A Bag-of-Paths Node Criticality Measure

Bertrand Lebichot[1] & Marco Saerens

Machine Learning Group – ICTEAM & LSM, Université catholique de Louvain
Place des Doyens 1, B-1348 Louvain-la-Neuve, Belgium

Abstract

This work compares several node (and network) criticality measures quantifying to which extend each node is critical with respect to the communication flow between nodes of the network, and introduces a new measure based on the Bag-of-Paths (BoP) framework. Network disconnection simulation experiments show that the new BoP measure outperforms all the other tested measures on a sample of Erdős-Rényi and Albert-Barabási graphs. Furthermore, a faster (but still $O(n^3)$), approximate, BoP criticality relying on the Sherman-Morrison rank-one update of a matrix is introduced for tackling larger networks. This approximate measure shows similar performances as the original, exact, one.

Keywords: Criticality measure, network vulnerability, vital nodes, graph mining, network science, network data analysis, betweenness centrality.

1. Introduction

The analysis and the modeling of network data has become a popular research topic in the last decade and is now often referred to as *link analysis* (in computer science) and *network science* (in physics). Network data appear in virtually every field of science and is therefore studied in many different disciplines, such as social sciences, applied mathematics, physics, computer science, chemistry, biology, economics, etc. Within this context, one important question that is often addressed is the following: Which node seems to be the most critical, or vital, in the network? The present work introduces such a new *node criticality measure*, also called *vulnerability*, quantifying to

[1]Corresponding author. Tel.: +32 10 47 83 77.
E-mail address: bertrand.lebichot@uclouvain.be,
URL: http://www.isys.ucl.ac.be/staff/lebichot/

which extend the deletion of each node hurts the connectivity within the network in a broad sense, e.g., in terms of communication, proximity, or movement. Criticality measures are often considered as a subset of centrality measures, which are frequently used as a proxy for quantifying criticality. Interested readers are invited to consult the recent comprehensive review [1].

Indeed, a huge number of centrality measures have been defined in various fields, starting from social science (see, e.g., [2, 3, 4, 5, 6, 7] and [8] for a survey). These quantities assign a score to each node of the graph G which reflects the extent to which this node is "central" by exploiting the structure of the graph G, or with respect to the communication flow between nodes. Centrality measures tend to answer the following questions [9]: What is the most representative, or central, node within a given graph (closeness centrality)? How critical is a given node with respect to the information flow in a network (criticality)? Which node is the most peripheral in a social network (eccentricity)? Which node is the most important intermediary in the network (betweenness centrality)? Centrality scores try to answer to these questions by proposing measures modeling and quantifying these different, somewhat vague, properties of the nodes.

Notice that, in general, these centrality measures are computed on undirected graphs, or, when dealing with a directed graph, by ignoring the direction of edges. They are therefore denoted as "undirectional" [10]. Measures defined on directed graphs – and therefore directional – are often called *importance* or *prestige* measures. They capture to which extend a node is "important", "prominent", or "prestigious" with respect to the whole directed graph by considering directed edges as representing some kind of endorsement. However, this kind of measure will not be discussed here.

This manuscript introduces a new, efficient and effective, criticality measure: the bag-of-paths (BoP) criticality. The quantity relies on the bag-of-paths framework assigning a Gibbs-Boltzmann distribution on the set of paths in the network [11, 12, 13]. This framework already allowed to define new distance measures between nodes interpolating between two well-known distances, the shortest-path distance and the resistance distance (or commute-time distance) [11]. In this context, the BoP criticality of a node measures the impact of the node deletion on the total accessibility between nodes within the network. More specifically, it is defined as the Kullback-Leibler divergence between the bag-of-paths probabilities, quantifying relative accessibilities, computed before and after removal of a node of interest. The larger this decrease in accessibility, the higher the impact of the node deletion, and thus the higher its criticality.

The novelty of the approach introduced in this paper can be under-

stood as follows. Most of the traditional criticality measures are essentially based on two different paradigms about the communication occurring in the network: optimal communication based on shortest paths and random communication based on a random walk on the graph. For instance, the Wiener index (described later in this paper) is based on shortest paths and the Kirchhoff index on random walks. However, both the shortest path and the random walk have some drawbacks [14]: shortest paths do not integrate the amount of connectivity between the two nodes whereas random walks quickly loose the notion of proximity to the initial node when the graph becomes larger [15, 16].

Contrary to traditional measures, our criticality measure integrates both proximity and amount of connectivity in the bag-of-paths framework [11]. Nodes that are both close and highly connected are qualified as highly *accessible*. Our introduced bag-of-paths measures aim to quantify the accessibility between the nodes. When the temperature of the model is low (close to zero), communication occurs through a random walk, while for large temperatures, short paths are promoted.

The introduced measure is compared experimentally to already developed criticality measures as well as a sample of popular centrality measures, briefly reviewed in this paper. All those measures are compared through a Kendall's correlation analysis and a *disconnection methodology* [17, 18] in Section 5. This empirical analysis is performed on a large number, and two types, of randomly generated graphs (see Subsection 5.1).

In summary, this work has the following main contributions,

- A new criticality measure, showing good performance in the identification of the most critical nodes of a network, is introduced.
- All methods are compared experimentally using two disconnection strategies on a large number of randomly generated graphs.

Finally, the paper is organized as follows: First, the underlying background and various notations are discussed in Section 2, then Section 3 introduces ten centrality and criticality measures (some being quite well-known). The bag-of-paths (BoP) model described in [11] is summarized and the new BoP criticality measure is derived in Section 4. Finally, those measures are assessed and compared in Section 5.

2. Background and Notation

This section aims to introduce the necessary background and notation used in this paper. Consider a weighted undirected graph or network, $G =$

$\{\mathcal{V}, \mathcal{E}\}$, strongly connected with a set of n nodes $\mathcal{V}$ (or vertices) and a set of edges $\mathcal{E}$ (or arcs, links). The $n \times n$ symmetric adjacency matrix of the graph, containing non-negative affinities between nodes, is denoted as $\mathbf{A}$, with elements $[\mathbf{A}]_{ij} = a_{ij} \geq 0$.

$\mathbf{A}^{\mathrm{T}}$ will refer to the transpose of $\mathbf{A}$, $\mathbf{A}^{(-j)}$ is a $(n-1) \times (n-1)$ matrix obtained from $\mathbf{A}$ by removing its jth row and its jth column, $\mathbf{e}$ is a column vector full of ones and $\mathbf{e}_j$ is the jth column vector of the identity matrix $\mathbf{I}$. Except explicitly stated, all lower-case bold letters represent column vectors while upper-case bold letters are matrices.

Moreover, to each edge between node i and j is associated a non-negative number $c_{ij} \geq 0$. This number represents the immediate cost of transition from node i to j. If there is no link between i and j, the cost is assumed to take a large value, denoted by $c_{ij} = \infty$. The cost matrix $\mathbf{C}$ is an $n \times n$ matrix containing the c_{ij} as elements. Costs are usually set independently of the adjacency matrix: they are quantifying the cost of a transition according to the problem at hand. For example, costs can be set in function of some properties, or features, of the nodes (or the edges) in order to bias the probability distribution of choosing a path to follow. In the case of a social network, we may, for instance, want to bias the paths in function of the education level of the persons, therefore favoring paths visiting highly educated persons. Now, if there is no reason to introduce a cost, we can simply set $c_{ij} = 1$ (paths are penalized by their length) or $c_{ij} = 1/a_{ij}$ (in this case, a_{ij} is viewed as a conductance and c_{ij} as a resistance) – this last setting will be used in the experimental section.

We also introduce the Laplacian matrix $\mathbf{L}$ of the graph, defined in the usual manner and needed below,

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \tag{1}$$

where $\mathbf{D} = \mathbf{Diag}(\mathbf{Ae})$ is the diagonal (out)degree matrix of the graph G containing the $a_{i\bullet}$ on its diagonal. One interesting property of $\mathbf{L}$ is that its eigenvalues provide important information about the connectivity of the graph [19].

One of the most interesting accessibility measure of the graph G, the so-called connectivity, is often defined as the minimum number of nodes that need to be removed to separate it into two disconnected sub-graphs [20, 21]. Unfortunately, this quantity is hard to compute and cannot be easily exploited in practice for this reason. Beside this, it can be shown that the number of zero eigenvalues of $\mathbf{L}$ is equal to the number of disconnected subgraphs, or connected components, of G [19]. Then, for a connected graph

the smallest eigenvalue of $\mathbf{L}$ is called the algebraic connectivity or spectral gap and has been shown to be a good indicator of its overall "connectedness" (G is disconnected when its algebraic connectivity is equal to zero). Finally, the Moore-Penrose pseudoinverse of $\mathbf{L}$ is denoted as $\mathbf{L}^+$, and contains elements l_{ij}^+. Due to the properties of the Moore-Penrose pseudoinverse, its largest eigenvalue is the algebraic connectivity.

In addition, a *natural random walk* on G is defined in the standard way. In node i, the random walker chooses the next edge to follow according to reference transition probabilities

$$p_{ij}^{\text{ref}} = \frac{a_{ij}}{\sum_{j'=1}^{n} a_{ij'}} \tag{2}$$

The $n \times n$ matrix $\mathbf{P}^{\text{ref}}$, containing transition probabilities p_{ij}^{ref}, is stochastic and is simply equal to $\mathbf{P}^{\text{ref}} = \mathbf{D}^{-1}\mathbf{A}$. Note that this can lead to a division by zero if a node i is isolated or is a dangling node; we therefore assume that the graph is strongly connected. $\mathbf{P}^{\text{ref}}$ represents the probability of jumping from any node i to node $j \in \mathcal{S}ucc(i)$, the set of successor nodes of i. In other words, the random walker chooses to follow an edge with a likelihood proportional to the affinity (apart from the sum-to-one normalization), therefore favoring edges with a large associated affinity.

A path $\wp$ (also called a walk) is a sequence of transitions to adjacent nodes on G (loops are allowed), initiated from a starting node s, and stopping in an ending node e. The total cost of a path $\wp$, $\tilde{c}(\wp)$, is defined as the sum of the individual transition costs c_{ij} along $\wp$.

3. Related Work

In this paper, a large set of criticality measures will be compared experimentally, and briefly reviewed in this section (see [8, 14] for a more thorough description of these measures). It is convenient to categorize them into three classes: node betweenness centrality measures, global graph criticality measures, and node criticality measures.

3.1. Node betweenness centralities

As already mentioned, the concept of criticality is closely related to the concept of betweenness centrality; we therefore also investigate a few of the most well-known betweenness and centrality measures. The measure is defined on each node, identified by its index j.

- The simple *node degree*, or *edge connection* (EC). This quantity is simply the number of nodes connected to a node j, weighted by edge weights in the case of a weighted graph. It is obtained by summing the entries on the jth row of the adjacency matrix $\mathbf{A}$. The idea is that if a node has a high degree, it is more likely to hurt or disconnect the graph when removed. It can be computed by

$$\mathrm{EC}_j = \mathbf{e}_j^{\mathrm{T}} \mathbf{A} \mathbf{e} \tag{3}$$

- The famous *shortest path betweenness* (SPB), introduced by Freeman [2]. It counts the proportion of shortest paths connecting any two nodes i and k, and passing through an intermediate node j of interest (with $i \neq j \neq k \neq i$). The idea is that if a node contributes to a large number of shortest paths, it can be considered as an important intermediary between nodes when the information is spread "optimally" along shortest paths. More precisely,

$$\mathrm{SPB}_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} \frac{\eta(j \in \mathcal{P}_{ik}^*)}{|\mathcal{P}_{ik}^*|} \tag{4}$$

  where $\mathcal{P}_{ik}^*$ is the set of all shortest paths from i to k, $|\mathcal{P}_{ik}^*|$ is the total number of such shortest paths $\wp_{ik}^*$ and $\eta(j \in \mathcal{P}_{ik}^*) = \sum_{\wp_{ik}^* \in \mathcal{P}_{ik}^*} \delta(j \in \wp_{ik}^*)$ is the total number of such paths visiting node j. We used Brandes' algorithm [22] to compute the SPB of each node of the graph.

- The *random walk betweenness* (RWB), introduced by Newman [3] and closely related to Brandes' electrical centrality [23]. Newman introduced the current flow betweenness centrality, which measures the centrality of a node as the total sum of electrical currents that flow through it, when considering all node pairs as source-destination pairs with a unit current flow. The current flow betweenness is also called the *random walk betweenness centrality* because of the well-known connection between electric current flows and random walks [24, 14]. The idea is thus the same as for the SPB, but taking into account a random walk-based diffusion of information instead of shortest paths. Notice that Brandes and Fleischer [23] proposed a more efficient algorithm computing the random walk betweenness for all nodes of a network. The properties and computation of the current flow betweenness have also been discussed by Bozzo and Franceschet [25]. Kivimaki et al. proposed a new betweenness measure interpolating between the shortest path betweenness and the random walk betweenness [26].

- *Estrada's centrality* (EST). In [4], Estrada et al. defined a centrality measure called "subgraph centrality" for a weighted undirected graph or subgraph. It summarizes simply as

$$\mathrm{EST}_j = \mathbf{e}_j^{\mathrm{T}} \left(\sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!} \right) \mathbf{e}_j = \mathbf{e}_j^{\mathrm{T}} \, \mathbf{diag}(\mathrm{expm}(\mathbf{A})) \tag{5}$$

  where $\mathrm{expm}(\mathbf{A})$ is the matrix exponential of $\mathbf{A}$ and $\mathbf{diag}(\mathbf{X})$ extract the main diagonal of $\mathbf{X}$. It is well-known that element $a_{ij}^{(k)} = [\mathbf{A}^k]_{ij}$ of matrix $\mathbf{A}^k$ ($\mathbf{A}$ to the power k) is the weighted number of paths between node i and node j with exactly k steps (length k). The subgraph centrality measure therefore integrates a contribution from all paths connecting node j to himself, discounting paths according to their number of steps (it favors shorter paths in terms of length). The intuition is that a node should have a high centrality score if the closed paths (cycles) starting from it are short and are visiting many different nodes [4].

3.2. Node criticalities

We now introduce the node criticalities studied in this work. As for the betweenness, the criticality measure is defined on each node j.

- *Wehmuth's criticality K* (WK) is introduced in [5],

$$\mathrm{WK}_j = \frac{\lambda_2^{(j)}}{\log_2(d_j)} \tag{6}$$

  where $\lambda_2^{(j)}$ is the algebraic connectivity of the h-neighbourhood of node j (the subnetwork composed by all nodes within h hops, or steps, of node j) and d_j is the degree of node j. Recall that the algebraic connectivity is the second smallest eigenvalue of the Laplacian matrix $\mathbf{L}$. The idea is to take advantage of the algebraic connectivity property; the higher the value of $\lambda_2^{(j)}$, the higher the connectivity/density of the subnetwork. Then, $\lambda_2^{(j)}$ is divided by the logarithm of the node degree as locally computed algebraic connectivities show a bias towards higher values on nodes with high degree. This bias causes $\lambda_2^{(j)}$ to be over-sensitive to the presence of hubs [5].

- *Klein's edge criticality* (KLE). Klein derived the analytical form of this node criticality measure for several global measures, including the

Wiener index and the Kirchhoff index [6]. We will use the measure based on the Kirchhoff index here ([6]; see also [14]),

$$\mathrm{KLE}_j = \sum_{i=1}^{n} a_{ij}(\mathbf{e}_i - \mathbf{e}_j)^{\mathrm{T}}(\mathbf{L}^{+})^2(\mathbf{e}_i - \mathbf{e}_j) \tag{7}$$

The intuition behind the measure is the following. Klein's edge (i, j) criticality is defined as the sensitivity of the global network criticality index (here the Kirchhoff index – defined in the next subsection) with respect to the increase in the resistance of the edge (i, j) [6]. In other words, it quantifies the impact of an increase in this resistance on the global network. Edges having a high impact on the global network criticality hurt most the network and are considered as highly critical. Then, edge criticality is summed up over incident edges to provide a node criticality.

3.3. Global network criticalities

The following *global* criticality indexes are defined on the whole network G. They quantify the extend to which the network as a whole is efficient, that is, highly interconnected and cohesive, with high accessibility. For a communication network, this measure can be, e.g., the "Wiener index" – the sum of the shortest-path distances (which can be travel time, travel cost, etc.) between all pairs of nodes. An effective network is characterized by a *low value* of the Wiener index as, then, distances between nodes are small in average.

The impact of a node of interest on the global network accessibility measure – the *derived node criticality* – is then quantified by evaluating the marginal loss in global accessibility when the node of interest is not operating, i.e., has simply been removed. This measure therefore reports how critical the node is, relative to the entire graph. To evaluate the criticality of a particular node j in a fixed graph G, the difference between the global criticality after deleting this node j, $\mathrm{cr}(G \setminus j)$, and the initial global network criticality, $\mathrm{cr}(G)$, is computed,

$$\mathrm{cr}_j = \mathrm{cr}(G \setminus j) - \mathrm{cr}(G) \tag{8}$$

and the higher this value, the more critical node j is. Here, $G \setminus j$ is graph G whose node j and incident edges have been removed.

This node criticality will be computed on several well-known global criticality measures which are described now. We could also normalize the

quantity when it corresponds to a sum over all pairs of nodes by something like $\mathrm{cr}(G \setminus i)/((n-1)(n-2)) - \mathrm{cr}(G)/(n(n-1))$. However, this does not change the ranking of the nodes as the second term is a constant.

- The *Wiener index* (WIE) is defined as the sum of the shortest-path distances between all node pairs (see, e.g., [8]),

$$\mathrm{WIE}(G) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \Delta_{ij}^{\mathrm{SP}} \tag{9}$$

  where $\Delta_{ij}^{\mathrm{SP}}$ is the shortest-path distance. The underlying idea is that if the sum of the distances between every node pairs is small, the network is more likely to be well-connected.

- The *Kirchhoff index* (KIR) is similar to the Wiener index but uses the resistance distance (the effective resistance, proportional to the commute-time distance based on a random walk on the graph) [27], instead of the shortest path distance, and has been recently used by Tizghadam and al. in network theory for quantifying the robustness of a communication network [7]. It can be easily computed by

$$\mathrm{KIR}(G) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \Delta_{ij}^{\mathrm{ER}} \tag{10}$$

  where $\Delta_{ij}^{\mathrm{ER}}$ is now the effective resistance between i and j, with $\Delta_{ii}^{\mathrm{ER}} = 0$ for each node i. The idea is thus the same as for WIE, but with a different concept of distance.

- The *Kemeny index* (KEM) represents the expected number of steps needed by a random walker for reaching an arbitrary node from some arbitrary starting node [28], when the starting and ending nodes are selected according to the equilibrium distribution of the Markov chain. Indeed, for an irreducible, aperiodic, Markov chain, it is known (see, e.g., [29]) that the stationary distribution exists and is independent of the initial state i. More precisely, the Kemeny index is

$$\mathrm{KEM}(G) = \sum_{i=1}^{n}\pi_i \sum_{j=1}^{n}\pi_j m_{ij} = \sum_{j=1}^{n}\pi_j m_{ij} \tag{11}$$

  where m_{ij} is the average first-passage time between node i and node j and $\boldsymbol{\pi}$ is the stationary distribution. Equation (11) holds because

it can be shown that the quantity $\sum_{j=1}^{n} \pi_j m_{ij}$ is independent of the starting node i [30]. This index measures the relative accessibility of all pairs of nodes, putting more weight on the long-term frequently visited nodes according to the stationary distribution.

- The *Shield value* (SHV) has recently been introduced [31]:

$$\mathrm{SHV}(G) = \lambda_1 \tag{12}$$

where λ_1 is the dominant eigenvalue of the adjacency matrix $\mathbf{A}$. It is closely related to the loop capacity and the path capacity of the graph, that is, the number of loops and paths of finite length. The higher λ_1, the more loops and long path in the graph. As for Estrada's centrality, the underlying idea is that if a graph has many such loops and paths then it is more likely to be well connected. The more the deletion of a node lowers this value, the less the graph becomes connected, and therefore the larger its criticality value.

4. The Proposed bag-of-paths Criticality

We now derive a new node criticality measure called the bag-of-paths criticality (BPC). It is based on computing the effect of a node removal in a bag-of-paths model (BoP). This framework was recently introduced in [11] (see also [32] for a related work) for computing distances on graphs, and used for semi-supervised classification tasks in [11, 12]. In order to make the paper as self-contained as possible, we briefly review this framework first in this section. The BoP criticality and its fast approximation are derived in the next two subsections. Finally, an illustrative example is shown in Subsection 4.4.

4.1. The bag-of-paths model

The BoP framework is based on the probability of drawing a path $i \rightsquigarrow j$ starting at a node i and ending in a node j from a virtual bag containing all possible paths in the network [11]. Let us define $\mathcal{P}_{ij}$ as the set of all paths connecting node i to node j, including loops. We further define the set of all paths through the network as $\mathcal{P} = \bigcup_{i,j=1}^{n} \mathcal{P}_{ij}$.

The potentially infinite set of paths in the graph is enumerated and a probability distribution is assigned to the set of individual paths $\mathcal{P}$, considered independently. This probability distribution on the set $\mathcal{P}$ represents the probability of drawing a path $\wp \in \mathcal{P}$ from the bag, and is defined as

the probability distribution $\mathrm{P}(\cdot)$ minimizing the total expected cost along path $\wp$, $\mathbb{E}\left[\tilde{c}(\wp)\right]$, among all the distributions having a fixed relative entropy J_0 with respect to a reference distribution, for instance the natural random walk on the graph (defined by Equation (2)). The quantity $\tilde{c}(\wp)$ is the cumulated cost along path $\wp$.

This choice naturally defines a probability distribution on the set of paths such that "long" (high cost) paths occur with a low probability while "short" (low cost) paths occur with a high probability. In other words, we are seeking path probabilities, $\mathrm{P}(\wp), \wp \in \mathcal{P}$, minimizing the total expected cost subject to a constant relative entropy constraint,

$$\left| \begin{array}{ll} \underset{\{\mathrm{P}(\wp)\}}{\text{minimize}} & \displaystyle\sum_{\wp \in \mathcal{P}} \mathrm{P}(\wp)\, \tilde{c}(\wp) \\ \text{subject to} & \sum_{\wp \in \mathcal{P}} \mathrm{P}(\wp) \ln(\mathrm{P}(\wp)/\mathrm{P}^{\mathrm{ref}}(\wp)) = J_0 \\ & \sum_{\wp \in \mathcal{P}} \mathrm{P}(\wp) = 1 \end{array} \right. \tag{13}$$

where $\mathrm{P}^{\mathrm{ref}}(\wp)$ represents the probability of following the path $\wp$ when walking according to the natural random walk reference distribution (see Equation (2)). More precisely, $\mathrm{P}^{\mathrm{ref}}(\wp)$ is proportional to $\tilde{\pi}^{\mathrm{ref}}(\wp)$, which is the product of the transition probabilities p_{ij}^{ref} along the path $\wp$ – the *likelihood* of the path (see [11] for details). Here, $J_0 > 0$ is provided a priori by the user, according to the desired degree of randomness, or exploration, he is willing to concede. Note also that, normally, a non-negativity constraint should be added, but this is not necessary since the resulting probabilities will automatically be non-negative.

As well-known (see [33, 34] and [11, 32, 35] for maximum entropy distributions over paths), this problem is similar to a standard maximum entropy one and can be solved by introducing the following Lagrange function integrating equality constraints

$$\mathscr{L} = \sum_{\wp \in \mathcal{P}} \mathrm{P}(\wp)\tilde{c}(\wp) + \lambda \left[\sum_{\wp \in \mathcal{P}} \mathrm{P}(\wp) \ln \left(\frac{\mathrm{P}(\wp)}{\mathrm{P}^{\mathrm{ref}}(\wp)} \right) - J_0 \right] + \mu \left[\sum_{\wp \in \mathcal{P}} \mathrm{P}(\wp) - 1 \right]$$

and optimizing over the set of path probabilities $\{\mathrm{P}(\wp)\}_{\wp \in \mathcal{P}}$ (partial derivatives set to zero). The Lagrange parameters are then deduced after imposing the constraints.

The result of the minimization of (13) is a Gibbs-Boltzmann probability

distribution [11, 32, 35]:

$$\mathrm{P}(\wp) = \frac{\tilde{\pi}^{\mathrm{ref}}(\wp)\exp\left[-\theta\tilde{c}(\wp)\right]}{\sum_{\wp'\in\mathcal{P}}\tilde{\pi}^{\mathrm{ref}}(\wp')\exp[-\theta\tilde{c}(\wp')]} \tag{14}$$

where $\theta = 1/T$ plays the role of an inverse temperature, exp is the elementwise exponential and $\tilde{\pi}(\wp)$ is the likelihood of the path $\wp$, according to the natural random walk on G (the reference random walk) defined earlier in this section.

As expected, short paths $\wp$ (having a low $\tilde{c}(\wp)$) are favoured in that they have a larger probability of being chosen. Moreover, from Equation (14), we clearly observe that when $\theta \to 0^+$, paths probabilities reduce to the probabilities generated by the natural random walk on the graph. In this case, $J_0 \to 0$ and paths are chosen according to their likelihood in a natural random walk. On the other hand, when θ is large, the probability distribution defined by Equation (14) is biased towards short paths (shortest ones are more likely). Notice that, in the sequel, it will be assumed that the user provides the value of the parameter θ instead of J_0, with $\theta > 0$.

The *bag-of-paths probability* [11], $\mathrm{P}(s = i, e = j)$, is an important quantity defined on the set of (starting, ending) nodes of the paths. It corresponds to the probability of drawing a path starting in node i and ending in node j from the virtual bag-of-paths:

$$\mathrm{P}(s = i, e = j) = \frac{\sum_{\wp\in\mathcal{P}_{ij}}\tilde{\pi}^{\mathrm{ref}}(\wp)\exp[-\theta\tilde{c}(\wp)]}{\sum_{\wp'\in\mathcal{P}}\tilde{\pi}^{\mathrm{ref}}(\wp')\exp[-\theta\tilde{c}(\wp')]} \tag{15}$$

where $\mathcal{P}_{ij}$ is the set of paths connecting the starting node i to the ending node j.

In [11], it is shown that this bag-of-paths probability can be computed in matrix form by

$$\mathrm{P}(s = i, e = j) = \frac{z_{ij}}{\sum_{i',j'=1}^{n} z_{i'j'}}, \text{ with } \mathbf{Z} = (\mathbf{I} - \mathbf{W})^{-1} \tag{16}$$

where z_{ij} is the element i, j of matrix $\mathbf{Z}$, called the *fundamental matrix* and

$$\mathbf{W} = \mathbf{P}^{\mathrm{ref}} \circ \exp[-\theta\mathbf{C}] \tag{17}$$

with $\circ$ being the elementwise (Hadamard) product.

Notice that $\mathrm{P}(s = i, e = j)$ is not symmetric. These probabilities quantify the *relative accessibility* between the nodes and it was shown that minus their logarithm, $-\log \mathrm{P}(s = i, e = j)$, defines a useful distance measure between nodes [11]. By construction this probability is high when the two nodes i and j are *highly connected* (there are many terms in the numerator of Equation (15)) by *low-cost* paths (each term of the numerator is large). In other words, it accurately captures the intuitive notion of relative accessibility. These BoP probabilities will serve as a basis for defining the BoP criticality.

Note that the BoP probabilities can also be used to define some betweenness measures [26] which are related to well-known centrality/betweenness measures in some sense: if $\theta \to \infty$ the betweenness tends to be highly correlated with Freeman's betweenness [2] (only shortest paths are considered), while if $\theta \to 0^+$, the betweenness tends to be highly correlated with Newman's random walk betweenness [3] described in Subsection 3.1 (based on a random walk on G).

4.2. The bag-of-paths criticality: basic, standard, case (BPC)

We now derive a closed-form formula for computing these probabilities when an intermediate node j is deleted from the graph. Then, our BoP criticality measure for node j will be the relative entropy (or Kullback-Leibler divergence) between the bag-of-paths probabilities – the relative accessibility (see Equation (16)) – before and after removing node j from G. It therefore quantifies to which extend the relative accessibility is affected by the deletion of node j.

The intuition is the following. The bag-of-paths criticality measures the global impact of a node deletion on the total relative accessibility of the nodes in the network

- by computing this accessibility *before* and *after* node deletion,
- and then by computing their difference by means of the Kullback-Leibler divergence.
- This difference computes the *loss in accessibility* when deleting each node in turn.

Thus, in this work, a critical node is defined as a node whose deletion greatly affects the relative accessibility between the remaining nodes. This criticality measure will be referred as BPC. We now detail its derivation.

4.2.1. Reducing the support of the bag-of-paths probability distribution

First, let us introduce some new notation. In Equation (16), z_{ik} will be denoted as $z_{ik}(\mathbf{A})$ and $\mathbf{Z}$ as $\mathbf{Z}(\mathbf{A})$ since they are based on adjacency matrix $\mathbf{A}$. Then, as our criticality measure relies on the deletion of a node (say, node j), we will need to reduce the support of the bag-of-paths probability distribution to $\mathcal{V} \setminus j$ (the set of nodes of G, with node j removed) by eliminating paths starting or ending in j.

To do this, we introduce $\mathbf{Z}^{(-j)}(\mathbf{A})$, which is $\mathbf{Z}$ based on $\mathbf{A}$ (the original graph), but where the jth column and the jth row of $\mathbf{Z}$ *have been removed*. Then, $z_{ik}^{(-j)}(\mathbf{A})$ with $i \neq j$ and $k \neq j$ is its i, k element. We further define $\mathrm{P}_{ik}^{(-j)}(\mathbf{A}) = \mathrm{P}^{(-j)}(s = i, e = k)$ with support $\mathcal{V} \setminus j$ based on the elements of $\mathbf{Z}^{(-j)}(\mathbf{A})$,

$$\mathrm{P}_{ik}^{(-j)}(\mathbf{A}) = \frac{z_{ik}^{(-j)}(\mathbf{A})}{\sum_{\substack{i',k'=1 \\ i',k' \neq j}}^{n} z_{i'k'}^{(-j)}(\mathbf{A})}, \text{ with } i, k \neq j \tag{18}$$

which corresponds to the BoP probabilities (see Equation (16)) based on the whole original graph ($\mathbf{A}$), but where the *support of the discrete probability distribution is reduced to the set of nodes different from j* – we do not consider node j as a potential source or destination node.

In practice, from this last equation, we observe that this can be done by putting both row j and column j of $\mathbf{Z}$ to 0 and then summing over its elements, as will be done in Algorithm 1 (line 7).

4.2.2. Computing $\mathbf{Z}$ after deleting node j from $\mathbf{A}$

We now turn to the computation of the fundamental matrix of the graph after deleting node j. In this context, it is important not to confuse $\mathbf{Z}^{(-j)}(\mathbf{A})$ (introduced in the previous subsection) with $\mathbf{Z}(\mathbf{A}^{(-j)})$, which is defined as matrix $\mathbf{Z}$ computed from Equation (16), but based this time on $\mathbf{A}^{(-j)}$: the adjacency matrix $\mathbf{A}$ whose jth row and column have been removed (node j is deleted so that paths in the graph cannot visit this node any more). In other words, $\mathbf{A}^{(-j)}$ is the adjacency matrix of $G \setminus j$. Thus the $z_{ik}(\mathbf{A}^{(-j)})$ with $i \neq j$ and $k \neq j$ are the elements of $\mathbf{Z}(\mathbf{A}^{(-j)})$. Notice that $\mathbf{Z}^{(-j)}(\mathbf{A})$ and $\mathbf{Z}(\mathbf{A}^{(-j)})$ have the same size; both are $(n-1) \times (n-1)$ square matrices (node j is dismissed in both cases).

Then, from Equation (16), we define the bag-of-paths probabilities $\mathrm{P}(s =$

$i, e = k|s, e \neq j)$ based on $\mathbf{A}^{(-j)}$ as

$$\mathrm{P}_{ik}(\mathbf{A}^{(-j)}) = \frac{z_{ik}(\mathbf{A}^{(-j)})}{\sum\limits_{\substack{i',k'=1 \\ i',k' \neq j}}^{n} z_{i'k'}(\mathbf{A}^{(-j)})}, \text{ with } i, k \neq j \tag{19}$$

corresponding to the graph G with node j removed.

4.2.3. The bag-of-paths criticality

Finally, the **bag-of-paths criticality** (BPC) is the *Kullback-Leibler divergence* between the bag-of-paths probabilities, quantifying relative accessibilities, before and after node removal,

$$\mathrm{cr}_j = \sum_{\substack{i,k=1 \\ i,k \neq j}}^{n} \mathrm{P}_{ik}^{(-j)}(\mathbf{A}) \log\left(\frac{\mathrm{P}_{ik}^{(-j)}(\mathbf{A})}{\mathrm{P}_{ik}(\mathbf{A}^{(-j)})}\right) \tag{20}$$

and the larger this divergence, the larger the impact of the deletion of node j on the overall accessibility.

Note that computing the bag-of-paths criticality for all the n nodes has a time complexity of about $O(n^3 + n(n-1)^3)$. The first term corresponds to the evaluation of $\mathrm{P}^{(-j)}(\mathbf{A})$ (which requires a matrix inversion) and the second term to n evaluations of $\mathrm{P}(\mathbf{A}^{(-j)})$ (inversion of n matrices, after deleting each node). This leads to an overall $O(n^4)$ time complexity. We now turn to a fast approximation of this quantity.

4.3. The bag-of-paths criticality: a fast approximation (BPCf)

In this subsection, we will modify the bag-of-paths criticality to obtain a $O(n^3)$ time complexity instead of $O(n^4)$. It relies on the efficient approximation of the entries of $\mathbf{Z}^{(-j)}$ in terms of the fundamental matrix $\mathbf{Z} = (\mathbf{I} - \mathbf{W})^{-1}$. This version will be referred as BPCf.

4.3.1. The fast, approximate, bag-of-paths criticality

Let us first define

- $\mathbf{z}_j^{\mathrm{c}} = \mathbf{col}_j(\mathbf{Z}) = \mathbf{Z}\mathbf{e}_j$ and $\mathbf{z}_j^{\mathrm{r}} = \mathbf{row}_j(\mathbf{Z}) = \mathbf{e}_j^{\mathrm{T}}\mathbf{Z}$
- $\mathbf{w}_j^{\mathrm{c}} = \mathbf{col}_j(\mathbf{W}) = \mathbf{W}\mathbf{e}_j$ and $\mathbf{w}_j^{\mathrm{r}} = \mathbf{row}_j(\mathbf{W}) = \mathbf{e}_j^{\mathrm{T}}\mathbf{W}$

where $\mathbf{col}_j$ and $\mathbf{row}_j$ are respectively the jth column (a column vector) and the jth row (a row vector) of the matrix.

The main idea behind the approximation is to set row j of matrix $\mathbf{W}$ to zero[2] (providing $\mathbf{W}^{(-j)}$), instead of deleting row and column j of adjacency matrix $\mathbf{A}$, as required by the exact bag-of-paths criticality (see Equation (20)). Indeed, this approximation appears to be much simpler that the original problem and reduces the set of paths to paths avoiding j (as if node j was deleted), as shown below. Then, the bag-of-paths criticality is approximated by the Kullback-Leibler divergence between the bag-of-paths probabilities, as before,

$$\mathrm{cr}_j^{\mathrm{f}} = \sum_{\substack{i,k=1 \\ i,k\neq j}}^{n} \mathrm{P}_{ik}^{(-j)}(\mathbf{A}) \log\left(\frac{\mathrm{P}_{ik}^{(-j)}(\mathbf{A})}{\mathrm{P}_{ik}(\mathbf{W}^{(-j)})}\right) \tag{21}$$

using this time $\mathbf{W}^{(-j)}$ for computing the fundamental matrix and the bag-of-paths probabilities (instead of $\mathbf{A}^{(-j)}$ in Equation (20)). However, this only results in an *approximation* of the exact solution, as discussed later in Subsection 4.3.4. We now detail how to approximate efficiently the bag-of-paths probabilities from the matrix $\mathbf{W}^{(-j)}$.

4.3.2. Computing $\mathbf{Z}$ *after setting row* j *of* $\mathbf{W}$ *to zero*

Indeed, turning node j into a killing, absorbing, node (no outgoing link from this node) can be achieved by defining a new matrix $\mathbf{W}^{(-j)} = \mathbf{W} - \mathbf{e}_j \mathbf{w}_j^{\mathrm{r}}$ as $\mathbf{W}$ is the elementwise (Hadamard) product between $\mathbf{P}^{\mathrm{ref}}$ and $\mathbf{C}$ (see Equation (17)). Doing so, row j of $\mathbf{W}$ is set to zero, meaning that node j cannot be an intermediate node anymore, as if node j was deleted. Thus paths connecting i and k (with $i, k \neq j$) cannot visit j any more: this node is excluded from the paths. Moreover, this actually corresponds to a simple rank-one matrix update.

By exploiting this property, it will now be shown that we obtain an extremely simple formula for the update of the fundamental matrix:

$$\mathbf{Z}(\mathbf{W}^{(-j)}) = (\mathbf{I} - \mathbf{W}^{(-j)})^{-1} = \mathbf{Z} - \frac{\mathbf{z}_j^{\mathrm{c}} \mathbf{z}_j^{\mathrm{r}}}{z_{jj}} \tag{22}$$

[2]Note that we obtain the same result if we set both row j and column j of $\mathbf{W}$ to zero – this way of doing is equivalent to deleting row and column j of $\mathbf{W}$ and computing the fundamental matrix and the bag-of-paths probabilities from this reduced matrix. However, setting only row j to zero is simpler and provides the same results.

where only the entries $i, k \neq j$ of $\mathbf{Z}(\mathbf{W}^{(-j)})$ are meaningful. Recall that $\mathbf{z}_j^{\mathrm{c}}$ is a column vector while $\mathbf{z}_j^{\mathrm{r}}$ is a row vector. The rest of the subsection is dedicated to the derivation of this result and can be skipped at first reading.

Indeed, this results from a simple application of the Sherman-Morrison formula (see, e.g., [36, 37, 38]) for the inverse of a rank-one update of a matrix: if $\mathbf{c}$ and $\mathbf{d}$ are column vectors,

$$(\mathbf{B} + \mathbf{c}\mathbf{d}^{\mathrm{T}})^{-1} = \mathbf{B}^{-1} - \frac{\mathbf{B}^{-1}\mathbf{c}\mathbf{d}^{\mathrm{T}}\mathbf{B}^{-1}}{1 + \mathbf{d}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{c}} \tag{23}$$

Now, from $\mathbf{W}^{(-j)} = \mathbf{W} - \mathbf{e}_j\mathbf{w}_j^{\mathrm{r}}$, we have $(\mathbf{I} - \mathbf{W}^{(-j)}) = (\mathbf{I} - \mathbf{W}) + \mathbf{e}_j\mathbf{w}_j^{\mathrm{r}}$. By setting $\mathbf{B}^{-1} = \mathbf{Z}$, $\mathbf{B} = (\mathbf{I} - \mathbf{W})$, $\mathbf{c} = \mathbf{e}_j$ and $\mathbf{d} = (\mathbf{w}_j^{\mathrm{r}})^{\mathrm{T}}$ in Equation (23), we obtain for (22)

$$\mathbf{Z}(\mathbf{W}^{(-j)}) = (\mathbf{I} - \mathbf{W}^{(-j)})^{-1} = \mathbf{Z} - \frac{\mathbf{Z}\mathbf{e}_j\mathbf{w}_j^{\mathrm{r}}\mathbf{Z}}{1 + \mathbf{w}_j^{\mathrm{r}}\mathbf{Z}\mathbf{e}_j} \tag{24}$$

Let us first compute the term $\mathbf{w}_j^{\mathrm{r}}\mathbf{Z}$ appearing both in the numerator and the denominator of the previous equation. Since $\mathbf{Z} = (\mathbf{I}-\mathbf{W})^{-1}$, $(\mathbf{I}-\mathbf{W})\mathbf{Z} = \mathbf{I}$, and thus

$$\begin{aligned}
\mathbf{w}_j^{\mathrm{r}}\mathbf{Z} &= ((\mathbf{w}_j^{\mathrm{r}})^{\mathrm{T}} - \mathbf{e}_j + \mathbf{e}_j)^{\mathrm{T}}\mathbf{Z} \\
&= -(\mathbf{e}_j - (\mathbf{w}_j^{\mathrm{r}})^{\mathrm{T}})^{\mathrm{T}}\mathbf{Z} + \mathbf{e}_j^{\mathrm{T}}\mathbf{Z} \\
&= -\mathbf{e}_j^{\mathrm{T}} + \mathbf{z}_j^{\mathrm{r}} = \mathbf{z}_j^{\mathrm{r}} - \mathbf{e}_j^{\mathrm{T}}
\end{aligned} \tag{25}$$

From Equation (25), the denominator of the second term in the right-hand side of Equation (24) becomes

$$1 + \mathbf{w}_j^{\mathrm{r}}\mathbf{Z}\mathbf{e}_j = 1 + \left(\mathbf{z}_j^{\mathrm{r}} - \mathbf{e}_j^{\mathrm{T}}\right)\mathbf{e}_j = \mathbf{z}_j^{\mathrm{r}}\mathbf{e}_j = z_{jj} \tag{26}$$

Moreover, also from (25), the numerator of the second term in the right-hand side of Equation (24) is

$$\mathbf{Z}\mathbf{e}_j\mathbf{w}_j^{\mathrm{r}}\mathbf{Z} = \mathbf{z}_j^{\mathrm{c}}\left(\mathbf{z}_j^{\mathrm{r}} - \mathbf{e}_j^{\mathrm{T}}\right) \tag{27}$$

We substitute the results (26) and (27) in the denominator and the numerator of Equation (24), providing

$$\mathbf{Z}(\mathbf{W}^{(-j)}) = \mathbf{Z} - \frac{\mathbf{z}_j^{\mathrm{c}}\left(\mathbf{z}_j^{\mathrm{r}} - \mathbf{e}_j^{\mathrm{T}}\right)}{z_{jj}} \tag{28}$$

However, row and column j should neither be taken into account, nor used, and can therefore be put to zero. Indeed, since the last term of the numerator in Equation (28), $\mathbf{z}_j^{\mathrm{c}}\mathbf{e}_j^{\mathrm{T}}$, only updates the jth column, it can safely be ignored (this column j is useless and will never be used, as it corresponds to the deleted node), resulting in redefining the quantity as

$$\mathbf{Z}(\mathbf{W}^{(-j)}) = (\mathbf{I} - \mathbf{W}^{(-j)})^{-1} = \mathbf{Z} - \frac{\mathbf{z}_j^{\mathrm{c}}\mathbf{z}_j^{\mathrm{r}}}{z_{jj}}$$

and now the jth row as well as the jth column of $\mathbf{Z}(\mathbf{W}^{(-j)})$ are equal to zero. Indeed, elementwise, this last equation reads $z_{ik}(\mathbf{W}^{(-j)}) = z_{ik} - z_{ij}z_{jk}/z_{jj}$, which is equal to zero both when $i = j$ and $k = j$. We therefore obtain exactly Equation (22). Thus, the fundamental matrix $\mathbf{Z}$ needs to be inverted *only once* and the elements $z_{ik}(\mathbf{A}^{(-j)})$ in Equation (19) are approximated by $z_{ik}(\mathbf{W}^{(-j)})$ for computing the approximate bag-of-paths probabilities.

The resulting matrix has a jth row as well as a jth column equal to zero and it can be shown that each element $z_{ik}(\mathbf{W}^{(-j)})$ of $\mathbf{Z}(\mathbf{W}^{(-j)})$ corresponds to

$$z_{ik}(\mathbf{W}^{(-j)}) = \sum_{\wp \in \mathcal{P}_{ik}^{(-j)}} \tilde{\pi}^{\mathrm{ref}}(\wp) \exp\left[-\theta c(\wp)\right] \tag{29}$$

where $\mathcal{P}_{ik}^{(-j)}$ is the set of paths avoiding node j.

4.3.3. The approximate bag-of-paths probabilities

The approximate bag-of-paths probabilities are computed from $\mathbf{W}^{(-j)}$ in the same way as for the standard bag-of-paths (see Equation (19)),

$$\mathrm{P}_{ik}(\mathbf{W}^{(-j)}) = \frac{z_{ik}(\mathbf{W}^{(-j)})}{\sum_{\substack{i',k'=1 \\ i',k' \neq j}}^{n} z_{i'k'}(\mathbf{W}^{(-j)})}, \text{ with } i,k \neq j \tag{30}$$

where the elements of the fundamental matrix are computed from Equation (22) this time.

Finally, the fast approximation of the criticality measure is computed from these approximate bag-of-paths probabilities through Equation (21). The algorithm is detailed in Algorithm 1, where the probabilities $\mathrm{P}_{ik}^{(-j)}(\mathbf{A})$ and $\mathrm{P}_{ik}(\mathbf{W}^{(-j)})$ are respectively gathered in matrices $\mathbf{\Pi}$ and $\mathbf{\Pi}^{(-j)}$.

4.3.4. Discussion of the approximation

It should be noted that this procedure only computes an *approximation* of the BoP probabilities $\mathrm{P}_{ik}(\mathbf{A}^{(-j)})$ (defined in Equation (19)) when removing an intermediate node j. Indeed, for computing the exact probabilities on the graph $G \setminus j$, the natural random walk transition probabilities (the reference probability matrix $\mathbf{P}^{\mathrm{ref}}$) should also be updated, as the edges entering node j cannot be followed any more. In our approximate procedure, these reference probabilities are *not updated* when computing $\mathbf{W}^{(-j)}$ (see Equation (17)), causing some (usually small) disturbance in comparison with explicitly deleting the node j and recomputing the quantities (including transition probabilities) from this new graph $G \setminus j$. Relative performances of the exact BoP criticality and the approximated BPCf criticality will be investigated in the experiments.

Note that the expression could be adapted to exactly reflect node deletion, but the update formula becomes much more complex and we did not observe any significant difference between the two approaches in our experiments (see the experimental section).

One way to render the procedure exact would be to instead minimize expected cost subject to a fixed *entropy* constraint (as in [35]), instead of the *Kullback-Leibler* divergence in Equation (13). This results in redefining the $\mathbf{W}$ matrix as

$$\mathbf{W} = \exp[-\theta \mathbf{C}] \tag{31}$$

instead of (17). This solves the problem of the $\mathbf{P}^{\mathrm{ref}}$ update since this transition matrix does not appear any more in the computation of $\mathbf{W}$ and $\mathbf{Z}$. However, experiments showed that this choice performs slightly worse (therefore not reported in the paper) than the approximate update introduced in this section.

An elementary study of the empirical time complexity of the two versions BPC and BPCf is reported in Figure 1. Recall that the overall complexity for BPC is $O(n^4)$ and $O(n^3)$ for BPCf. For a 3000-nodes graph, the saving factor is greater than 10. Notice that no sparse, approximate, or optimized, implementation were used in the study. The CPU is a simple Intel(R) Core(TM) i5-4310 at 2.00 GHz with 8 Go RAM and the programming language is Matlab.

4.4. Illustrative example

A small toy graph, depicted on Figure 2, is now used as an illustrative example. This graph has six nodes: the (rounded) BPC value for each node is 6.3, 8.5, 5.5, 6.2, 7.1, 6.3, respectively. It corresponds to the node ranking

Algorithm 1 Computing the approximate bag-of-paths criticality of the nodes of a graph.

Input:

– A weighted undirected graph G containing n nodes.
– The $n \times n$ adjacency matrix $\mathbf{A}$ associated to G, containing affinities.
– The $n \times n$ cost matrix $\mathbf{C}$ associated to G.
– The inverse temperature parameter θ.

Output:

– The $n \times 1$ approximate bag-of-paths criticality vector $\mathbf{cr}$ containing the change in the probability distribution of picking a path starting in node i and ending in node k, when each node j is deleted in turn.

1. $\mathbf{D} \leftarrow \mathbf{Diag}(\mathbf{Ae})$ {the row-normalization matrix; $\mathbf{e}$ is a column vector full of 1s}
2. $\mathbf{P}^{\mathrm{ref}} \leftarrow \mathbf{D}^{-1}\mathbf{A}$ {the reference transition probability matrix}
3. $\mathbf{W} \leftarrow \mathbf{P}^{\mathrm{ref}} \circ \exp\left[-\theta \mathbf{C}\right]$ {elementwise exponential and multiplication $\circ$}
4. $\mathbf{Z} \leftarrow (\mathbf{I} - \mathbf{W})^{-1}$ {the fundamental matrix}
5. **for** $j = 1$ to n **do** {compute criticality for each node j in turn}
6. $\quad \mathbf{z}_j^{\mathrm{r}} \leftarrow \mathbf{e}_j^{\mathrm{T}}\mathbf{Z}$ and $\mathbf{z}_j^{\mathrm{c}} \leftarrow \mathbf{Z}\mathbf{e}_j$ {copy row j and column j of $\mathbf{Z}$}
7. $\quad \mathbf{Z}' \leftarrow \mathbf{Z} - \mathbf{e}_j\mathbf{z}_j^{\mathrm{r}} - \mathbf{z}_j^{\mathrm{c}}\mathbf{e}_j^{\mathrm{T}} + z_{jj}\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}$ {set row j and column j of $\mathbf{Z}$ to 0 for disregarding paths starting and ending in j, but keeping those passing through j. Note that the last term is introduced because the diagonal element z_{jj} is subtracted twice.}
8. $\quad \mathbf{\Pi} \leftarrow \dfrac{\mathbf{Z}'}{\mathbf{e}^{\mathrm{T}}\mathbf{Z}'\mathbf{e}}$ {normalize in order to obtain the bag-of-paths probability matrix whose support is now $\mathcal{V} \setminus j$}
9. $\quad \mathbf{Z}^{(-j)} \leftarrow \mathbf{Z} - \dfrac{\mathbf{z}_j^{\mathrm{c}}\mathbf{z}_j^{\mathrm{r}}}{z_{jj}}$ {update of matrix $\mathbf{Z}$ when removing row j from $\mathbf{W}$}
10. $\quad \mathbf{\Pi}^{(-j)} \leftarrow \dfrac{\mathbf{Z}^{(-j)}}{\mathbf{e}^{\mathrm{T}}\mathbf{Z}^{(-j)}\mathbf{e}}$ {normalize in order to obtain the corresponding bag-of-paths probabilities after deletion of row j of $\mathbf{W}$}
11. $\quad$ Remove both row j and column j from $\mathbf{\Pi}$ and $\mathbf{\Pi}^{(-j)}$
12. $\quad \boldsymbol{\pi} \leftarrow \mathbf{vec}(\mathbf{\Pi})$ and $\boldsymbol{\pi}^{(-j)} \leftarrow \mathbf{vec}(\mathbf{\Pi}^{(-j)})$ {stack probabilities into column vectors by using the vec operator}
13. $\quad \mathrm{cr}_j \leftarrow (\boldsymbol{\pi}^{(-j)})^{\mathrm{T}} \log(\boldsymbol{\pi}^{(-j)} \div \boldsymbol{\pi})$ {compute Kullback-Leibler divergence with $\div$ being the elementwise division. It is assumed that $0 \log 0 = 0$ and $0 \log(0/0) = 0$}
14. **end for**
15. **return** $\mathbf{cr}$

2, 5, 6, 1, 4, 3 (where the largest score defines the most critical node), which seems legit. Conversely, the WIE criticality succeeds to identify node 2 as the most critical, but the second node in the ranking is node 3, which looks counter-intuitive.

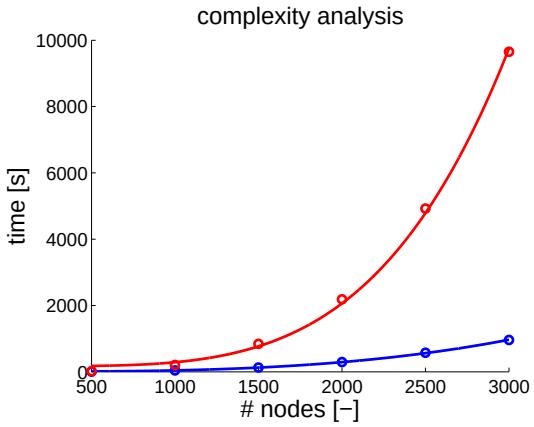

Figure 1: Empirical complexity analysis: computation time (in seconds) in function of network size (number of nodes). The overall complexity for BPC (upper curve) is $O(n^4)$ (a matrix inversion per node) and $O(n^3)$ for BPCf (lower curve, only one matrix inversion plus fast updates). We observe that BPCf scales better than BPC; for instance, for a 3000-nodes graph, the saving factor is larger than 10.

5. Experimental Comparisons

In this section, the bag-of-paths criticalities (both the exact one (BPC) and the fast approximate one (BPCf)) and the other centrality measures introduced in Section 3 are compared (see Table 1 for a reminder) on the two types of graphs described in subsection 5.1. To do so, we followed a common methodology [17, 18, 39, 40, 41] described in subsection 5.2 and we report first a simple correlation analysis between rankings in subsection 5.3. Then, results are compared and discussed in subsection 5.4.

5.1. Datasets

We used two well-known graph generators [42, 43] to build a set of 200 graphs: 100 are generated using Erdős-Rényi's model and an additional 100 using Albert-Barabási's model. Each of these models has different variants; the one we used is described below. The number of nodes is set randomly for each graph between 5 and 500.

- **Erdős-Rényi (ER) Graph Generator [43].** This model is also called the Poisson random graph generator because it generates a random graph with a Poisson node degree distribution. This type of graph is often used to study theoretical properties and behavior of networks [44]. A parameter $p \in]0, 1]$ is required. The model first generates an

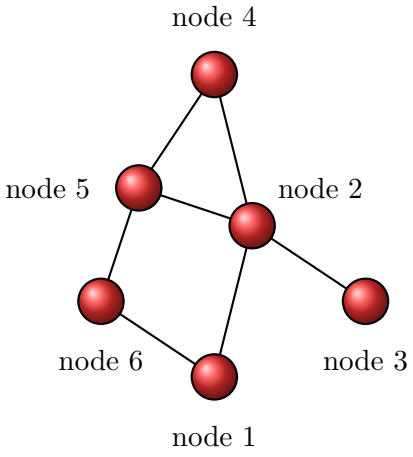

Figure 2: A small toy graph. The (rounded) BPC value for each node is 6.3, 8.5, 5.5, 6.2, 7.1, 6.3, respectively. It corresponds to node ranking 2, 5, 6, 1, 4, 3.

upper triangular random matrix (zeros on diagonal), then, for each entry of the matrix, it puts a 0 if the entry is smaller than p, and 1 otherwise. Then the matrix is symmetrized using $\mathbf{A} + \mathbf{A}^{\mathrm{T}}$. For our experiments, p was set to a random value for each graph, with $p \in]0, 1/2]$.

- **Albert-Barabási (AB) Graph Generator [42].** The model generates a random graph with a power law degree distribution. This kind of network is often observed in natural and human-generated systems, including the world wide web, citation networks, and social networks [44]. An integer parameter m is required. The model begins with an initial connected network of $m + 1$ nodes. Then, new nodes are added to the network, one at a time. Each new node is connected to m existing nodes with a probability that is proportional to the current degree of each node. The procedure stops when the desired number of nodes is reached. Heavily linked nodes ("hubs") tend to quickly accumulate even more links: the new nodes have a "preference" to attach themselves to these already heavily linked nodes. For our experiments, p was set to a random value for each graph with $m \in \{1, 2, 3, 4, 5, 6\}$. Many "natural" networks in real life behave like AB graphs (see for example [17] and citations inside).

Table 1: List of all measures compared in this study, together with their type, acronym and parameter. If a measure depends on a parameter, tested values as well as the most frequent value (mode) are reported. Notice that Shortest Path and Random Walk Betweenness algorithms are fast, optimized, versions. The other algorithms were implemented in Matlab, as described in Section 3. Further notice that the Matlab implementation of the matrix exponential is very efficient (it is used for computing Estrada's node betweenness).

Name	**Type**	**Acronym**	**Description**	**Param.**	**Tested values**	**Mode**	**Time**
Baseline (random disconnection)	-	BL	Subsection 5.2	none	-	-	$< 10^{-3}s$
Edge Connectivity	Node Betw.	EC	See Eq. 3	none	-	-	$< 10^{-3}s$
Shortest Path Betweenness	Node Betw.	SPB	See Eq. 4	none	-	-	$0.8s$
Random Walk Betweenness	Node Betw.	RWB	Subsection 3.1	none	-	-	$1s$
Estrada Index	Node Betw.	EST	See Eq. 5	none	-	-	$0.6s$
Wehmuth's K	Node Crit.	WK	See Eq. 6	h	[1,2,3,4,5,6]	1 (28%)	$342s$
Klein Index	Node Crit.	KLE	See Eq. 7	none	-	-	$1634s$
Wiener Index	Graph Crit.	WIE	See Eq. 9	none	-	-	$375s$
Kirchhoff Index	Graph Crit.	KIR	See Eq. 10	none	-	-	$884s$
Kemeny Index	Graph Crit.	KEM	See Eq. 11	none	-	-	$1000s$
Shield Value	Graph Crit.	SHV	See Eq. 12	none	-	-	$182s$
Bag-of-Paths criticality (fast)	Node Crit.	BPCf	See Eq. 20	θ	$10^{[-6,-3,-2,-1,0,1]}$	10 (44%)	$42s$
Bag-of-Paths criticality (standard)	Node Crit.	BPC	See Eq. 22	θ	$10^{[-6,-3,-2,-1,0,1]}$	1 (39%)	$205s$

5.2. Disconnection strategies

To study the performances of the different centrality/criticality measures, we simulate the effect of network attacks consisting in deleting its nodes sequentially in the order provided by the measure – the most critical nodes being deleted first. This is a natural way of assessing node criticality [17, 18]. We then record, for each network and each measure, the results of this sequential node deletion by measuring its gradual impact on network connectivity. A good criticality measure hurts most the network by, e.g., disconnecting it in several connected components, each preferably having an equal size – a balanced partition.

In practice, we first compute a criticality ranking of all nodes according to each different centrality/criticality measure introduced in the previous section. This ranking can be achieved in two different way: (1) it is computed once for all from the whole graph G (one *single ranking*), or (2) it is re-computed after each node deletion. With this last option, the centrality/criticality measures must be re-computed $n-1$ times which is time-consuming. We therefore decided to update the ranking only 100 times in total (except, obviously, for graphs with $n < 100$ nodes). This last option will be referred to as *100-ranking*.

Recall that, to evaluate the criticality of a node j with respect to a global graph criticality measure, the difference between the graph criticality of $G \backslash j$ and the global graph G criticality is computed (see Equation (8)).

Once those node rankings have been computed for each measure, the

Table 2: Results obtained with the disconnection strategies described in Subsection 5.2. The *Friedman/Nemenyi ranking* over 100 graphs (AB and ER), according to two disconnection strategies (single ranking and 100-ranking), is presented, together with the mean ± standard deviation of the obtained relative biggest connected component *area under the curve* (AUC). Concerning the ranking, the critical difference is equal to 1.82, meaning that a measure is significantly better than another if their rank difference is larger than this amount. For the ranking, the larger is the better whereas, for AUC, smaller is better. In each column, the methods in bold are the best ones or are not significantly different from the overall best one.

100 AB graphs: single ranking			100 ER graphs: single ranking			100 AB graphs: 100-ranking			100 ER graphs: 100-ranking		
measure	ranking	AUC	measure	ranking	AUC	measure	ranking	AUC	measure	ranking	AUC
BPC	**11.750**	**0.3092 ±0.163**	**BPC**	**12.355**	**0.8634 ±0.180**	**BPCf**	**11.640**	**0.3174 ±0.155**	**BPC**	**12.555**	**0.7936 ±0.161**
BPCf	**11.285**	**0.3103 ±0.164**	BPCf	10.250	0.8773 ±0.182	**BPC**	**11.370**	**0.3185 ±0.156**	**BPCf**	**11.270**	**0.8063 ±0.163**
RWB	**10.425**	**0.3158 ±0.167**	SPB	9.590	0.8851 ±0.167	**WK**	**10.375**	**0.3249 ±0.159**	RWB	9.095	0.8186 ±0.160
KIR	9.435	0.4550 ±0.255	RWB	9.365	0.8827 ±0.175	EC	8.645	0.3392 ±0.162	KIR	8.630	0.8405 ±0.138
WK	8.805	0.3246 ±0.175	KIR	8.575	0.8954 ±0.150	RWB	8.475	0.3427 ±0.162	WK	8.275	0.8215 ±0.163
SPB	8.205	0.3283 ±0.172	WK	7.550	0.8937 ±0.168	EST	8.090	0.3423 ±0.167	SPB	7.955	0.8258 ±0.156
EC	7.815	0.3276 ±0.176	EC	7.290	0.8959 ±0.165	SPB	7.510	0.3523 ±0.164	EC	7.730	0.8272 ±0.159
KLE	6.940	0.3577 ±0.208	WIE	6.610	0.9112 ±0.131	KLE	5.945	0.3740 ±0.182	KEM	6.280	0.8467 ±0.143
WIE	4.385	0.5188 ±0.242	KEM	5.665	0.9092 ±0.145	KIR	5.290	0.5232 ±0.238	WIE	5.530	0.8719 ±0.112
KEM	4.260	0.5226 ±0.246	EST	4.230	0.9113 ±0.156	KEM	5.230	0.5172 ±0.228	EST	5.410	0.8396 ±0.156
EST	3.620	0.4666 ±0.224	SHV	3.780	0.9207 ±0.129	SHV	4.005	0.4179 ±0.169	SHV	3.730	0.8640 ±0.134
SHV	2.585	0.5035 ±0.185	BL	3.015	0.9366 ±0.104	WIE	2.635	0.5995 ±0.232	KLE	2.790	0.8771 ±0.150
BL	1.490	0.7078 ±0.193	KLE	2.725	0.9273 ±0.134	BL	1.795	0.6380 ±0.194	BL	1.910	0.8986 ±0.120

simulated attacks can start. Nodes are deleted in decreasing order of criticality. After each node deletion, the *Biggest Connected Component* size (BCC), i.e., the number of nodes contained in the largest connected component, is recorded [17, 18]. The smaller this value, the more effective the attack and thus the more effective the criticality index (see Figure 3 for an example). This performance measure quantifies to which extend the network is decomposed in several balanced parts (no "giant" component is left). If, for example, the node deletion strategy (the criticality ranking) is very inefficient, and it never disconnects the network, the BCC only decreases by one unit at a time. On the contrary, if it cuts the network into two equally sized parts, the BCC is divided by two, which corresponds to a large decrease.

By further normalizing with respect to the size of the graph, that is, dividing BCC by the current number of nodes, we get the *Relative Biggest Connected Component* size (RBCC) which will be the performance indicator used in the experiments. It is then possible to draw a plot of RBCC versus the number of deleted nodes $(1, 2, 3, \ldots, n)$ [17, 18]. Then, to summarize those plots, we sum up the *Area Under the Curve* (AUC). The smaller this AUC, the better the method since the deletion of the most critical nodes (according to the ranking) quickly disconnects the network into balanced components, leading to smaller RBCC (see the illustrative example in Figure

Table 3: Another perspective on the results obtained with the disconnection strategies described in Subsection 5.2. The *Friedman/Nemenyi ranking* over 100 graphs (AB or ER) is presented, together with the mean ± standard deviation of the RBBC *area under the curve* (AUC). Here, both strategies (single ranking and 100-ranking) are analyzed together. Concerning the ranking, the critical difference is equal to 3.97, meaning that a measure is significantly better than another if their rank difference is larger than this amount. For the ranking, the larger is the better while for AUC, the smaller is the better. In each column, the methods in bold are the best ones or are not significantly different from the overall best one.

100 AB graphs			100 ER graphs		
measure	ranking	AUC	measure	ranking	AUC
1-BPC	**23.185**	0.30923 ±0.16319	**100-BPC**	**25.190**	0.79360 ±0.16064
1-BPCf	**22.575**	0.31030 ±0.16351	**100-BPCf**	**23.955**	0.80633 ±0.16343
1-RWB	**21.045**	0.31579 ±0.16704	**100-RWB**	**21.660**	0.81857 ±0.15963
100-BPCf	**20.355**	0.31740 ±0.15509	**100-WK**	20.880	0.82151 ±0.16311
100-BPC	**20.075**	0.31847 ±0.15584	**100-KIR**	20.440	0.84045 ±0.13762
1-WK	18.875	0.32461 ±0.17465	**100-SPB**	20.250	0.82587 ±0.15663
1-KIR	18.565	0.45500 ±0.25486	**100-EC**	20.200	0.82723 ±0.15883
100-WK	18.235	0.32488 ±0.15874	**100-KEM**	18.275	0.84668 ±0.14274
1-SPB	17.665	0.32831 ±0.17209	**100-EST**	17.790	0.83957 ±0.15614
1-EC	17.585	0.32764 ±0.17627	**100-WIE**	16.420	0.87185 ±0.11230
100-EC	15.525	0.33924 ±0.16238	**100-SHV**	15.375	0.86398 ±0.13363
100-RWB	15.120	0.34273 ±0.16239	**1-BPC**	14.855	0.86344 ±0.18018
100-EST	14.890	0.34227 ±0.16667	**100-KLE**	13.595	0.87705 ±0.15034
1-KLE	14.780	0.35771 ±0.20768	**1-BPCf**	12.160	0.87733 ±0.18177
100-SPB	13.395	0.35227 ±0.16436	**100-BL**	11.245	0.89857 ±0.11965
100-KLE	10.675	0.37401 ±0.18217	**1-SPB**	11.120	0.88511 ±0.16666
100-KIR	10.085	0.52323 ±0.23805	**1-RWB**	11.005	0.88265 ±0.17446
100-KEM	10.065	0.51721 ±0.22753	**1-KIR**	9.655	0.89537 ±0.15000
1-WIE	8.685	0.51884 ±0.24191	**1-WK**	8.900	0.89371 ±0.16805
1-KEM	8.470	0.52258 ±0.24621	**1-EC**	8.470	0.89593 ±0.16539
100-SHV	7.905	0.41789 ±0.16913	**1-WIE**	7.290	0.91115 ±0.13097
1-EST	7.300	0.46664 ±0.22439	**1-KEM**	6.405	0.90924 ±0.14499
100-WIE	5.600	0.59947 ±0.23205	**1-EST**	5.040	0.91128 ±0.15548
1-SHV	4.815	0.50353 ±0.18480	**1-SHV**	4.390	0.92066 ±0.12886
100-BL	3.190	0.63796 ±0.19396	**1-BL**	3.300	0.93658 ±0.10436
1-BL	2.340	0.70782 ±0.19250	**1-KLE**	3.135	0.92732 ±0.13371

3).

Finally, we report our results as follows: we perform a Friedman/Nemenyi test [45] and, in addition, we also compute the mean and the standard deviation of the AUC across all of the AB and ER generated graphs, providing more detailed results. Results can be found on Table 2; the higher the ranking, the better the criticality measure.

If a parameter is present, it is tuned as follows: for each graph, a range of values is tested and the best one is chosen for the disconnection experiment (the size of the graph can influence the parameter choice). This reflects the case of a real attack (we assume that the attacker has access to the network structure and can test the effect of the different parameters). Parameters could be tuned again after each node deletion, but it would be too computationally intensive, so we did not investigate this approach. For information, best values of parameters h and θ are reported on Table 4.

Table 4: Number of times each value of the parameters is selected during the disconnection strategies described in Subsection 5.2. Note that only WK, BPCf and BPC need a parameter tuning. Bold values show the maximum per task and per measure.

measure	parameter value	100 AB graphs: single ranking	100 ER graphs: single ranking	100 AB graphs: 100-ranking	100 ER graphs: 100-ranking	sum over the 4 tasks
WK	$h = 1$	**28**	15	**52**	14	**111**
	$h = 2$	7	**68**	6	7	87
	$h = 3$	27	15	8	23	70
	$h = 4$	27	0	9	**26**	60
	$h = 5$	8	2	13	16	40
	$h = 6$	3	0	12	14	32
BPCf	$\theta = 10^{-6}$	6	24	**29**	10	61
	$\theta = 0.001$	6	0	9	3	18
	$\theta = 0.01$	6	1	12	5	24
	$\theta = 0.1$	8	2	19	14	44
	$\theta = 1$	18	6	22	29	76
	$\theta = 10$	**56**	**67**	9	**39**	**177**
BPC	$\theta = 10^{-6}$	16	21	8	15	60
	$\theta = 0.001$	6	1	4	4	14
	$\theta = 0.01$	17	3	2	1	23
	$\theta = 0.1$	24	8	17	18	66
	$\theta = 1$	12	**52**	**49**	**45**	**156**
	$\theta = 10$	**25**	15	20	17	81

For comparison, we also consider the case where nodes are simply removed at random and independently (BL for baseline). It corresponds to a random "failure" or "attack", which has been studied theoretically in the literature (see [17] for an example).

5.3. Preliminary exploration: correlation analysis

The different centrality/criticality measures were first compared by computing two Kendall's correlation tests between each ranking. This is reported on Table 5 for both a small and a larger value of the parameters of our centrality/criticality measures: θ (BPCf and BPC) and h (WK). The small θ and h were set to 10^{-6} and 1, respectively, while the larger θ and h were 10 and 6. To summarize and to make things more visual, dendrograms were built above with a Ward hierarchical clustering (see, e.g., [46, 47, 48]) based on Kendall's correlation matrices (Figure 4).

5.4. Results and discussion

Detailed results are presented in Table 2, and Table 1 lists the different tested methods together with their acronym. Note that, when performing the Friedman/Nemenyi test comparing the different rankings provided by the methods, the *critical difference* is equal to 1.82, meaning that a measure is considered as significantly better than another if its rank is larger than this amount.

Table 5: Mean Kendall's correlation between the investigated measures over our 200 graphs. Above the main diagonal: results for the larger θ and h (10 and 6). Below the main diagonal: results for the smaller θ and h (10^{-6} and 1).

	EC	SPB	SHV	WK	BPCf	WIE	KIR	KLE	EST	BL	KEM	RWB	BPC
EC	**1.0000**	0.8784	0.7352	0.9745	0.8525	0.7088	0.7562	0.3452	0.7761	-0.0030	0.6848	0.8895	0.7637
SPB	0.8784	**1.0000**	0.6367	0.8555	0.8169	0.6666	0.7365	0.3987	0.6732	-0.0047	0.6420	0.8790	0.7652
SHV	0.7352	0.6367	**1.0000**	0.7523	0.5391	0.7678	0.5992	0.1036	0.9476	-0.0072	0.7074	0.6008	0.4898
WK	0.3392	0.3116	0.0958	**1.0000**	0.8153	0.7470	0.7929	0.3286	0.7850	-0.0033	0.7206	0.8766	0.7377
BPCf	0.9231	0.8579	0.6607	0.2959	**1.0000**	0.5761	0.7689	0.4441	0.5600	-0.0022	0.5345	0.8695	0.8863
WIE	0.7088	0.6666	0.7678	0.0744	0.6873	**1.0000**	0.7821	0.1380	0.7559	-0.0073	0.8380	0.6384	0.5694
KIR	0.7562	0.7365	0.5992	0.1620	0.8185	0.7821	**1.0000**	0.3303	0.5923	-0.0066	0.7345	0.8055	0.7585
KLE	0.3452	0.3987	0.1036	0.3325	0.3747	0.1380	0.3303	**1.0000**	0.1314	0.0060	0.1075	0.4055	0.4675
EST	0.7761	0.6732	0.9476	0.1340	0.6864	0.7559	0.5923	0.1314	**1.0000**	-0.0032	0.6956	0.6350	0.5030
BL	0.0058	0.0047	0.0037	0.0003	0.0058	-0.0016	-0.0000	0.0100	0.0033	**1.0000**	-0.0057	-0.0050	-0.0064
KEM	0.6848	0.6420	0.7074	0.0847	0.6671	0.8380	0.7345	0.1075	0.6956	-0.0005	**1.0000**	0.6292	0.5038
RWB	0.8895	0.8790	0.6008	0.2945	0.9093	0.6384	0.8055	0.4055	0.6350	0.0068	0.6292	**1.0000**	0.7959
BPC	0.7566	0.7575	0.4538	0.3078	0.7480	0.4820	0.6818	0.5083	0.4753	0.0097	0.4451	0.7872	**1.0000**

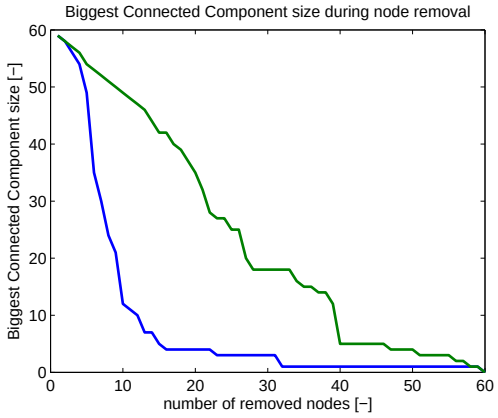

Figure 3: Example of Biggest Connected Component size recorded when nodes are sequentially removed by following two criticality rankings. The network is an Albert-Barabási (AB) 60-nodes graph. The two criticality rankings are BPC (lower curve) and BL (random baseline, upper curve) and are computed once before starting to remove nodes. The BPC ranking is more efficient in detecting the critical nodes, as their removal quickly disconnects the network in small pieces.

In this Table 2, we observe that, for three of the four considered tasks (for both disconnection strategies, single ranking and 100-ranking, on Albert-Barabási (AB) graphs and 100-ranking on Erdős-Rényi (ER) graphs but not for single ranking on ER graphs), the Friedman/Nemenyi test [45] cannot conclude that our proposed model (BPC) is better than its approximation (BPCf), and vice versa. On the ER graphs, single ranking (column 2 of Table 2), the results obtained by BPC and BPCf are significantly different but still

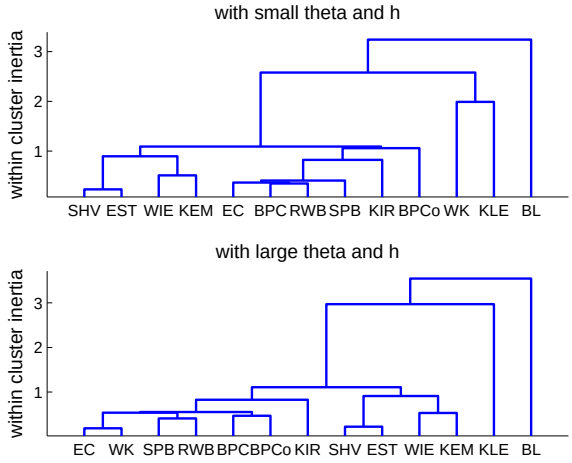

Figure 4: Ward dendrograms of studied criticality measures. Distances are based on Kendall's correlation of Table 5. The smaller the height (Y-axis) of joining branches, the closer the measures. As BPCf, BPC and WK depend on a parameter, two cases are considered: a larger value of the parameters and a smaller value. The small θ and h are 10^{-6} and 1, respectively, while the larger θ and h are 10 and 6.

close in comparison to the other criticalities. It means that the considered approximation seems reasonable, at least on the studied datasets.

We further observe from the same experiments (Table 2) that BPC is significantly better than all the other tested measures on ER graphs. On AB graphs, it cannot be concluded that BPC is significantly better than RWB in the case where only one ranking is performed (single ranking). This is probably related to the fact that BPC is based on a random walk, as RWB does. Moreover, if an updated ranking is used instead (100-ranking), then BPC is not significantly better than WK – while still obtaining better performances. We conclude that the introduced criticality measures (BPC and BPCf) perform well in all contexts as they always perform better (and, most of the time, significantly better) than the competing measures, at least on the investigated data sets. However, this advantage is not always statistically significant when compared to RWB (single ranking on AB graphs) and WK (100-ranking on AB graphs).

Besides this, when examining the results of the other criticality measures, we often find the RWB, KIR, WK and SPB measures in the top-5 best methods (Table 2). Note also that the EC (the degree) is quite efficient

combined with multiple ranking on AB graphs, given its simplicity. At the bottom of the rankings, KLE, WIE, KEM, EST, and SHV often appear to be even less effective than EC. Since EC is a really obvious measure that can be easily computed, it would certainly be interesting to use EC instead of other, more sophisticated, measures in many situations. In particular, EC is quite efficient on AB graphs, if recomputed after each node deletion. It can also be noted that KLE is not performing well on ER graphs (it can even be worse than the random baseline BL, but its mean AUC is still better). We unfortunately do not have a clear explanation of why this is the case. All these conclusions are confirmed in Table 3 where the results of both disconnection strategies (single ranking and ranking updated (100-ranking)) are pooled in order to have an idea of the best method, independently of the ranking strategy.

It is also interesting to identify the most chosen θ and h parameter values from Table 4. For h, it depends on to task to fulfill but the best h value is usually small (1 to 4), and for θ it is better to take a value between 1 and 10. Notice that BPCf still exhibits the best mean rank when its parameter is fixed (results not presented here; see the discussion at the end of this section).

From Table 5, it is clear that WK's correlation with the other measures varies a lot depending of the h value. On the other hand, BPC's and BPCf's correlation with the other measures are less dependent of θ. Note that it was expected that those measures should be highly correlated with RWB and EC when θ is small and with SPB when θ is large, as the bag-of-paths betweenness does [26]. However, we observe that this is not the case for a large θ: the criticality measures BPC and BPCf are still more correlated with RWB when $\theta = 10$. This suggests that the proposed measures capture different properties than the bag-of-paths betweenness.

In Figure 4, we once more notice that the behaviour of WK is strongly dependent of h. It turns out that with small h, its behavior is similar to KLE. When h is larger, the neighborhood is more and more likely to be close to the whole graph, therefore more and more correlated to EC. As from Table 5, BPC's and BPCf's behavior are less sensitive to θ.

From visual inspection of Figure 4, we can identify different clusters of measures:

- WIE, KEM, SHV and EST seem to form a cluster. This is a bit surprising as these measures are based on different properties of the graph, but still provide relatively similar results. Indeed, WIE is based on shortest paths, KEM is based on random walks, SHV is based on

an eigenvalue of $\mathbf{A}$ and EST on paths of different lengths.

- SPB, RWB, KIR, EC, BPCf, BPC are part of another cluster. The same observation can be made: if RWB, BPCf and BPC are based on random walks, SPB is based on shortest paths and KIR is based on the spectrum of the Laplacian matrix. Notice that SPB, RWB, KIR, BPCf and BPC tend to show good performances on Tables 2 and 3.

- KLE looks apart, but is correlated to WK when h is small.

- Finally, notice that the random baseline BL is the last merged measure in the two cases, which looks natural.

Before closing the discussion, let us comment on the presence of parameters. At first sight, it seems unfair to compare measures depending on a parameter (WH, BPC and BPCf) against measures free of parameter. Recall, however, that the attacker can adapt its behavior to the network structure, so that a parameter monitoring the smoothing scale can be considered as an advantage. Moreover, let us recall two facts about the parameter θ of BPC and BPCf. First, measures are not very sensitive to the parameter and, second, its optimal value (according to our experiments) is often close to 1 or 10. Therefore, it seems that we could also just fix this parameter. By the way, we reproduced the experiments by setting $\theta = 1$ and it turns out that BPC was still the best measure for three disconnection strategies while the BPCf was the best for the last one (experiments not reported here).

Finally, to summarize the results, methods can be sorted (the first been the best one) using Borda score ranking [49] based on our empirical results, providing a ranking from best to worse:

- If node ranking is updated after each node deletion, independently of the graph type: BPC, BPCf, RWB, WF, EC, SPB/KIR, EST, KEM, KLE, WIE, SHV.

- If node ranking is not updated after each node deletion, independently of the graph type: BPC, BPCf, RWB, KIR, SPB, WK, EC, KLE, KEM, WIE EST, SHV.

- Finally, independently of the graph type and update factor, the best methods are: BPC, BPCf, RWB, KIR, SPB, WK, EC, KLE, KEM, WIE EST, SHV.

These results are in accordance with the rest of this Section.

6. Conclusion

This paper investigated centrality/criticality measures on graphs through a node disconnection analysis and introduced a new criticality measure based on a bag-of-paths framework and its variant: the bag-of-paths criticality and its fast, approximate, version.

Comparisons based on node disconnection simulations performed on a large number of generated graphs show that those two bag-of-paths criticality measures outperform the other considered centrality/criticality measures. Friedman/Nemenyi tests confirm this fact statistically in most of the cases.

Of course the node disconnection analysis is only a proxy to determine if our criticalities are able to identify "critical" nodes. Our future work will mainly focus on testing the proposed measures on other tasks and to consider other strategies, such as disconnecting *groups of nodes* instead of one single node at each time.

Finally, a simple correlation analysis of those measure allowed to identify coherent groups, namely the WIE, KEM, SHV and EST versus the SPB, RWB, KIR, EC, BPCf and BPC (see Table 1 for acronyms). It was also shown that the choice of the θ parameter does not impact much the behavior of our two proposed criticality measures.

This study has also some limitations. It would, for instance, be useful to confirm the results on larger, real-world, networks – not only artificial graphs. Moreover, other criticality, vulnerability, and betweenness measures not considered here should be investigated as well [1, 8].

7. Acknowledgements

This work was partially supported by the Immediate and the Brufence projects funded by InnovIris (Brussels Region). We thank this institution for giving us the opportunity to conduct both fundamental and applied research.

References

[1] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, T. Zhou, Vital nodes identification in complex networks, Physics Reports 650 (2016) 1–63.

[2] L. Freeman, A set of measures of centrality based on betweenness, Sociometry 40 (1) (1977) 35–41.

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

[4] E. Estrada, D. J. Higham, N. Hatano, Communicability betweenness in complex networks, Physica A: Statistical Mechanics and its Applications 388 (5) (2009) 764–774.

[5] K. Wehmuth, A. Ziviani, Distributed location of the critical nodes to network robustness based on spectral analysis, in: Proceedings of the 7th Latin American Network Operations and Management Symposium (LANOMS), 2011, pp. 1–8.

[6] D. Klein, Centrality measure in graphs, Journal of Mathematical Chemistry 47 (2010) 1209–1223.

[7] A. Tizghadam, A. Leon-Garcia, Betweenness centrality and resistance distance in communication networks, IEEE Network 24 (6) (2010) 10–16.

[8] U. Brandes, T. Erlebach, Network analysis: methodological foundations, Springer-Verlag, 2005.

[9] E. D. Kolaczyk, Statistical analysis of network data: methods and models, Springer, 2009.

[10] S. Wasserman, K. Faust, Social network analysis: methods and applications, Cambridge University Press, 1994.

[11] K. Francoisse, I. Kivimäki, A. Mantrach, F. Rossi, M. Saerens, A bag-of-paths framework for network data analysis, Neural Networks 90 (2017) 90–111.

[12] B. Lebichot, I. Kivimäki, K. Françoisse, M. Saerens, Semi-supervised classification through the bag-of-paths group betweenness, IEEE Transactions on Neural Networks and Learning Systems 25 (2014) 1173–1186.

[13] R. Devooght, A. Mantrach, I. Kivimäki, H. Bersini, A. Jaimes, M. Saerens, Random walks based modularity: Application to semi-supervised learning, in: Proceedings of the 23rd International World Wide Web Conference (WWW '14), 2014, pp. 213–224.

[14] F. Fouss, M. Saerens, M. Shimbo, Algorithms and models for network data and link analysis, Cambridge University Press, 2016.

[15] U. von Luxburg, A. Radl, M. Hein, Getting lost in space: large sample analysis of the commute distance, Proceedings of the 23th Neural Information Processing Systems conference (NIPS 2010) (2010) 2622–2630.

[16] U. von Luxburg, A. Radl, M. Hein, Hitting and commute times in large random neighborhood graphs, Journal of Machine Learning Research 15 (1) (2014) 1751–1798.

[17] R. Albert, H. Jeong, A. L. Barabasi, Error and attack tolerance of complex networks, Nature 406 (6794) (2000) 378–382.

[18] P. Holme, B. J. Kim, C. N. Yoon, S. K. Han, Attack vulnerability of complex networks, Physical Review E 65 (5) (2002) 056109.

[19] F. R. Chung, Spectral graph theory, American Mathematical Society, 1997.

[20] F. Harary, Graph theory, Addison-Wesley, 1969.

[21] S. B. Seidman, Network structure and minimum degree, Social Networks 5 (3) (1983) 269–287.

[22] U. Brandes, A faster algorithm for betweenness centrality, The Journal of Mathematical Sociology 25 (2) (2001) 163–177.

[23] U. Brandes, D. Fleischer, Centrality measures based on current flow, in: Proceedings of the 22nd Annual Symposium on Theoretical Aspects of Computer Science (STACS), 2005, pp. 533–544.

[24] P. G. Doyle, J. L. Snell, Random walks and electric networks, The Mathematical Association of America, 1984.

[25] E. Bozzo, M. Franceschet, Resistance distance, closeness, and betweenness, Social Networks 35 (3) (2013) 460–469.

[26] I. Kivimäki, B. Lebichot, J. Saramaki, M. Saerens, Two betweenness centrality measures based on randomized shortest paths, Scientific Reports 6, Article number: 19668.

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

[28] J. G. Kemeny, J. L. Snell, A. Knapp, Denumerable Markov chains, Springer-Verlag, 1976.

[29] J. R. Norris, Markov chains, Cambridge University Press, 1997.

[30] P. G. Doyle, The Kemeny constant of a markov chain, Unpublished manuscript available at http://www.math.dartmouth.edu/ doyle (2009) 1–10.

[31] H. Tong, B. A. Prakash, C. Tsourakakis, T. Eliassi-Rad, C. Faloutsos, D. H. Chau, On the vulnerability of large graphs, in: Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM '10), 2010, pp. 1091–1096.

[32] A. Mantrach, L. Yen, J. Callut, K. Francoise, M. Shimbo, M. Saerens, The sum-over-paths covariance kernel: a novel covariance between nodes of a directed graph, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (6) (2010) 1112–1126.

[33] E. T. Jaynes, Information theory and statistical mechanics, Physical Review 106 (1957) 620–630.

[34] J. N. Kapur, H. K. Kesavan, Entropy optimization principles with applications, Academic Press, 1992.

[35] M. Saerens, Y. Achbany, F. Fouss, L. Yen, Randomized shortest-path problems: Two related models, Neural Computation 21 (8) (2009) 2363–2404.

[36] G. H. Golub, C. F. V. Loan, Matrix computations, 3th Ed., The Johns Hopkins University Press, 1996.

[37] C. D. Meyer, Matrix analysis and applied linear algebra, SIAM, 2000.

[38] G. Seber, A matrix handbook for statisticians, Wiley, 2008.

[39] E. Estrada, Network robustness to targeted attacks. the interplay of expansibility and degree distribution, The European Physical Journal B - Condensed Matter and Complex Systems 52 (4) (2006) 563–574.

[40] I. Petreska, I. Tomovski, E. Tenreiro, L. Kocarev, F. Bono, K. Poljansek, Application of modal analysis in assessing attack vulnerability of complex networks, Communications in Nonlinear Science and Numerical Simulation 15 (4) (2010) 1008–1018.

[41] A. Santiago, R. M. Benito, Robustness of heterogeneous complex networks, Physica A: Statistical Mechanics and its Applications 338 (2009) 2234–2242.

[42] A.-L. Barabasi, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.

[43] B. Bollobas, Random graphs, Cambridge University Press, 2001.

[44] M. Newman, Networks: an introduction, Oxford University Press, 2010.

[45] J. Demšar, Statistical comparisons of classifiers over multiple data sets, Journal of Machine Learning Research 7 (2006) 1–30.

[46] J. H. Ward, Hierarchical grouping to optimize an objective function, Journal of American Statistical Association 58 (1963) 236–244.

[47] S. Theodoridis, K. Koutroumbas, Pattern recognition, 4th ed., Academic Press, 2009.

[48] L. Lebart, A. Morineau, M. Piron, Statistique exploratoire multidimensionnelle, 4th ed., Dunod, 1995.

[49] A. D. Taylor, A. M. Pacelli, Mathematics and politics: strategy, voting, power, and proof, Springer-Verlag, 2008.