Ranking Vertices or Edges of a Network by Loops: A New Approach

Valery Van Kerrebroeck and Enzo Marinari
Dipartimento di Fisica, "Sapienza" Università di Roma, INFM-CNR and INFN, Piazzale Aldo Moro 2, 00185 Roma, Italy
(Received 10 April 2008; published 26 August 2008)

We introduce loop ranking, a new ranking measure based on the detection of closed paths, which can be computed in an efficient way. We analyze it with respect to several ranking measures which have been proposed in the past, and are widely used to capture the relative importance of the vertices in complex networks. We argue that loop ranking is a very appropriate measure to quantify the role of both vertices and edges in the network traffic.

DOI: 10.1103/PhysRevLett.101.098701 PACS numbers: 89.75.Hc, 02.70.−c

Finding the most important vertices is an important problem in complex network analysis [1]. In technological networks, such as the Internet, the main hubs (i.e., the vertices with many links to other vertices) play an important role in the stability of the network. Instead, the removal of any kind of species in food webs may cause the disintegration of the corresponding network [2]. The order of importance of the vertices is referred to as a *ranking*. In general the importance of a vertex, and thus its ranking, depends very much on the type of network which is under consideration.

The most simple topology based ranking measure is the degree, i.e., the number of neighboring vertices of a vertex. While the degree is easy to compute, it is not a very refined measure, as it solely depends on the local neighborhood around a vertex. Several more global measures [3–8] have been proposed which do take the overall structure of the network into account. Apart from finding a fast way to compute these measures, another challenge is usually to decide which of these measures are more appropriate for which type of network [9].

We start our reasoning from the observation that the link structure of complex networks describes the topology of interactions taking place in a dynamic complex system: the importance of a vertex (or edge) to the network traffic is related to the number of paths or walks it lies on [10].

The topology of most real-world networks is represented by a directed graph $G(V, E)$ which is completely characterized by its set V of N vertices i and by the set E of M directed edges $i \to j$, where i and j are said to be the starting and ending vertex of the directed edge $i \to j$. For undirected graphs either one of the ending vertices i and j of the edge $\{i, j\}$ can be the ending or starting vertices. We can associate a weight $r_{i\to j}$ to each one of the directed edges $i \to j$. For unweighted graphs all edge weights are uniformly equal to one. We denote the set of edges or the set of vertices of these edges distinct from i (as it is always clear from the context) of which i is an ending vertex or starting vertex as ∂_i^+ or ∂_i^-, respectively. Correspondingly, the in-degree d_i^+ and out-degree d_i^- of a vertex i are defined as the sum of the weights of the edges of which that vertex i is an ending vertex, or starting vertex, respectively.

Walks are defined as sequences of vertices $(i_1, \ldots, i_L)$, where for each couple of subsequent vertices i_{k-1} and i_k, for $k = 2, \ldots, L$, the directed edge $i_{k-1} \to i_k$ belongs to E. As such, they can cross the same edge or vertex infinitely many times. Instead, paths are defined as self-avoiding walks. In particular, a cycle or loop is a closed path. More formally, it is defined as a sequence $(i_1, i_2 \ldots, i_L, i_1)$ of vertices, where for all $k = 1, 2, \ldots, L$ these i_k are distinct from each other, and, for all $k = 2, \ldots, L$, each couple of subsequent vertices i_{k-1} and i_k are connected by a directed edge $i_{k-1} \to i_k$ belonging to E, as does $i_L \to i_1$. The weight of any of these subgraphs is defined as the product of the weights $r_{i\to j}$ of the edges composing that subgraph. In case of a cycle defined by a sequence $(i_1, i_2, \ldots, i_L, i_1)$, its corresponding weight is given by $w[C] = r_{i_L\to i_1} \prod_{k=2}^{L} r_{i_{k-1}\to i_k}$.

One measure which is widely used to find the most relevant vertices of a network is PageRank [5]. PageRank is an iteratively computed ranking measure where the PageRank of a given vertex depends on the PageRank of its neighboring vertices. More formally, the PageRank $\mathcal{P}(i)$ of a vertex i is defined as

$$\mathcal{P}(i) = c \sum_{j\in\partial_i^+} \frac{\mathcal{P}(j)}{d_j^-} + \frac{1-c}{N}, \tag{1}$$

where c is a damping factor chosen in the interval $]0, 1]$. In matrix form this becomes $\mathcal{P} = cC^T\mathcal{P} + \frac{1-c}{N}\delta$, where $\mathcal{P}$ is a N-dimensional vector, δ is a N-dimensional vector with all elements equal to one, and the elements C_{ij} of the $N \times N$ matrix C are equal to $1/d_i^-$ if the edge $i \to j$ belongs to E, and zero otherwise. As the sum of the entries of a column of this matrix C is equal to one, it can be interpreted as a Markov matrix. The resulting PageRank is then proportional to the probability with which a random walker will come across a given vertex. As such, PageRank is an importance measure for vertices which, being based on random walks, takes the overall structure of the network into account. However, the question naturally arises whether it is not preferable to emulate the behavior of a more efficient *self-avoiding* random walker.

Motivated by the latter observation, we introduce a new ranking measure based on paths rather than walks. Several centrality measures based on either the number or length of shortest paths passing through or ending at a given vertex have already been proposed [7]. In particular, betweenness centrality $\mathcal{B}$ (BC) of a given vertex (or edge) is the fraction of shortest paths on which that vertex (or edge) lies. Defining $\sigma_{k,l}$ as the number of shortest paths between k and l, and $\sigma_{k,l}(i)$ and $\sigma_{k,l}(i \to j)$ the number of these passing through i or $i \to j$, we have

$$\mathcal{B}(i) = \sum_{k,l(\neq i)\in V} \frac{\sigma_{k,l}(i)}{\sigma_{k,l}}, \quad \mathcal{B}(i \to j) = \sum_{k,l\in V} \frac{\sigma_{k,l}(i \to j)}{\sigma_{k,l}}, \quad (2)$$

for the BC of vertex i and the edge $i \to j$, respectively. One fundamental problem that prevents these measures from becoming widely used in real network analysis is that they cannot be computed as fast as, for example, PageRank [7]. Moreover, a measure based on shortest paths only may not be adequate enough as also longer paths (with possibly higher weights) could add to the centrality of a given vertex, or edge [10].

We propose a ranking based on the presence of closed paths, through a given vertex. We consider the probability of presence of cycles [11–14], rather than all, i.e., also open, paths, for a specific reason; it allows us to compute ranking by means of belief propagation (BP), a distributed, message passing algorithm which converges in linear time in the system size to the marginal probabilities of presence of these cycles. This restriction to closed paths results in a ranking which reflects the geometric position of each one of the vertices, and gives a subjective view of how each vertex sees the overall network based on paths.

In particular, we propose a ranking based on the sums of the weights of these cycles. As such it represents the probability with which a self-avoiding walker returns to the same vertex while exploring the network, taking the weight of each path into account. We define the loop ranking $\mathcal{L}$ of a vertex i, or edge $i \to j$, as the sum of the weights w of all cycles C passing through that vertex i, or edge $i \to j$ respectively, i.e.,

$$\mathcal{L}(i) = \sum_{C \ni i} w[C] \quad \text{and} \quad \mathcal{L}(i \to j) = \sum_{C \ni (i \to j)} w[C]. \quad (3)$$

On random graphs loop ranking is connected to ranking by degrees, since vertices with the highest value of $(d_i^- d_i^+)$ live on many loops [15]: the most relevant information given by loop ranking is about nonrandom correlations that can be present in real networks. In practice, we do not compute the actual loop ranking, but rather the marginal expressing the probability with which a cycle passes through a given vertex, i.e., $\mathcal{L}(i)/\sum_C w[C]$, which produces the same ordering. The latter can be obtained by reformulating the problem of identifying all cycles of a given graph as a constraint satisfaction problem [12–14].

We define an appropriate phase space in which subgraphs such as cycles are represented by a unique configuration. We associate with each edge $(i \to j)$ an Ising-like variable $S_{i\to j}$, where $S_{i\to j}$ takes on the value zero or one if the corresponding edge $(i \to j)$ belongs, or does not belong to the considered subgraph, respectively. In this way, we establish the desired one-to-one correspondence between any simple subgraph of the original graph G and all configurations defined by any of 2^M sequences $\underline{S} = (S_1, \ldots, S_M)$. For simplicity, we also introduce the notation $\underline{S}_i$ which denotes the set of all edge variables $S_{i\to j}$ and $S_{j\to i}$ of which i is a starting or ending vertex.

We can now define the probability law

$$\mathrm{Prob}[\underline{S}] = \frac{1}{Z} \prod_{(i\to j)\in E} (r_{i\to j})^{S_{i\to j}} \prod_{i\in V} f_i(\underline{S}_i), \quad (4)$$

where the local constraints $f_i(\underline{S}_i)$ are equal to 1 if $\sum_{j\in\partial_i^+} S_{ij} = \sum_{j\in\partial_i^-} S_{ij}$ is 0 or 1 and are equal to 0 otherwise, and Z is a normalization constant. The complete set of local constraints f_i ensures that only configurations representing subgraphs composed of (possibly vertex disjoint) directed cycles have a nonzero probability (4). The first product makes the probability proportional to the weight of the subgraph they represent.

An approximation to the marginals of (4) can be obtained using a local Monte Carlo algorithm [13]. However, for factorizable probability laws, such as (4), they can also easily be computed by means of message passing algorithms, such as belief and survey propagation [16]. BP is a distributed, iterative algorithm which is intrinsically linear in the system size. It requires the introduction of $2M$ real-valued message variables, M of type $x_{i\to j}[S_{i\to j}]$ in the same direction of the edges, and M of type $y_{j\to i}[S_{i\to j}]$ in the opposite direction. Initially, they all take on a random value in the interval [0, 1]. Each BP iteration consists in an update of these $2M$ variables. Assuming (4), the update rules are

$$x_{i\to j}[S_{i\to j}] = \frac{\sum_{k\in\partial_i^+} x_{k\to i}[S_{k\to i}]}{1 + \sum_{k\in\partial_i^+} x_{k\to i}[S_{k\to i}] \sum_{k'\in\partial_i^- \setminus j} y_{k'\to i}[S_{i\to k'}]},$$

$$y_{j\to i}[S_{i\to j}] = \frac{\sum_{k\in\partial_j^-} y_{k\to j}[S_{j\to k}]}{1 + \sum_{k\in\partial_j^+\setminus i} x_{k\to j}[S_{k\to j}] \sum_{k'\in\partial_j^-} y_{k'\to j}[S_{j\to k'}]}.$$

On acyclic graphs, the successive repetition of these BP iteration steps always leads to a fixed point solution. For generic graphs containing cycles, BP does not necessarily converge [17]. However, at least for sparse graphs, which do not contain too many small loops and locally are tree-like, usually it does (for recent results see [18]). Thus we expect the BP algorithm to give reasonable results for typical directed real-world networks, where the number of loops is smaller than in the random case [15].

Once the fixed point has been reached, the values of the message variables can be used to obtain the marginal probabilities. We are interested in the vertex and edge marginals expressing the probability with which a cycle contains that particular vertex or edge. Upon convergence of BP, these marginals can be obtained as

$$p_i = \frac{\sum_{k\in\partial_i^+} x_{k\to i}\sum_{k'\in\partial_i^-} y_{k'\to i}}{1+\sum_{k\in\partial_i^+} x_{k\to i}\sum_{k'\in\partial_i^-} y_{k'\to i}},$$
$$\text{and}\quad p_{i\to j} = \frac{x_{i\to j}y_{j\to i}}{1+x_{i\to j}y_{j\to i}},$$

respectively. On a generic, cyclic graph, the above expressions are an approximation to the actual marginals. However, in general, these approximations are very reasonable to work with. The marginals of (4) express the probability with which a vertex, or edge, is part of a subgraph composed of possibly several vertex disjoint directed cycles. Thus, exact marginals of (4) are possibly an overestimation of the desired marginals expressing the probability of presence of a (single) closed path. However, for weighted graphs where the edge weights have been rescaled such that they all lie in the interval [0, 1], this effect is largely reduced [14].

To get hints about which is a good ranking measure for dynamic networks, we have looked at a number of examples and particular cases. We discuss here explicitly the case of two directed small world networks that we consider very telling. We assume them to be unweighted for simplicity. We first consider the graph shown in Fig. 1. If we only take the outer ring of edges into account, all information can be exchanged between any two vertices in two ways, either in clockwise or counterclockwise direction: in such a case all vertices and edges are considered to be equally important.

The presence of the directed "short-cut" edges (which are all outbound from vertex 0) reduces the length of the shortest paths, a feature which is typical of most real-world networks. The resulting "smaller world" (from the point of view of a single vertex) has a large impact on the mobility inside a network. In particular, the vertex 0 will play a more crucial role than other vertices in dispatching packages along the network. Similarly, though in a minor way, the ending vertices of the extra outbound short cuts, i.e., 3, 5, and 8, should play a more important role in the traffic along this network. Also, we expect the presence of these edges

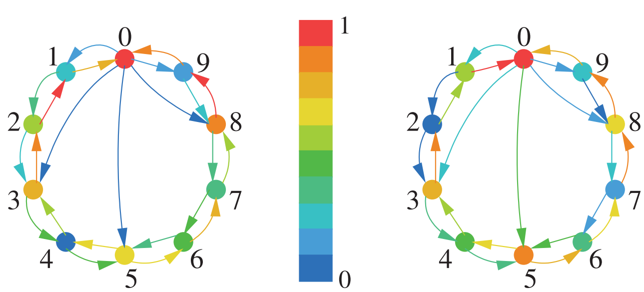

FIG. 1 (color). Directed small world network G_1. The various colors express the ranking of the vertices and edges based on loop ranking (on the left) and betweenness centrality (on the right).

to break the symmetry of the role of the single edges in the network flow.

The results according to the various ranking schemes and corresponding ordering for the vertices of graph G_1 are reported in Table I. We have rescaled the PageRank, loop ranking, and BC results such that they all lie in the interval [0, 1]. For these rather small graphs which contain a number of small loops BP was always found to converge. Clearly, the different ranking schemes attribute various degrees of importance to the vertex 0. As the PageRank of a vertex depends primarily on the number of edges directed towards that vertex and the rescaled PageRank they transmit (rescaled by the out-degree of their respective starting vertices), vertex 0 of graph G_1 has a relative low PageRank. Instead, as it does lie on most of the (closed) paths of graph G_1, its central role is acknowledged by both the loop ranking as the BC. Moreover, the latter two rankings recognize the increased role of the vertices 3, 5, and 8 with respect to the other vertices, while PageRank makes no clearcut difference between them. For the other vertices, the ordering does differ depending on which ranking is considered, as loop ranking depends on all closed paths, while the BC is only based on the number of shortest paths.

The ordering produced by loop ranking and the BC is schematically presented by Fig. 1. As in case of the vertices, loop ranking and the BC do not produce the exact same ordering of edges, but there are no essential huge shifts between the two corresponding rankings. Note that edges with high ranking usually connect one vertex with high and another with low ranking. The presence of these edges has been observed in the case of protein networks [19], where it was argued that they play a crucial role in the overall robustness of the network.

Graph G_2, shown in Fig. 2, represents a slightly different type of small world network than G_1. The extra shortcut edges are in this case pointing towards vertex 0, increasing its role in dispatching information along the network. We

TABLE I. Ranking of the vertices of the small world graph G_1 and G_2 shown in Figs. 1 and 2, respectively, according to the various ranking measures. The corresponding values of their PageRank, loop ranking, and betweenness centrality, rescaled such that they lie in the interval [0, 1] [e.g., $(\mathcal{L}(i) - \mathcal{L}_{\min})/(\mathcal{L}_{\max} - \mathcal{L}_{\min})$], are also included.

$\mathcal{P}_{G_1}(i)$		$\mathcal{L}_{G_1}(i)$		$\mathcal{B}_{G_1}(i)$		$\mathcal{P}_{G_2}(i)$		$\mathcal{L}_{G_2}(i)$		$\mathcal{B}_{G_2}(i)$	
5	1.00	0	1.00	0	1.00	0	1.00	0	1.00	0	1.00
4	0.88	8	0.57	5	0.47	1	0.62	8	0.57	5	0.45
3	0.83	3	0.47	3	0.31	9	0.53	3	0.47	3	0.28
6	0.81	5	0.41	8	0.26	2	0.31	5	0.41	8	0.25
7	0.68	2	0.30	1	0.17	8	0.31	2	0.30	1	0.20
8	0.58	6	0.14	4	0.10	3	0.17	6	0.14	4	0.10
2	0.43	7	0.07	6	0.10	7	0.09	7	0.07	6	0.09
0	0.08	1	0.05	9	0.06	5	0.05	1	0.05	9	0.06
9	0.07	9	0.03	7	0.05	6	0.04	9	0.03	7	0.03
1	0.00	4	0.00	2	0.00	4	0.00	4	0.00	2	0.00

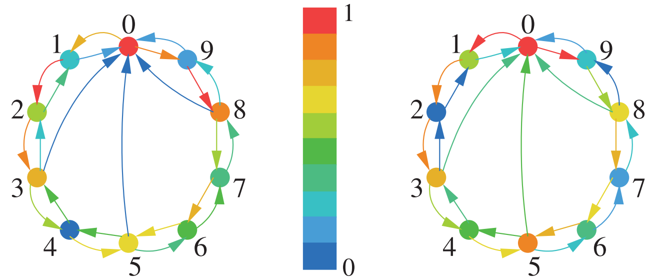

FIG. 2 (color). Directed small world network G_2; loop ranking (left) and betweenness centrality (right).

also expect the starting vertices 3, 5, 8 of the shortcut edges to play an increased role.

Our results are also in Table I. All rankings recognize the vertex 0 as the more important one. Loop ranking and BC also capture the special status of the vertices 3, 5, and 8, even if in a different order. This is not the case for PageRank, due to the fact that it does not consider the out-degree to be very relevant. A comparison of the loop ranking of the two graphs shows that the in- and outbound edges of the vertices are treated in an identical way, resulting in the same set of loop ranking values for all vertices. This is not the case for the BC, as it does not rely on all, but only on the shortest paths.

Figure 2 also includes a schematic representation of the edge and vertex ranking according to loop ranking and the BC of graph G_2. In the case of graph G_2, the edges leading away from vertex 0 play a more significant role, while in graph G_1 the edges going to vertex 0 are more important for obvious reasons.

We have also compared loop ranking to subgraph centrality (SC), the quantity introduced in [8]: here nodes are characterized by their participation in all subgraphs in the network. SC shares some features with the other approaches (for example the vertex 0 is always the more important), but has some different attitudes (for example recognizing an important role for vertex 9, that is very low in all other rankings: this is reasonable given the definition of SC) that differ from the other approaches.

In conclusion, loop ranking reflects the role of vertices and edges during the dissipation of information along the network. We have discussed unweighted small world networks to allow for an easier comparison between the various ranking methods. Note that loop ranking can naturally be extended to weighted networks, which is a clear advantage in the analysis of real-world networks.

While PageRank is more sensitive to the in-degree of a given vertex, loop ranking treats all paths (passing in either direction through a given vertex or edge) in an equivalent way. It produces a slightly different ordering of importance with respect to the BC as it takes all paths (with their relative weight) into account. Another advantage of the path based rankings we considered here is that they produce both results for the vertices as the edges. Moreover, loop ranking can be computed by means of BP, which has a linear time complexity in the system size, and this is a remarkable practical advantage. This should also allow for an easier dynamical analysis of rankings, an aspect which has already been studied more carefully for the BC [20].

A limitation of loop ranking is that it is only based on loops. As such, it can only produce results regarding the vertices and edges belonging to the 2-core of the graph, i.e., the subgraph for which all vertices have an in- and out-degree of at least one. It is reasonable to assume that the 2-core contains those vertices and edges which are important to the traffic flow: integration with different schemes could eventually allow to design ranking methods optimized for different applications.

We thank EVERGROW, IP 1935 in the complex systems initiative, EU VI FP, and D. Bickson, F. Ricci-Tersenghi, and G. Semerjian for discussions.

[1] R. Albert, H. Jeong, and A.-L. Barabasi, Nature (London) **406**, 378 (2000).
[2] J. Camacho, R. Guimerá, and L. A. Nunes Amaral, Phys. Rev. Lett. **88**, 228102 (2002).
[3] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, in *Proceedings of the 12th International World Wide Web Conference* (ACM Press, New York, 2003), p. 640.
[4] J. M. Kleinberg, J. ACM **46**, 604 (1999).
[5] L. Page, S. Brin, R. Motwani and T. Winograd, *Stanford Digital Library Technologies Project* (Stanford University, Palo Alto, CA, 1998).
[6] D. Bickson, D. Malkhi, and L. Zhou, in *Proceedings of the 7th IEEE International Conference on Peer-to-Peer Computing* (IEEE Computer Society, Washington, DC, 2007), p. 211.
[7] U. Brandes, J. Math. Sociol. **25**, 163 (2001).
[8] E. Estrada and J. A. Rodriguez-Velasquez, Phys. Rev. E **71**, 056103 (2005).
[9] M. Jungsbluth, B. Burghardt, and A. K. Hartmann, Physica (Amsterdam) **381A**, 444 (2007).
[10] A. Fekete, G. Vattay, and L.Kocarev, Phys. Rev. E **73**, 046102 (2006).
[11] G. Bianconi and M. Marsili, J. Stat. Mech. (2005) P06005.
[12] E. Marinari and G. Semerjian, J. Stat. Mech. (2006) P06019.
[13] E. Marinari, G. Semerjian, and V. VanKerrebroeck, Phys. Rev. E **75**, 066708 (2007).
[14] E. Marinari and V. Van Kerrebroeck, J. Phys. Conf. Ser. **95**, 012014 (2008).
[15] G. Bianconi, N. Gulbahce, and A. E. Motter, Phys. Rev. Lett. **100**, 118701 (2008).
[16] M. Mézard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002).
[17] J. S. Yedidia, W. T. Freeman, and Y. Weiss, Adv. Neural Inf. Process. Syst. **13**, 689 (2001).
[18] A. Dembo and A. Montanari, Physica (Amsterdam) **40E**, 1530 (2008).
[19] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
[20] C. Demetrescu, S. Emiliozzi, and G. F. Italiano, J. ACM **51**, 968 (2004).