lambda_i$, for $i = 2, \ldots, n$, where λ_i are the eigenvalues of the Laplacian G. It follows that:

$$\delta = \frac{2}{n}\sum_{i=2}^{n}\frac{1}{\lambda_i}.$$

Hence the typical resistance distance between two nodes in a network is proportional to the arithmetic mean of the eigenvalues of G^+, which is the reciprocal of the harmonic mean of the eigenvalues of G. Recall that λ_2, the second smallest eigenvalue of the Laplacian, is known as the *algebraic connectivity* of the network and it is a measure of how easily a network can be divided into two disconnected components: the closer λ_2 to 0, the easier the network can be divided (in particular, if $\lambda_2 = 0$, then the network is not connected). However, if λ_2 is close to 0, then $1/\lambda_2$ is much larger than 0, and hence the average resistance distance is also large. Reasonably, networks that are easily separable are networks with large average (resistance) distances among nodes.

In general, the typical resistance distance of a network is mainly determined by the magnitude of the small eigenvalues of G, which correspond to the large eigenvalues of G^+. Zhan et al. (2010) showed that, for many network models including scale-free networks, there are strict similarities between the Laplacian eigenvalue distribution and the node degree distribution. Since the distribution of node degrees in a scale-free network is a power law, the same long-tailed distribution is expected for the Laplacian eigenvalues of a scale-free network. This means that most of the eigenvalues of the Laplacian have small values (the trivial many), and a significant few of them have very large values (the vital few). Interestingly, the magnitude of the typical resistance distance of a network is mainly determined by the trivial many eigenvalues of the Laplacian.

5. Current-flow closeness centrality

Closeness measures the mean distance from a node to other nodes of the network. A central node with respect to this measure is a vertex that is separated from others by only a short distance on average. Such a vertex might have better access to information (or to whatever else flows on the network) at other nodes or more direct influence on other nodes. For instance, in a social network, the opinions of a central actor might reach others in the community more quickly and, similarly, the viewpoints of the community actors might arrive earlier to the central actor.

The definition of closeness is parametric in terms of that of distance among nodes. Current-flow closeness centrality uses the notion of resistance distance discussed in Section 4.2. Hence, given two nodes i and j, the distance between them is the resistance distance $R_{i,j}$. The mean distance d_i of node i from the other nodes is then defined by:

$$d_i = \frac{\sum_j R_{i,j}}{n}, \tag{13}$$

and current-flow closeness centrality for node i is:

$$c_i = \frac{1}{d_i} = \frac{n}{\sum_j R_{i,j}}. \tag{14}$$

The lower the distances from a node to the other nodes, the higher the centrality of than node. Using Eq. (10) to express the resistance distance in terms of the generalized inverse of the Laplacian, we have that the sum of all distances from node i is equal to:

$$\sum_j R_{i,j} = nG^+_{i,i} + Tr(G^+) - 2\sum_j G^+_{i,j} = nG^+_{i,i} + Tr(G^+). \tag{15}$$

In the above equality we have used the fact that all rows of the generalized inverse G^+ sum to 0 (Eq. (2)). It follows that the mean distance of node i is given by:

$$d_i = G^+_{i,i} + \frac{Tr(G^+)}{n} = G^+_{i,i} + \frac{\delta}{2}, \tag{16}$$

where we recall from Section 4.3 that $\delta = 2Tr(G^+)/n$ is the mean resistance distance of the network. The above formulation of the average distance d_i of a node in a network is interesting since it distinguishes two contributions: a node-level component $G^+_{i,i}$, which indicates how close node i is within the given network, and a network-level component proportional to δ, which is symptomatic of how large is the network overall, independently of the particular node i.

Eq. (16) turns out to be useful to provide a precise interpretation of the elements of the generalized inverse G^+ of the Laplacian matrix. The diagonal element $G^+_{i,i}$ tells us about how close is node i, in terms of resistance distance, with respect to the rest of the graph. What about the off-label elements $G^+_{i,j}$? From Eq. (10) we have that:

$$R_{i,j} = G^+_{i,i} + G^+_{j,j} - 2G^+_{i,j} \geq 0.$$

It follows that $G^+_{i,j}$ tells us something about the resistance proximity of nodes i and j: it is high (in particular positive) when the nodes i and j are close (their resistance distance is low), it is low (in particular negative) when they are distant (their resistance distance is high). See Fig. 2 for an example.

It is worth noticing that only the diagonal of matrix G^+ is necessary to compute all closeness centrality scores. By virtue of Eq. 4, the diagonal of G^+ can be obtained by computing the diagonal of the inverse of a perturbed Laplacian matrix. The problem of computing the diagonal of the inverse of a matrix is well studied in the literature; for an up to date overview see Tang and Saad (2012). One of the most promising approaches, in our view, is the use of Gauss quadrature formulas in order to obtain lower and upper bounds on the sought diagonal entries (Golub and Meurant, 2010).

We next show that the notion of information centrality proposed by Stephenson and Zelen (1989) is exactly that of current-flow closeness defined above. To be sure, this has been already shown by Brandes and Fleischer (2005), but, we think, using a longer and harder proof. Stephenson and Zelen (1989) define the information between nodes i and j as:

$$I_{i,j} = \frac{1}{C_{i,i} + C_{j,j} - 2C_{i,j}}, \tag{17}$$

where $C = (G + J)^{-1}$, with G the Laplacian of the graph and J a matrix having all elements unity. Furthermore, they define the information centrality I_i of node i as the harmonic mean of information between i and other nodes in the network:

$$I_i = \frac{n}{\sum_j 1/I_{i,j}}. \tag{18}$$

Theorem 1. *Information centrality is the same as current-flow closeness centrality.*

Proof. We have to show that, for each node i of a graph, information centrality I_i defined by Eq. (18) equals current-flow closeness centrality c_i defined by Eq. (14). To this end, we prove that $R_{i,j} = C_{i,i} + C_{j,j} - 2C_{i,j}$, that is $I_{i,j} = 1/R_{i,j}$, from which the thesis follows by definition of information centrality and current-flow closeness centrality.

By virtue of Eq. (10) we know that $R_{i,j} = G^+_{i,i} + G^+_{j,j} - 2G^+_{i,j}$. Since, as we are going to prove below, matrices C and G^+ differ by a constant matrix, and this constant cancels in the sum $G^+_{i,i} + G^+_{j,j} - 2G^+_{i,j}$, then we have that $R_{i,j} = C_{i,i} + C_{j,j} - 2C_{i,j}$. To see that matrices C and G^+ differ by a constant, we show that $C = G^+ + 1/n^2 J$, hence $C - G^+$ is the constant matrix with elements equal to $1/n^2$. Using Eqs. (2) and (3), we have:

$$(G^+ + 1/n^2 J)(G + J) = I - 1/nJ + O + O + 1/nJ = I,$$

and hence $G^+ + 1/n^2 J = (G + J)^{-1} = C$.
□

We conclude this section by computing closeness centrality on a real network. The instance is a social network of dolphins (*Tursiops truncatus*) belonging to a community that lives in the fjord of Doubtful Sound in New Zealand. The unusual conditions of this fjord, with relatively cool water and a layer of fresh water on the surface, have limited the departure of dolphins and the arrival of new individuals in the group, facilitating a strong social relationship within the dolphin community. The network is an undirected unweighted graph containing 62 dolphins and 159 non-directional connections between pairs of dolphins. Two dolphins are joined by an edge if, during the observation period lasted from 1994 to 2001, they were spotted together more often than expected by chance. This network has been extensively studied by David Lusseau and co-authors, for instance see (Lusseau and Newman, 2004).

Fig. 4 depicts the dolphin social network where the size of the node is proportional to its current-flow closeness (graph on the left) or to its shortest-path closeness (graph on the right). The network can be broadly divided into two communities of different size, linked by a bridge of nodes. Each community is composed of a cluster of densely interconnected nodes and a periphery of more isolated nodes. As expected, both types of closeness give more importance to nodes in the core of the communities with respect to the peripheral ones. Furthermore, nodes lying in the bridge between the two communities are also central. However, there are also differences between the rankings corresponding to the two notions of closeness, as witnessed by the following basic statistics. The top-3 rankings according to the two closeness measures share only one dolphin, and the top-10 share 6 dolphins (but

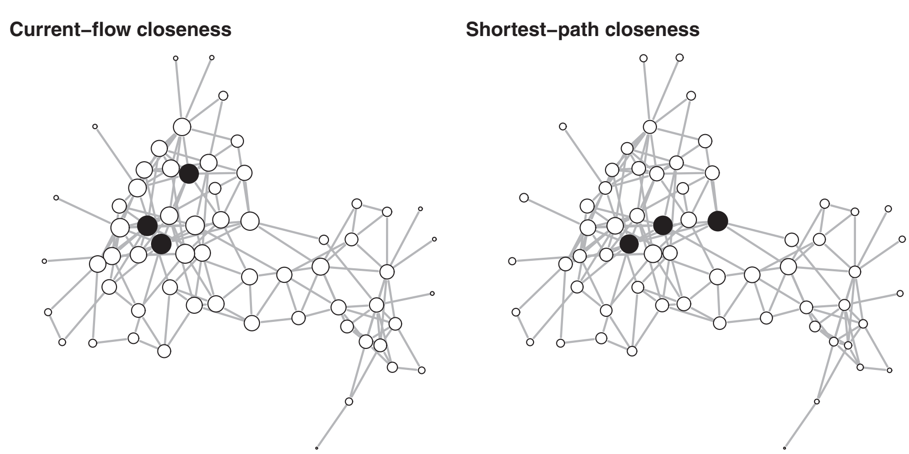


Fig. 4. A dolphin social network. The size of the nodes is proportional to the current-flow closeness (graph on the left) or to the shortest-path closeness (graph on the right) of the node. Black nodes are the top-3 leaders of the closeness rankings (only one is shared).

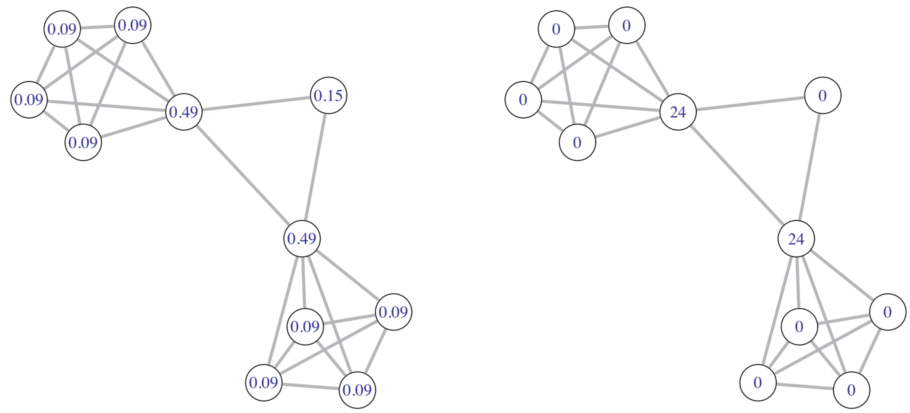


Fig. 5. A network with two clear communities linked together by a closed triad of nodes. In the graph on the left nodes are labelled with their current-flow betweenness, while in the graph on the right with their shortest-path betweenness (in this calculation, the end-points of a path are not considered part of the path).

only one with the same rank). The mean change of rank between the two entire rankings is 6.4 (10% of the ranking length), with a maximum rank change of 27. Only 4 dolphins maintain the same rank in the two compilations. The Pearson correlation coefficient between the two rankings is 0.88, and the association between current-flow closeness and node degree (0.89) is higher than the association between shortest-path closeness and node degree (0.71).

6. Current-flow betweenness centrality

Betweenness measures the extent to which a node lies on paths between other nodes. Nodes with high betweenness might have considerable influence within a network by virtue of their control over information (or over whatever else flows on the network) passing between others. They are also the ones whose removal from the network will most disrupt communications between other vertices because they lie on the largest number of paths between other nodes.

Typically, only geodesic paths are considered in the definition, obtaining a measure that is called shortest-path betweenness. Here, we study current-flow betweenness, which includes contributions of all paths, although longer paths give a lesser contribution (Newman, 2005; Brandes and Fleischer, 2005). For a given node, current-flow betweenness measures the current flow that passes through the vertex when a unit of current is injected in a source node and removed from a target node, averaged over all source–target pairs. Equivalently, it is equal to the number of times that a random walk starting at a source node and ending at a target node passes through the node on its journey, averaged over all source–target pairs.

The difference between shortest-path and current-flow betweenness is well illustrated in the example depicted in Fig. 5, which is borrowed from Newman (2005). The network depicts two communities (cliques) linked together by a closed triad of nodes. Let us call A the node of the triad that is not part of any community, and B and C the other two nodes of the triad. In the shortest-path case, the betweenness of A is clearly null. Indeed, no shortest path goes trough A, since the way through B and C is shorter. On the other hand, the current-flow betweenness of A is significant, although smaller than that of B and C, since paths that include A are longer than path that include B or C. Similarly, the shortest-path betweenness of nodes in the communities that are not in the triad is null, while their current-flow counterpart is positive.

We next give the precise definition of current-flow betweenness centrality. As observed in Section 4.1, given a source s and a target t, the absolute current flow through edge (i, j) is the quantity $A_{i,j}|v_i^{(s,t)} - v_j^{(s,t)}|$. By Kirchhoff's law the current that enters a node is equal to the current that leaves the node. Hence, the current flow $F_i^{(s,t)}$ through a node i different from the source s and a target t is half of the absolute flow on the edges incident in i:

$$F_i^{(s,t)} = \frac{1}{2}\sum_j A_{i,j}|v_i^{(s,t)} - v_j^{(s,t)}|. \tag{19}$$

Moreover, the current flows $F_s^{(s,t)}$ and $F_t^{(s,t)}$ through both s and t are set to 1, whenever end-points of a path are considered part of the path (this is our choice in the rest of this paper), or it is set to 0 otherwise. Fig. 6 gives an example. Notice that the flow from A to H through node G is 1 (all paths from A to H pass eventually through G), the flow through F is 0.4 (a proper subset of the paths from A to H go through F and these paths are generally longer than for G), and the flow through E is 0.13 (a proper subset of the paths from A to H go through E and these paths are generally longer than for F).

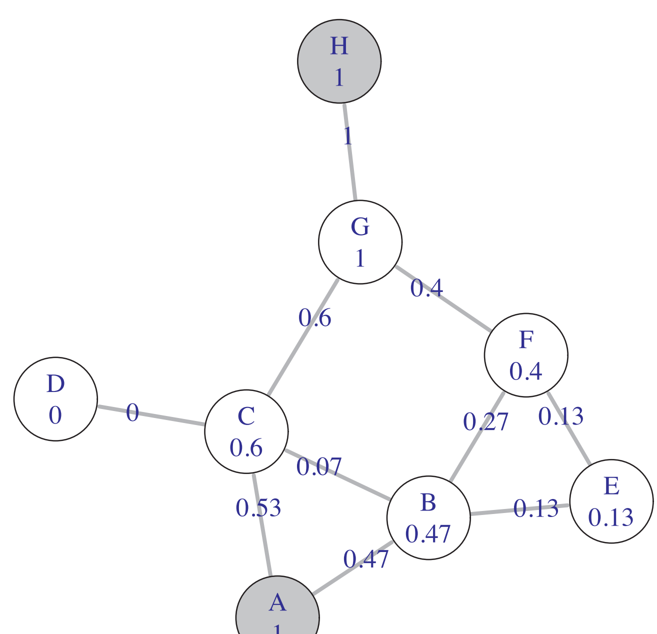


Fig. 6. A resistor network with all resistances equal to unity (this is the same network of Fig. 2). Each node is now labelled with the value of flow through it when a unit current is injected in node A and removed from node H. Each edge is labelled with the absolute current flowing on it. Values are approximated with precision of two digits.

Please cite this article in press as: Bozzo, E., Franceschet, M., Resistance distance, closeness, and betweenness. Soc. Netw. (2013), http://dx.doi.org/10.1016/j.socnet.2013.05.003

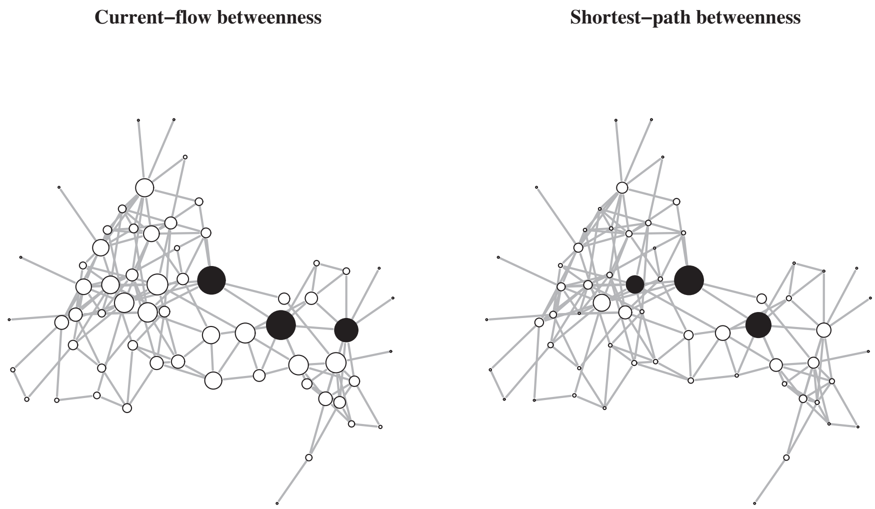


Fig. 7. A dolphin social network. The size of the nodes is proportional to the current-flow betweenness (graph on the left) or to the shortest-path betweenness (graph on the right) of the node. Black nodes are the top-3 leaders of the closeness rankings (two are shared).

Finally, the current-flow betweenness centrality b_i of node i is the flow through i averaged over all source–target pairs (s, t):

$$b_i = \frac{\sum_{s<t} F_i^{(s,t)}}{(1/2)n(n-1)}. \quad (20)$$

Since, by Eq. (8), the potential $v_i^{(s,t)} = G^+_{i,s} - G^+_{i,t}$, with G^+ the generalized inverse of the graph Laplacian, Eq. (19) can be expressed in terms of elements of G^+ as follows:

$$F_i^{(s,t)} = \frac{1}{2}\sum_j A_{i,j}|G^+_{i,s} - G^+_{j,s} + G^+_{j,t} - G^+_{i,t}|. \quad (21)$$

Notice that $F_i^{(s,t)} = F_i^{(t,s)}$. We can also rewrite Eq. (19) in terms of the resistance distance matrix R. Let us denote with e_i the i-th canonical vector such that $e_i(i) = 1$ and $e_i(j) = 0$ if $j \neq i$. Then Eq. (21) can be rewritten as:

$$F_i^{(s,t)} = \frac{1}{2}\sum_j A_{i,j}|(e_i - e_j)^T G^+(e_s - e_t)|.$$

Using Eq. (11) and the fact that $(e_i - e_j)^T e = e^T(e_s - e_t) = 0$, it follows that

$$(e_i - e_j)^T R(e_s - e_t) = -2(e_i - e_j)^T G^+(e_s - e_t),$$

and hence Eq. (19) can be expressed in terms of the resistance distance matrix as follows:

$$F_i^{(s,t)} = \frac{1}{4}\sum_j A_{i,j}|R_{i,s} - R_{j,s} + R_{j,t} - R_{i,t}|. \quad (22)$$

The above formulation gives an original interpretation of the flow through the edge (i, j) in terms of resistance distance among nodes i and j with respect to s and t. If the resistance distances of nodes i and j with respect to s (or with respect to t) are close, that is, i and j are almost equi-distant from s (or from t), then the potentials of i and j are also near, and hence there is a small potential difference between the nodes, which means that there is a small flow through the edge linking them. On the other hand, if nodes i and j have different distances from s (or from t), then their potential difference is high, which induces a large flow through the edge connecting them.

The computational complexity of current-flow betweenness centrality is as follows. Matrix G^+ can be computed by inverting a perturbed Laplacian matrix as given in Eq. (4). This costs $O(n^3)$ operations and uses $O(n^2)$ memory locations. Once we have matrix G^+, we can solve Eq. (20); this costs $O(n^2(n+m))$. This complexity can be improved to $O(mn \log n)$, as shown in Brandes and Fleischer (2005). Hence, computing current-flow betweenness centrality costs $O(n^3 + mn \log n)$ operations ($O(n^3)$ if the graph is sparse) and uses $O(n^2)$ memory locations. These costs are prohibitive if the network is relatively large. Bozzo and Franceschet (2012) devise methods for finding approximations of the generalized inverse of the graph Laplacian matrix, and hence of the resistance distance matrix, using only few eigenpairs of the Laplacian matrix. The few eigenpairs that are necessary to run the methods can be stored with a linear amount of memory in the number of nodes of the graph and, in the realistic case of sparse networks, they can be efficiently computed using one of the many methods for retrieving few eigenpairs of sparse matrices that abound in the literature. The devised approximations can be applied to estimate current-flow betweenness centrality scores when the exact computation is unfeasible.

We conclude this section by discussing Fig. 7, which depicts the already mentioned dolphin social network where the size of the node is proportional to its current-flow betweenness (graph on the left) or to its shortest-path betweenness (graph on the right). Notice that current-flow betweenness gives a high centrality to all nodes on the bridge between the two main communities of the network; on the other hand, shortest-path betweenness awards only those nodes on the bridge that belong to the shortest paths between the two communities (the lower side of the bridge). Moreover, current-flow betweenness attributes centrality to highly interconnected nodes of the two communities, since these nodes lie on many (non-shortest) paths between other nodes. The mean change of rank between the two rankings is 3.7 (6% of the ranking length), with a maximum rank change of 21. The Pearson correlation coefficient between the two rankings is 0.89, and, as observed for closeness, the association between current-flow betweenness and node degree (0.81) is higher than the association between shortest-path betweenness and node degree (0.59).

7. Conclusion

We have interpreted current-flow closeness and betweenness centrality in terms of resistance distances among nodes of the graph, which can be expressed in terms of the generalized inverse of the Laplacian matrix of a graph. This interpretation provided a simple proof that current-flow closeness and information centrality are the same measure. The computation of current-flow centralities is intensive and not feasible if the graph is large. Our interpretation allows to approximate current-flow centralities using approximations of the generalized inverse of the Laplacian (Bozzo and Franceschet, 2012). On the other hand, our investigation sheds some light on the graph-theoretic meaning of entries of the generalized inverse of the Laplacian matrix.

Current-flow centralities are relevant when information (or whatever else) flows on the network without following an optimal path. However, when information is spread on the network minimizing the length of the journey, geodetic centralities are to be preferred, also because they are less intensive to compute both in terms of time and space required.

We think that resistance distance is an interesting (and maybe still underestimated) metric on networks. It has a strong mathematical background and persuasive interpretations in both information and electrical network theory. In this contribution, we also made an attempt to revitalize resistance distance in the context of (social) network analysis. We are convinced that linking seemingly unrelated concepts is part of the research endeavor.

References

Backstrom, L., Boldi, P., Rosa, M., Ugander, J., Vigna, S., 2012. Four degrees of separation. In: ACM Web Science, pp. 45–54.

Balaban, A.T., Klein, D.J., 2002. Co-authorship, rational Erdős numbers, and resistance distances in graphs. Scientometrics 55 (1), 59–70.

Bavelas, A., 1948. A mathematical model for group structure. Human Organizations 7 (3), 16–30.

Beauchamp, A.M., 1965. An improved index of centrality. Behavioral Science 10 (2), 161–163.

Ben-Israel, A., Greville, T.N., 2003. Generalized Inverses. Theory and Applications, 2nd ed. Springer, New York.

Borgatti, S.P., 2005. Centrality and network flow. Social Networks 27 (1), 55–71.

Borgatti, S.P., Everett, M.G., 2006. A graph-theoretic perspective on centrality. Social Networks 28 (4), 466–484.

Bozzo, E., Franceschet, M., 2012. Approximations of the generalized inverse of the graph Laplacian matrix. Internet Mathematics 8 (4), 1–26.

Brandes, U., Erlebach, T. (Eds.), 2005. Network Analysis: Methodological Foundations. Vol. 3418 of Lecture Notes in Computer Science. Springer, Berlin Heidelberg.

Brandes, U., Fleischer, D., 2005. Centrality measures based on current flow. In: STACS, pp. 533–544.

Erdős, P., 1972. On the fundamental problem of mathematics. The American Mathematical Monthly 79 (2), 149–150.

Freeman, L.C., 1977. A set of measures of centrality based upon betweenness. Sociometry 40, 35–41.

Freeman, L.C., Borgatti, S.P., White, D.R., 1991. Centrality in valued graphs: a measure of betweenness based on network flow. Social Networks 13 (2), 141–154.

Ghosh, A., Boyd, S., Saberi, A., 2008. Minimizing effective resistance of a graph. SIAM Review 50 (1), 37–66.

Golub, G.H., Meurant, G., 2010. Matrices, moments and quadrature with applications. In: Princeton Series in Applied Mathematics. Princeton University Press, Princeton, Oxford.

Klein, D.J., Randić, M., 1993. Resistance distance. Journal of Mathematical Chemistry 12 (1), 81–95.

Lusseau, D., Newman, M.E.J., 2004. Identifying the role that animals play in their social networks. Proceedings of the Royal Society B: Biological Sciences 271, S477–S548.

Milgram, S., 1967. The small world problem. Physiology Today 2, 60–67.

Newman, M., 2005. A measure of betweenness centrality based on random walks. Social Networks 27 (1), 39–54.

Newman, M.E.J., 2010. Networks: An introduction. Oxford University Press, Oxford.

Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. Physical Review E 69, 026113.

Noh, J.D., Rieger, H., 2004. Random walks on complex networks. Physical Review Letters 92, 118701.

Sabidussi, G., 1966. The centrality index of a graph. Psychometrika 31, 581–603.

Shimbel, A., 1953. Structural parameters of communication networks. Bulletin of Mathematical Biology 15, 501–507.

Stephenson, K., Zelen, M., 1989. Rethinking centrality: methods and examples. Social Networks 11 (1), 1–37.

Tang, J.M., Saad, Y., 2012. A probing method for computing the diagonal of a matrix inverse. Numerical Linear Algebra with Applications 19 (3), 485–501.

Zhan, C., Chen, G., Yeung, L.F., 2010. On the distributions of Laplacian eigenvalues versus node degrees in complex networks. Physica A: Statistical Mechanics and its Applications 389 (8), 1779–1788.

Please cite this article in press as: Bozzo, E., Franceschet, M., Resistance distance, closeness, and betweenness. Soc. Netw. (2013), http://dx.doi.org/10.1016/j.socnet.2013.05.003