Estimating Graph Robustness Through the Randic Index

Pasquale De Meo, Fabrizio Messina, Domenico Rosaci, Giuseppe M. L. Sarné, and Athanasios V. Vasilakos

***Abstract*—Graph robustness—the ability of a graph to preserve its connectivity after the loss of nodes and edges—has been extensively studied to quantify how social, biological, physical, and technical systems withstand to external damages. In this paper, we prove that graph robustness can be quickly estimated through the Randic index, a parameter introduced in chemistry to study organic compounds. We prove that Erdos–Renyj (ER) graphs are a good specimen of robust graphs because they lack of a clear modular structure; we derive an analytical expression for the Randic index of ER graphs and use ER graphs as an effective term of comparison to decide about graph robustness. Experiments on real datasets from different domains (scientific collaboration networks, content-sharing systems, co-purchase networks from an e-commerce platform, and a road network) show that real-life large graphs are more robust than ER ones with the same number of nodes and edges. We also observe that if node degree distribution closely follows a power law, then few edges contribute for more than half of the Randic index, thus indicating that the selective removal of those edges has devastating impact on graph robustness. Finally, we describe sampling-based algorithms to efficiently but accurately approximate the Randic index.**

***Index Terms*—Complex network analytics, data analytics, data-driven complex systems modeling, Erdos–Renyj (ER) random graphs, graph robustness, Randic index.**

I. Introduction

A. Motivations

GRAPH robustness (also known as resilience) is the ability of a graph to preserve the connectivity after the failure of some of its nodes and/or edges [1]–[4]. Due to the ability of graphs to represent many physical, social, and biological systems, the study of graph robustness holds a widely recognised position in many branches of science and engineering: the functioning of the Internet [4], social media [5], trust networks [6], airline routes [7], metabolic networks [8], and electrical power grids [9]—just to name few application domains—may be severely impaired if some of the system components are no longer connected.

One of the most celebrated results in the assessment of graph robustness is reported in [3], which describes a process where nodes are iteratively removed from a graph and the variations of some topological properties (like the size of the largest connected component or the diameter) are correspondingly measured. One of the main findings of [3] is that graphs displaying a heavy-tailed degree distribution are highly resilient to the random removal of nodes but they are extremely fragile if attacks target at deleting high degree nodes.

The procedure described in [3] does not scale well on real-life graphs and, therefore, we are in an urgent need of drawing up fast methods to estimate of graph robustness.

B. Main Contributions

In this paper, we make a step toward the definition of an index to quickly evaluate the robustness of a graph.

As already observed in [10], a good specimen of robust graph is given by the so-called *expanders*, i.e., graphs which are sparse but, simultaneously, well connected. In an expander graph, each node has a relatively large neighborhood and, then, we are forced to delete a relatively large fraction of nodes/edges to fragment it into many separated components.

An interesting class of expander graphs is given by Erdos–Renyj (ER) random graphs [1]: for a fixed integer n and $p \in (0, 1]$, the symbol $G(n, p)$ identifies a family of graphs with n nodes and the property that two arbitrarily chosen nodes are connected with probability p (called *edge probability*). If $p \to 1$, an ER graph degenerates to a *complete graph* K_n, i.e., a graph in which all pair of nodes are connected. Due to the existence of multiple backup paths between any pair of nodes, the complete graph is the strongest example of robust graph we can think of.

In this paper, we started studying the robustness of an arbitrary graph G by grounding on some well-known results from graph theory and combinatorics [11], [12]. In agreement with past studies, we focus on a special matrix $\mathcal{L}(G)$ associated with G, called *normalized Laplacian*, and on the distribution of its eigenvalues $\lambda_1, \ldots, \lambda_n$ (also called ℓ-eigenvalues).

Manuscript received March 6, 2017; revised June 30, 2017 and September 18, 2017; accepted October 10, 2017. This paper was recommended by Associate Editor J. Liu. *(Corresponding author: Fabrizio Messina.)*

P. De Meo is with the Department of Ancient and Modern Civilizations, University of Messina, 98168 Messina, Italy (e-mail: pdemeo@unime.it).

F. Messina is with the Department of Mathematics and Computer Science, University of Catania, 95125 Catania, Italy (e-mail: messina@dmi.unict.it).

D. Rosaci is with the DIIES Department, University of Reggio Calabria, 89122 Reggio Calabria, Italy (e-mail: domenico.rosaci@unirc.it).

G. M. L. Sarné is with the DICEAM Department, University of Reggio Calabria, 89122 Reggio Calabria, Italy (e-mail: sarne@unirc.it).

A. V. Vasilakos is with the Laboratory of Networks and Cybersecurity, Innopolis University, Innopolis 420500, Russia (e-mail: t.vasilakos@innopolis.ru).

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2017.2763578

We define the spread $S(G)$ as the average square difference of the ℓ-eigenvalues from their mean $\overline{\lambda}$

$$S(G) = \frac{1}{n}\sum_{i=1}^{n}\left(\lambda_i - \overline{\lambda}\right)^2 \quad \text{with} \quad \overline{\lambda} = \frac{1}{n}\sum_{i=1}^{n}\lambda_i$$

and we prove that $2/n \leq S(G) \leq 1$. We also prove that $S(G)$ achieves its lowest value if and only if G coincides with K_n. Consequently, the lower $S(G)$, the more robust G. However, the calculation of $S(G)$ does not require to calculate the whole spectrum of $\mathcal{L}(G)$, which may be prohibitively time-consuming even on moderately large graphs: in fact, we show that $S(G)$ is proportional to the *Randic index* of a graph [12]–[14], a parameter introduced in theoretical chemistry to study organic compounds.

The calculation of the Randic index $R(G)$ takes linear time on the number of the edges of G and it can be easily recomputed in an incremental way if new edges/nodes are added or deleted from G. Due to these features, the Randic index is a computationally feasible tool for assessing the robustness of large graphs.

Unfortunately, two major drawbacks impede us to directly adopt the spread $S(G)$ [and, equivalently, the corresponding Randic index $R(G)$] to evaluate the robustness of G: first, in fact, we should determine a threshold under which the gap $S(G) - S(K_n)$ can be classified as small but the definition of such a threshold is hard and it critically depends on the application domain we are working in. Second, real graphs are far from being assimilable to complete graphs and, therefore, a real graph could showcase a large gap $S(G) - S(K_n)$ even if the real system it represents has been designed to be highly resistant to adversarial attacks.

To overcome the problems above, we suggest to use ER graphs as *reference models* to evaluate graph robustness. More specifically, given a graph G with m edges we build a family of ER graphs with n nodes and edge probability p equal to $p = 2m/(n(n-1))$. We derive the expected value of the Randic index $R(G_{\mathrm{ER}})$ associated with $G(n, p)$ and we calculate the ratio $\gamma = R(G)/R(G_{\mathrm{ER}})$; intuitively, the lower γ, the more robust G is if compared with an arbitrary graph in $G(n, p)$. In the light of the robustness of $G(n, p)$, we can conclude that G is, in its turn, robust. We study also how to extend our procedure on very large graphs. To this purpose we made use of *sampling algorithms* [15]–[18], to get a quick estimation of the Randic index.

We experimentally test our approach on four real graphs describing: 1) a snapshot of the friendship network in YouTube; 2) a co-authorship network extracted from the DBLP computer science bibliography; 3) a co-purchase network from Amazon; and 4) the road networks associated with the State of the California in USA.

The main findings of this paper are as follows.

1) From our experiments, real graphs were associated with a value of γ ranging from 0.42 to 0.56, thus showcasing a high level of robustness. We also showed that the graphs with the largest value of γ (that are, of course, the less robust graphs) also displayed a clear modular structure.
2) We studied how edges contributed to the Randic index in real graphs. We found that if node degree distribution closely mirrors a power law then a small fraction of edges (around 5%) contributed more than 55% to the Randic index.
3) We experimentally studied the scalability of our sampling algorithms to calculate the Randic index and we found that these techniques, even on modest hardware platforms, were able to yield an accurate Randic index approximation in less than one minute.

C. Plan of This Paper

The plan of this paper is as follows. We start by discussing related literature in Section II. In Section III, we illustrate the relation between the Randic index and the spread of the ℓ-eigenvalues. Section IV is devoted to explain why (and how) ER graphs can be used as proxies to evaluate network robustness; in the same section we derive an analytical expression for the expected value of the Randic index of an ER graph. In Section V, we describe sampling-based algorithms to quickly approximate the Randic index of a graph, while in Section VI we elaborate on the experiments we carried out to test the effectiveness and efficiency of our approach. Finally, in Section VII we draw the conclusion.

II. Related Work

In this section, we review some approaches to evaluate the robustness of a graph. We may broadly classify existing approaches in two categories, namely approaches based on topological analysis (see Section II-A) and approaches relying on the eigenvalues of some matrices associated with a graph (see Section II-B).

A. Assessing Graph Robustness via Topological Analysis

The robustness of a system (and, equivalently, of the graph associated with that system) quantifies the ability of that system to withstand an external damage [2]–[4].

Albert *et al.* [3] focused on *scale-free networks*, i.e., on graphs in which node degree distribution follows a power law. They considered two types of adversarial attacks: the first one randomly deletes nodes along with the edges they emanate and the second one targets at removing first high degree nodes. The main finding of [3] is that scale-free networks are extremely robust under random node failures, but they are highly vulnerable if high degree nodes are removed. Wang *et al.* [19] suggested to alter the network topology by properly adding edges (*link-addition strategy*). The scheme proposed in [19] relies on the *structural holes theory* [20] and it only requires the knowledge of local topology. Experiments provide evidence that the manipulation of graph topology through the clever insertion of edges positively affects graph robustness while mitigating network congestion.

The importance of topological information to assess graph robustness has been highlighted in many application domains like the structural analysis of terrorist [21] and criminal organizations [22], power grid infrastructures [23], military [24], and transportation networks [25].

This paper significantly differs from those reported in this section. First, in fact, we do not make any assumption on the

topology of a graph but we consider datasets spanning different domains which induced graphs may deeply differ in terms of their topological organization.

Second, our main goal is to define a global index to estimate graph robustness instead of simulating the loss of nodes and edges from the graph itself and measuring how these operations affect graph connectivity.

B. Spectral Approaches to Assessing Graph Robustness

A second group of approaches—called *spectrum-based*—rely on the analysis of the eigenvalues of the adjacency matrix or the Laplacian [11] of a graph to evaluate its robustness.

We start presenting robustness metrics derived from the adjacency matrix $\mathbf{A}$ of a graph. The first metric we introduce is the *spectral radius*, i.e., the largest eigenvalue of $\mathbf{A}$. The spectral radius controls the speed at which a dynamic process, like a virus, spreads over a computer network [26], [27]. It has been employed to evaluate the robustness of a graph in [28]–[30]. A further measure to consider is the *spectral gap*, defined as the difference between the largest and the second largest eigenvalue of $\mathbf{A}$. Graphs with a large spectral gap are also good expanders [31] and, therefore, spectral gap is a measure of robustness in [32].

A third metric to mention is *natural connectivity* [33], which is related to the number of alternative paths between any pair of nodes in a graph. Large values of natural connectivity imply the existence of more than one route to get from one node to another: if some nodes/edges were deleted from a graph (thus making the paths containing them unusable), there would exist alternative ways to connect a source node with a target one and this, ultimately, is an index of network robustness [34], [35].

A second group of spectrum-based robustness measures derives from the Laplacian $\mathbf{L}$ of a graph, defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$; here $\mathbf{D}$ is a diagonal matrix such that $\mathbf{D}_{uu}$ equals the degree of u. Among robustness metrics based on the Laplacian we recall the *algebraic connectivity* [36], which is the first nonzero eigenvalue of the Laplacian. Large values of algebraic connectivity imply higher difficulty in breaking a graph into disconnected components and, as such, it has been used to assess graph robustness [37].

We also cite *the Kirchhoff index* $\mathrm{Kf}(G)$ [38] which is defined as follows: we can view a graph G as an electrical circuit in which each node uniquely corresponds to a *point* in the circuit and an edge from node u to node v corresponds to a resistor of 1 ohm. If we connect our circuit to a voltage, the *effective resistance* $\kappa(u, v)$ is a measure of the difficulty of the current to flow from u to v. Observe that effective resistance can be calculated by means of the Kirchhoff's law. The $\mathrm{Kf}(G)$ can be written as the pairwise sum of effective resistances, that is

$$\mathrm{Kf}(G) = \sum_{u \in V} \sum_{v \in V} \kappa(u, v). \tag{1}$$

From a spectral standpoint, $\mathrm{Kf}(G)$ is equal to the product of the number n of nodes and the sum of the inverse of Laplacian eigenvalues [39], [40]; it constitutes a valid parameter to assess graph robustness: the smaller $\mathrm{Kf}(G)$, the more robust a graph is [39], [41].

The calculation of the Kirchhoff index is an active research field with many interesting results: for instance, Liu *et al.* [42], [43] provided some bounds on the Kirchhoff index in case of special graphs and Liu *et al.* [44], [45] studied how to recalculate the Kirchhoff index if some nodes/edges are inserted.

Starting from the metrics defined above, some authors studied how to modify graph topology with the goal of increasing graph robustness. To this end, Chan and Akoglu [46] developed a general framework to modify an input graph by edge rewiring in such a way as to yield the largest increase in robustness under a specified budget (intended as the largest number of allowed rewirings). Arrigo and Benzi [47] described some heuristics to add, delete, or rewire a limited number of edges in a sparse graph to optimize its natural connectivity.

Our approach belongs to the class spectrum-based metrics; unlike the approaches discussed in this section we focus on the normalized Laplacian (see Definition 1) and on the spread $S(G)$ of its eigenvalues around their mean. We provide tight lower and upper bounds on $S(G)$ which allowed us to quantify the robustness of G.

As a further difference, the calculation of the metrics above can be time expensive and, then, suitable approximation methods are required [32]. We do not need to resort to approximate strategies to calculate $S(G)$ because we proved that the calculation of $S(G)$ is equivalent to the calculation of the Randic index [13], which takes linear time in the number of graph edges. As such our approach is viable also on large real graphs. We also observed (see Section VI-A) that graphs lacking of a clear community structure are also the most robust ones (and showcase the lowest values of the Randic index).

III. Randic Index and Graph Robustness

In this section, we describe our approach to evaluating the robustness of a graph. In this paper, we will concentrate on *undirected graphs* (shortly, *graphs*). A graph G is a pair $G = \langle N, E \rangle$, where $N = \{1, \ldots, n\}$ is the set of nodes and $E \subseteq N \times N$ is the set of edges. Two nodes u and v are connected if $\langle u, v \rangle \in E$; the graph G is of *order* n if $|N| = n$ and it has *size* m if $|E| = m$. The degree d_u of a node u is equal to the number of edges incident onto u. A graph G is *connected* if there exists at least one path between any pair of nodes. In the following, we will only consider connected graphs. A graph is *complete* if any node is connected to all other nodes, which implies $d_u = n - 1$ for any node u. A complete graph of order n is indicated as K_n and it contains $([n(n-1)]/2)$ edges.

A graph G of order n is associated with an adjacency matrix $\mathbf{A}$, with $\mathbf{A}_{uv} = 1$ if and only if $\langle u, v \rangle \in E$, 0 otherwise. Two further matrices, which allow to capture the structure of a graph, are the *Laplacian* and the *normalized Laplacian* [11], [12], [48]–[51].

Definition 1 (Laplacian and Normalized Laplacian): Let $G = \langle N, E \rangle$ be a graph and let $\mathbf{D}$ be a diagonal matrix (whose rows and columns are indexed by the nodes of G) such that $\mathbf{D}_{uu} = d_u$. The Laplacian $\mathbf{L}(G)$ of G is

$$\mathbf{L}(G) = \mathbf{D} - \mathbf{A}.$$

The normalized Laplacian $\mathcal{L}(G)$ of G is defined as

$$\mathcal{L}(G) = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}. \tag{2}$$

Here $\mathbf{I}$ is the identity matrix and $\mathbf{D}^{-(1/2)}$ is a diagonal matrix such that $\mathbf{D}_{ii}^{-(1/2)} = (1/\sqrt{d_i})$.

Many authors [11], [12], [48], [50], [51] studied how the eigenvalues $\lambda_1, \ldots, \lambda_n$ (also called ℓ-eigenvalues) of $\mathcal{L}(G)$ are related to the structure of G. Some of the most interesting facts about ℓ-eigenvalues are reported below.

Theorem 1 (Properties of ℓ-Eigenvalues): Let G be a connected graph of order n. The following facts hold true.

1) $\mathcal{L}$ is a positive semidefinite matrix.
2) Let $\lambda_1 \leq \lambda_2 \leq, \ldots \leq \lambda_n \leq 2$ be the ℓ-eigenvalues sorted in increasing order of magnitude. We have that $\lambda_1 = 0$, i.e., 0 is the lowest eigenvalue and it is associated with the eigenvector $\mathbf{1}$.[1]
3) $\sum_{i=1}^{n} \lambda_i = n$, i.e., the sum of the ℓ-eigenvalues equals the order of G.
4) $\lambda_2 = \ldots = \lambda_n$ if and only if G is the complete graph K_n.

Proof: See [11] and [48]. ■

From Theorem 1, we have the following result.

Corollary 1: Let G be a connected graph of order n with $n \geq 2$. Then: 1) $\overline{\lambda} = (1/n)\sum_{i=1}^{n} \lambda_i = 1$ and 2) $\lambda_2 = \ldots = \lambda_n = (n/n-1)$ if and only if G is the complete graph of order n.

Proof: From Theorem 1, part 3, we have

$$\overline{\lambda} = \frac{1}{n}\sum_{i=1}^{n} \lambda_i = \frac{1}{n} \times n = 1.$$

From Theorem 1, part 4 we have that $\lambda_2 = \ldots = \lambda_n = \lambda^\star$ if and only if G coincides with K_n. From parts 2 and 3, we get

$$n = \sum_{i=1}^{n} \lambda_i = \lambda_1 + \sum_{i=2}^{n} \lambda_i = \sum_{i=2}^{n} \lambda_i = (n-1)\lambda^\star$$

which ends the proof. ■

We now discuss how the distribution of ℓ-eigenvalues relates to the robustness of G. To do so, let us define the *spread* $S(G)$ of ℓ-eigenvalues as the mean square deviation of ℓ-eigenvalues from their average (that, according to Corollary 1, is equal to $\overline{\lambda} = 1$)

$$S(G) = \frac{1}{n}\sum_{i=1}^{n}(\lambda_i - 1)^2 = \frac{1}{n} + \frac{1}{n}\sum_{i=2}^{n}(\lambda_i - 1)^2. \tag{3}$$

The following theorem provides a lower and an upper bound on $S(G)$.

Theorem 2: Let G be a connected graph of order n with spread $S(G)$. Then the following bounds hold:

$$\frac{2}{n} \leq S(G) \leq 1$$

and $S(G) = (2/n)$ if and only if G is the complete graph with n vertices.

Proof: First, we prove that $S(G) \leq 1$. From Theorem 1, part 2, we have $0 \leq \lambda_i \leq 2$, we get $(\lambda_i - 1)^2 \leq 1$ which, in combination with (3), yields the following upper bound on $S(G)$:

$$S(G) = \frac{1}{n} + \frac{1}{n}\sum_{i=2}^{n}(\lambda_i - 1)^2 \leq \frac{1}{n} + \frac{n-1}{n} = 1.$$

To prove that $S(G) \geq 2/n$, we observe that

$$\begin{aligned}\sum_{i=1}^{n}(\lambda_i - 1)^2 &= \sum_{i=1}^{n}\left(\lambda_i^2 - 2\lambda_i + 1\right) \\ &= \sum_{i=1}^{n}\lambda_i^2 - 2\sum_{i=1}^{n}\lambda_i + n = \sum_{i=1}^{n}\lambda_i^2 - 2n + n \\ &= \sum_{i=1}^{n}\lambda_i^2 - n\end{aligned}$$

which implies

$$S(G) = \frac{1}{n}\sum_{i=1}^{n}(\lambda_i - 1)^2 = \frac{1}{n}\left(\sum_{i=1}^{n}\lambda_i^2 - n\right) = \frac{1}{n}\sum_{i=1}^{n}\lambda_i^2 - 1.$$

Provided that n is fixed, the lowest value of $S(G)$ can be obtained by solving the following optimization problem:

$$\underset{\lambda_i}{\text{minimize}} \quad \sum_{i=1}^{n}\lambda_i^2 \quad \text{subject to} \quad \sum_{i=1}^{n}\lambda_i = n. \tag{4}$$

With the change of variables $\Lambda_i = \lambda_i/n$, we get the following optimization problem:

$$\underset{\Lambda_i}{\text{minimize}} \quad \sum_{i=1}^{n}\Lambda_i^2 \quad \text{subject to} \quad \sum_{i=1}^{n}\Lambda_i = 1. \tag{5}$$

If we apply the Lagrange multiplier technique [52], we obtain that the optimal solution is $\Lambda_i = 1/n$, i.e., $\lambda_i = n/(n-1)$ for any i. Because of Corollary 1, the condition $\lambda_2 = \lambda_3 = \cdots = \lambda_n = n/(n-1)$ is valid if and only if G coincides with K_n. In this case

$$S(K_n) = \frac{1}{n} + \frac{1}{n}\sum_{i=2}^{n}\left(\frac{n}{n-1} - 1\right)^2 = \frac{2}{n}.$$

■

The spread $S(G)$ achieves its lower bound if G coincides with K_n, which is the most robust graph we can think of. Consequently, the gap $S(G) - S(K_n)$ specifies how $S(G)$ is robust if compared to K_n.

Unfortunately, calculating $S(G)$ implies the calculation of the ℓ-eigenvalues and such a procedure is computationally expensive on even moderate size graphs. In practice, the computation of $S(G)$ is equivalent to the calculation of the *Randic index* [13], [14], a parameter widely used in theoretical chemistry to study organic compounds. The Randic index $R(G)$ of G is defined as follows.

Definition 2 (Randic Index): Let G be a connected graph of order n. The Randic index $R(G)$ associated with G is defined as

$$R(G) = \sum_{\langle u,v\rangle \in E} \frac{1}{d_u d_v}. \tag{6}$$

The Randic index $R(G)$ is related to the ℓ-eigenvalues, as specified in the next theorem.

[1] Here $\mathbf{1}$ is the vector with all of its entries equal to 1.

Theorem 3: Let G be a connected graph of order n. Let $0 = \lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$ be the ℓ-eigenvalues sorted in increasing order of magnitude. The Randic index $R(G)$ of G verifies the following equality:

$$\sum_{i=1}^{n} \lambda_i^2 = n + 2R(G). \tag{7}$$

Proof: See [12]. ■

Now, we use the above Theorem 3 to prove Theorem 4, which says us that there exists a relation between $S(G)$ and the Randic index.

Theorem 4: Let G be a connected graph of order n. The Randic index $R(G)$ and the spread $S(G)$ verify the following equality:

$$S(G) = \frac{R(G)}{n}. \tag{8}$$

Proof: From the definition of $S(G)$, we get

$$S(G) = \frac{1}{n}\sum_{i=1}^{n}(\lambda_i - 1)^2 = \frac{1}{n}\left(\sum_{i=1}^{n}\lambda_i^2 - n\right) = \frac{2}{n}R(G).$$

Here, we rely on the equality $\sum_{i=1}^{n}\lambda_i^2 - 2n + n = \sum_{i=1}^{n}\lambda_i^2 - n$ already established in Theorem 2. ■

By Theorem 4, we can calculate the spread of G by computing its Randic index. Such a procedure takes $O(m)$ steps, where m is the number of edges of the graph G, and, therefore, it is suitable also on large graphs. The Randic index will vary from 1 (in case of K_n) to n and the lower $R(G)$, the more robust G is.

IV. Using Random Graphs As Proxies to Measure Robustness

A. Why ER Graphs Are Effective in Estimating Graph Robustness

Our procedure to assess the robustness of a graph G of order n requires to calculate the gap $R(G) - R(K_n) = R(G) - 1$ (see Theorems 2 and 4; the smaller the gap, the better G will resemble K_n and, then, the more robust G. However, graphs describing real-life systems are far from being assimilable to complete graphs because they have much less edges than a complete graph. Therefore, the graph associated with a real-system may showcase a large Randic index despite the underlying system has been designed to be robust to node and edge failures. As such, comparing G with K_n may be ineffective to decide about the robustness of G.

To overcome this problem, we suggest to take Erdos-Renyi (ER) random graphs [1] as term of comparison.

Given an integer $n > 0$ and a real number $p \in (0, 1]$ an ER graph $G(n, p)$ is a family of graphs of order n in which nodes are connected independently and uniformly at random with probability p. In an ER graph $G(n, p)$, the expected number m of edges is $m = \binom{n}{2}p$. Random graphs display good *expansion properties* [10]: roughly speaking, a graph is said to be a good expander if it is sparse and well connected. In words, in an expander, any subset of nodes has a relatively large number of neighbors and, then, if we would remove a relatively large fraction of edges we are expected to not fragment G into disjoint components. Expansion properties are closely related to robustness [32] and then, we can take ER graphs as a reference to evaluate graph robustness.

Based on the considerations above, our strategy to evaluate the robustness of a graph G requires to calculate the ratio $\gamma_{G,G_{\mathrm{ER}}} = R(G)/R(G_{\mathrm{ER}})$, i.e., the ratio of the Randic index of G to the expected value of the Randic index of an ER graph of the same order and size of G. The parameter $\gamma_{G,G_{\mathrm{ER}}}$ quantifies the gain, in terms of the Randic index, of selecting G rather than drawing a specimen from $G(n, p)$ at random; of course, the lower $\gamma_{G,G_{\mathrm{ER}}}$, the more robust G.

In what follows, we provide an analytical expression for the expected value of the Randic index of an ER graph with n nodes and edge probability equal to p.

B. Randic Index of ER Graph

In this section, we derive an analytical expression for the Randic index of ER graphs.

Let M be the binomial random variable defining the number of edges in an ER graph $G(n, p)$; it is easy to check that $\mathbb{E}(M) = \binom{n}{2}p$ [1]. The Randic index of an ER graph is a random variable, whose expected value is as follows.

Theorem 5: Let G be an ER graph of order n and edge probability $p \in (0, 1]$. The expected value $\mathbb{E}(R(G))$ of its Randic index is

$$\mathbb{E}(R(G)) = \frac{\left(1 - (1-p)^{n+1}\right)^2}{p}. \tag{9}$$

Theorem 5 tells us that the Randic index of an ER graph diverges if $p \to 0$, independently of n. If $p \to 1$, then $R(G)$ approaches 0 as quick as $(1/p)$.

We provide a proof of Theorem 5 in the supplementary material.

V. Fast Computation of the Randic Index

In this section, we describe a fast procedure to approximate the Randic index of a graph.

As said before, the computation of the Randic index scales linearly with the graph size; however, many real-world systems are described through gigantic graphs, thus making the calculation of the Randic index infeasible. Many algorithms are available to sample subgraphs from large graphs [15]–[18], [53]–[57]. Sampled subgraphs can be used to get a fast (but often realistic) estimation of the graph properties (like degree distribution or diameter) as well as to quantify the strengths and limitations of an algorithm before deploying it on the market [17], [53].

Given sampling algorithm $\mathcal{S}$, we define the *relative error* ε of $\mathcal{S}$ as follows.

Definition 3: Let $G = \langle N, E\rangle$ be a graph and let $\mathcal{S}$ be a sampling algorithm which, if applied to G, generates the graph $G^\star = \langle N^\star, E^\star\rangle$. Let $R(G)$ and $R(G^\star)$ be the Randic indices of G and $G^\star$, respectively. The *relative error* ε of $\mathcal{S}$ is defined as follows:

$$\varepsilon = \frac{|R(G) - R(G^\star)|}{\max\{R(G), R(G^\star)\}}. \tag{10}$$

Algorithm 1 Induced Edge Sampling Algorithm

```
1: procedure INDUCED_EDGE_SAMPLING
2: Require: G: a graph, φ*: a real number in [0, 1]
3: Ensure: A graph G* = ⟨N*, E*⟩, subgraph of G
4:     G* ← EDGE_SAMPLING(G, φ*)
5:     N* ← NODES(G*)
6:     E* ← EDGES(G*)
7:     G' ← SUBGRAPH(G, N*) with G' = ⟨N', E'⟩
8:     for e ∈ E' − E* do
9:         ADD_EDGE(G*, e)
10:    end for
11: return G*
12: end procedure
```

Observe that $0 \leq \varepsilon \leq 1$ and the lower ε, the more accurate $\mathcal{S}$.

A popular strategy to sample large graphs is based on *random walks* [15], [16], [53], [57]. Methods based on random walks have been successfully applied in many application domains like the Internet but they may fail to produce accurate results in other ones: for instance, in case of Twitter, it is a well-known fact that samples collected via random walks are biased toward users with a large number of followers or users who frequently tweets [58]–[61].

In many real-life systems, uniformly sampling edges/nodes is conducive to surprisingly effective results [62]; therefore, we applied node and edge sampling algorithms to efficiently calculate $R(G)$.

The NODE_SAMPLING procedure takes a graph $G = \langle N, E\rangle$ and a real number $\phi^\star \in [0, 1]$. Here, $\phi^\star$—called *sample rate*—specifies the fraction of nodes to sample from G. NODE_SAMPLING consists of two steps.

- Step 1: Select, *uniformly at random*, $\lceil \phi^\star \times |N| \rceil$ nodes from N and let $N^\star$ be the set of sampled nodes. Each node $u \in N$ will *uniquely correspond* to a node $u^\star \in N^\star$.
- Step 2: Draw an edge between two nodes $u^\star$ and $v^\star$ *if and only if* there is an edge from u to v in E and inserts it in $E^\star$.

NODE_SAMPLING returns $G^\star = \langle N^\star, E^\star\rangle$ as output.

The procedure for edge sampling is defined in a similar way. We assume that a *primitive procedure*—called EDGE_SAMPLING—is available. As in node sampling, the EDGE_SAMPLING procedure takes a graph $G = \langle N, E\rangle$ and a real number $\phi^\star \in [0, 1]$ and it produces a graph $G^\star = \langle N^\star, E^\star\rangle$ (with $|N^\star|$ of order $\lceil \phi^\star \times |N| \rceil$ nodes) as output. The EDGE_SAMPLING operator chooses, *uniformly at random*, an edge $e = \langle u, v\rangle$ from E and inserts it in $E^\star$; nodes u and v are inserted into $N^\star$ too.

The EDGE_SAMPLING procedure preferentially selects nodes with large degree (which are those ones contributing the less to the value of the Randic index) and this may cause an underestimation of the Randic index. To overcome this problem, we implemented a slight modification of the EDGE_SAMPLING procedure, which is called INDUCED_EDGE_SAMPLING and sketched in Algorithm 1.

The INDUCED_EDGE_SAMPLING procedure builds a subgraph $G^\star = \langle N^\star, E^\star\rangle$ from G by calling the EDGE_SAMPLING procedure. Then, it calls the procedure SUBGRAPH (line 7), which returns the subgraph G' of G induced by the nodes in $N^\star$. Finally, edges present in E' but absent in $E^\star$ are inserted in $G^\star$.

The time complexity of the INDUCED_EDGE_SAMPLING procedure is parametric in $\phi^\star$ and, to this extent, we will experimentally investigate its efficiency in Section VI-C. By construction, the INDUCED_EDGE_SAMPLING procedure will build up a graph of order $|N^\star| = \lceil \phi^\star \times |N| \rceil$ and size varying from $|N^\star|$ to $|N^\star|^2$. Therefore, the time complexity of INDUCED_EDGE_SAMPLING procedure ranges from $O(|N^\star|)$ (in case of sparse induced subgraphs) to $O(|N^\star|^2)$ (in case of dense induced subgraphs). This implies that the smaller $|N^\star|$ (or, equivalently, the closer $\phi^\star$ to 0), the faster our sampling approach.

VI. EXPERIMENTAL EVALUATION

In this section, we report on the experiments we carried out to test the effectiveness of Randic index in assessing graph robustness. We designed our experiments to answer the following questions.

- Q_1: How robust—in term of Randic index—are real graphs if compared with ER graphs of the same order and size?
- Q_2: What are the edges influencing the most the Randic index of real graphs?
- Q_3: Are the sampling techniques described in Section V effective and efficient in approximating the Randic index?

We used four real-life datasets (all datasets have been taken from http://konect.uni-koblenz.de/) to perform our experimental analysis. The first dataset (`Amazon`) is a co-purchase network extracted from Amazon: nodes identify products and edges specify that two products have been frequently bought together. The second dataset (`California`) describes the road network of the State of California in USA: nodes of the network are the intersections between roads and road endpoints, and the edges are road segments between intersections and road endpoints. The third dataset (`DBLP`) is a co-authorship network extracted from the DBLP archive: nodes are authors and two nodes are connected if the corresponding authors have published at least one paper together. The fourth dataset (`Youtube`) is the friendship network extracted from Youtube: nodes identify users and two nodes are connected through an edge if the corresponding users are friends. Table I lists the order n and the size m of each of these graphs. Experiments were performed on a machine equipped with a CPU Intel Xeon, model E31220 @ 3.10 GHz and 4 GB of RAM.

A. Comparing Real Graphs With Artificial Ones

In this section, we describe the experiments to answer $\mathbf{Q}_1$.

For each graph G in our dataset we calculated the ratio $\gamma_{G,G_{\text{ER}}}$ introduced in Section IV. In Table I, we report the edge probability p of the corresponding ER graphs (fourth column) as well as the value of $\gamma_{G,G_{\text{ER}}}$ (fifth column).

TABLE I
DATASETS EMPLOYED IN THIS PAPER AND VALUES OF γ

Real Dataset				ER Graph	
Dataset	Nodes (n)	Edges (m)	Q	p	$\gamma_{G,G_{ER}}$
Amazon	334 863	925 872	0.6695	1.65×10^{-5}	0.4834
Calif.	1 965 206	2 766 607	0.7829	1.43×10^{-6}	0.5646
DBLP	317 080	1 049 866	0.7972	2.09×10^{-5}	0.5384
YouT.	1 134 890	2 987 624	0.6767	4.64×10^{-6}	0.4206

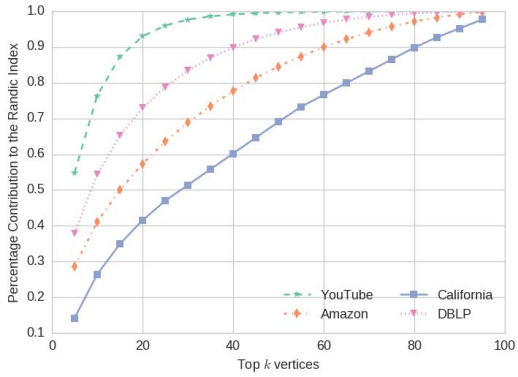

Fig. 1. Horizontal axis: percentage of top k edges, sorted on the basis of their inverse product connectivity. Vertical axis: how these edges contribute to the whole Randic index.

From our tests, we can observe that $\gamma_{G,G_{\mathrm{ER}}}$ varies from 0.42 to 0.56, which implies that the Randic index of each of the input graphs is less than the average value of the expected value of the Randic index of an ER graph of the same order and size. Because of the lower the Randic index the more robust a graph and because ER graphs are good specimen of robust graphs, we may conclude that real system display a high level of robustness.

To get an explanation about the robustness of graphs associated with real-life systems, we inspected the community structure of our input graph. A community (or cluster or module) in a graph [63] is a group of nodes displaying a high intracommunity edge density and a low intercommunity edge density. A clear community structure is responsible for low level of robustness [32]: in fact, if a graph would display a clear community structure, we could fragment it into many disjoint components by deleting edges running from a community to another one. In the light of these observations, we applied the label propagation algorithm (LPA) [64] to find out communities. We ended up deciding on LPA because of its accuracy and running time which is linear in the graph size, thus making LPA suitable to quickly process even large graphs. We used the Q modularity score [63], [65] to evaluate the quality of the obtained partitioning. The Q modularity score ranges from $-(1/2)$ to 1 [66], and the higher Q, the stronger the division of a graph in communities is.

We observe that Q varies from 0.66 to 0.79 (see Table I, third column); values of Q appear to be directly related to those of $\gamma_{G,G_{\mathrm{ER}}}$, thus signalling that the most robust graphs are also those graphs showcasing the lowest values of Q.

B. Contribution of Edges to Randic Index

Many studies in complex network theory witness the unequal contribution that nodes and edges have on graph topology: for instance, in the World Wide Web graph, node degree follows a power law distribution because only few nodes collects most of the edges [67], [68].

Our second question is, therefore, on the contribution that edges have on the Randic index: because node degrees are unevenly distributed, we hypothesize that some edges have a bigger impact on $R(G)$ than others. To run our analysis, we sorted edges $\langle u, v\rangle$ in decreasing order of the factor—called *inverse product connectivity*—$1/(d_u d_v)$. We evaluated the contribution that the fraction formed by the k% of edges with the largest inverse product connectivity had on $R(G)$ (see Fig. 1).

We notice that, in all graphs under inquiry, a small fraction of edges significantly impacted on the overall Randic index: for instance, in case of YouTube, the top 5% edges contributed for about 55% of the Randic index. A less evident impact was recorded in case of California: the top 20% vertices contributed slightly more than 40% to the Randic index. To explain the differences emerging across different graphs, we studied the node degree distribution in each of these graphs. We used the method described in [69] to check if the probability $P(x)$ that a node has degree equal to x is shaped as $P(x) \propto x^{\alpha}$, being α a suitable coefficient to estimate from empirical data. In general α should vary from 1.8 to 3.5 to claim that node degree actually follows a power law distribution.

From our test, only node degree distribution in YouTube can be classified as a power law ($\alpha = 2.14$); this depends on the fact that in YouTube there are few users collecting most of the friendship relationships while the vast majority of user has very few connections. If we would cut from the YouTube social graph those edges linking low degree nodes, we would disconnect the social graph into several components and this would exert a big variation on the Randic index. As an opposite case, let us consider the case of California dataset: here node degree distribution does not follow a power law ($\alpha = 8.99$) and, as expected, the impact of the edges with the largest inverse product connectivity factor on the Randic index is much softer than that observed in YouTube. This is likely to depend on the mechanism underlying the design of a road network: any point in the network should be well connected to the others to avoid congestion, improve environment conditions and ensure alternative routes in case of incidents.

These requirements often conflict with budget constraints which limit the number and extension of roads we are allowed to build. Node degree distribution in a road network is more regular than in YouTube and, then, there are less critical edges which significantly contribute to the Randic index. An intermediate configuration occurs in case of DBLP ($\alpha = 3.26$) and Amazon ($\alpha = 3.58$).

C. Accuracy and Efficiency of Sampling Techniques

In this section, we analyze the effectiveness and efficiency of the sampling algorithms described in Section V.

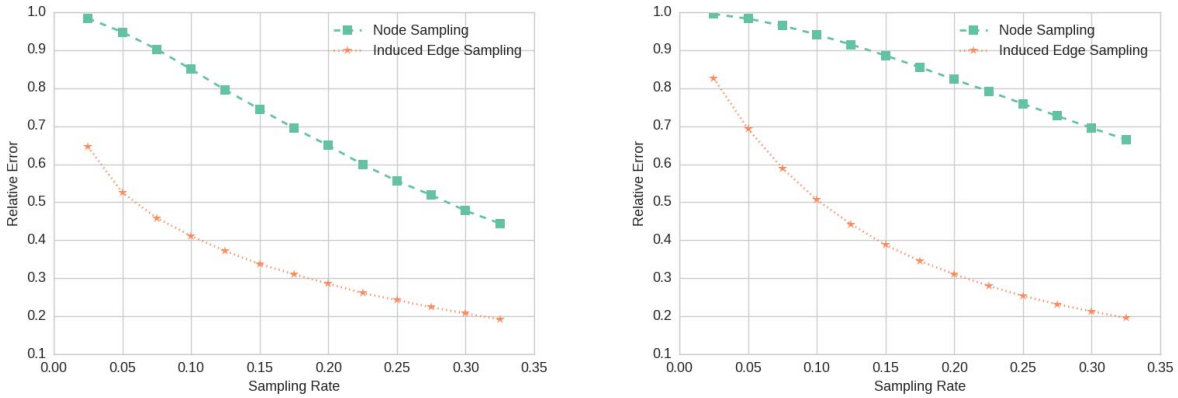

Fig. 2. Left: relative error ε versus $\phi^\star$ for NODE_SAMPLING/INDUCED_EDGE_SAMPLING procedures on `Amazon` graph. Right: relative error ε versus $\phi^\star$ for NODE_SAMPLING/INDUCED_EDGE_SAMPLING procedures on `California` graph.

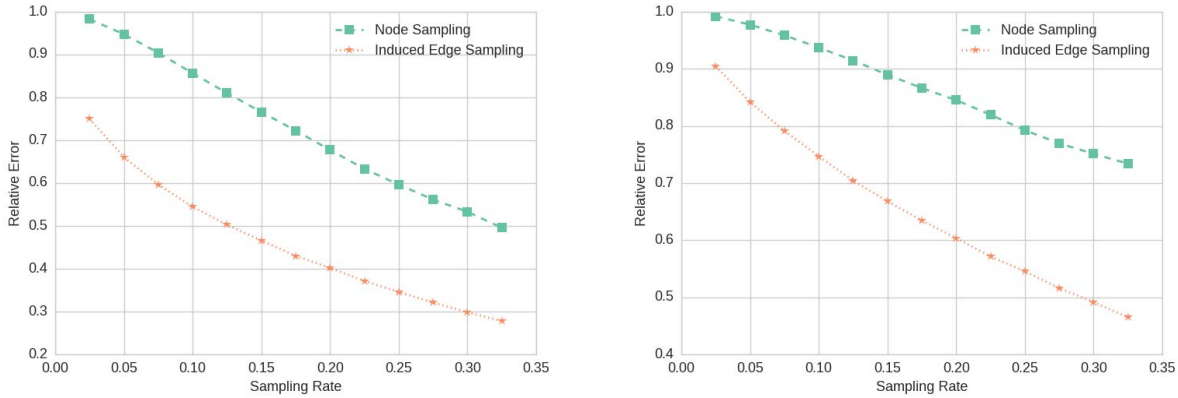

Fig. 3. Left: relative error ε versus $\phi^\star$ for NODE_SAMPLING/INDUCED_EDGE_SAMPLING procedures on `DBLP` graph. Right: relative error ε versus $\phi^\star$ for NODE_SAMPLING/INDUCED_EDGE_SAMPLING procedures on `YouTube` graph.

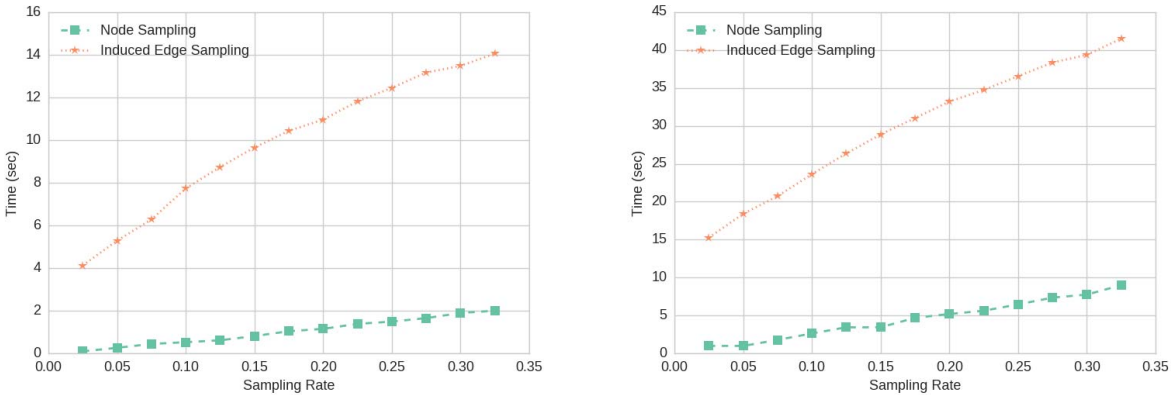

Fig. 4. Left: running time versus $\phi^\star$ for NODE_SAMPLING/INDUCED_EDGE_SAMPLING procedures on `Amazon` graph. Right: running time versus $\phi^\star$ for NODE_SAMPLING/INDUCED_EDGE_SAMPLING procedures on `California` graph.

We considered a sample rate $\phi^\star$ varying from 0.01 to 0.3 and we measured the relative error ε [see (10)] for both NODE_SAMPLING and INDUCED_EDGE_SAMPLING procedures. The obtained results are reported in Figs. 2 and 3. First, observe that INDUCED_EDGE_SAMPLING always outperforms NODE_SAMPLING in terms of accuracy; in addition, as $\phi^\star$ gets larger and larger, ε for INDUCED_EDGE_SAMPLING decays faster than in NODE_SAMPLING. Under the same sample rate $\phi^\star$, the NODE_SAMPLING procedure fails to catch many of the edges that, instead, are sampled by the INDUCED_EDGE_SAMPLING procedure; differences in sampled subgraphs have a clear impact on the estimated Randic index, as Figs. 2 and 3 show. Independently of $\phi^\star$, the lowest relative error ε is observed on `YouTube` dataset.

We also measured the running time needed for a single run of procedures NODE_SAMPLING and INDUCED_EDGE_SAMPLING by averaging over a total of ten runs (see Figs. 4 and 5).

NODE_SAMPLING is 4–20 times faster than INDUCED_EDGE_SAMPLING for small values of $\phi^\star$.

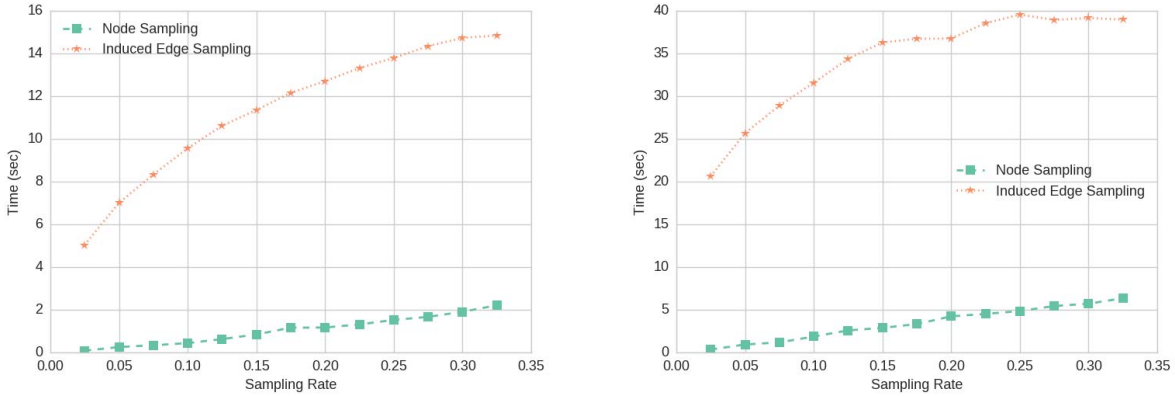

Fig. 5. Left: running time versus $\phi^\star$ for NODE_SAMPLING/INDUCED_EDGE_SAMPLING procedures on DBLP graph. Right: running time versus $\phi^\star$ for NODE_SAMPLING/INDUCED_EDGE_SAMPLING procedures on YouTube graph.

The largest gap is observed in DBLP and YouTube, i.e., on graphs best resembling power law networks. In this case, if we would select few high degree nodes we would be able to collect a large fraction of edges and the graph sampled by the INDUCED_EDGE_SAMPLING procedure would contain much more edges than that generated by NODE_SAMPLING. This explains observed differences in running times.

VII. CONCLUSION

In this paper, we describe an approach to estimating the robustness of large graphs through the Randic index, a parameter introduced in chemistry to study organic compounds. We proved that graph robustness can be measured through the spread of the eigenvalues of normalized Laplacian associated with the graph around their mean. We showed that the calculation of the spread is equivalent to calculate the Randic index of a graph. Then, we observed that, due to their good expansion properties, ER graphs are good specimen of robust graphs and we derived an analytical expression of the expected Randic index of an ER graph of given order and size. We described efficient sampling-based algorithms to get a quick estimation of the Randic index on large graphs.

A main takeaway of our analysis is that graphs lacking of a clear community structure are also the most robust ones. In addition, if node degree distribution approximates well a power law distribution, then a small fraction of the edges account for more than half of the Randic index.

As for our future work, we plan to study the relation between the modular structure of a graph and its robustness and we target at designing novel community detection algorithms based on the notion of Randic index, also in case of graphs in which nodes are equipped with labels [70]. We also plan to study in detail the relationship between Randic index and other graph robustness metrics.

REFERENCES

[1] M. E. J. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, 2003.
[2] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Network robustness and fragility: Percolation on random graphs," *Phys. Rev. Lett.*, vol. 85, no. 25, pp. 5468–5471, 2000.
[3] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.
[4] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, "Resilience of the Internet to random breakdowns," *Phys. Rev. Lett.*, vol. 85, no. 21, pp. 4626–4628, 2000.
[5] P. Andriotis, G. Oikonomou, T. Tryfonas, and S. Li, "Highlighting relationships of a smartphone's social ecosystem in potentially large investigations," *IEEE Trans. Cybern.*, vol. 46, no. 9, pp. 1974–1985, Sep. 2016.
[6] P. De Meo, K. Musial-Gabrys, D. Rosaci, G. M. L. Sarnè, and L. Aroyo, "Using centrality measures to predict helpfulness-based reputation in trust networks," *ACM Trans. Internet Technol.*, vol. 17, no. 1, p. 8, 2017.
[7] O. Lordan, J. M. Sallan, and P. Simo, "Study of the topology and robustness of airline route networks from the complex network approach: A survey and research agenda," *J. Transp. Geograph.*, vol. 37, pp. 112–120, May 2014.
[8] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. Barabási, "The large-scale organization of metabolic networks," *Nature*, vol. 407, no. 6804, pp. 651–654, 2000.
[9] D. Chassin and C. C. Posse, "Evaluating North American electric grid reliability using the Barabási–Albert network model," *Phys. A Stat. Mech. Appl.*, vol. 355, no. 2, pp. 667–677, 2005.
[10] S. Hoory, N. Linial, and A. Wigderson, "Expander graphs and their applications," *Bull. Amer. Math. Soc.*, vol. 43, no. 4, pp. 439–561, 2006.
[11] F. R. K. Chung, *Spectral Graph Theory*, vol. 92. Providence, RI, USA: Amer. Math. Soc., 1997.
[12] M. Cavers, "The normalized Laplacian matrix and general Randic index of graphs," Ph.D. dissertation, Faculty Graduate Stud. Res., Univ. at Regina, Regina, SK, Canada, 2010.
[13] M. Randic, "Characterization of molecular branching," *J. Amer. Chem. Soc.*, vol. 97, no. 23, pp. 6609–6615, 1975.
[14] X. Li and Y. T. Shi, "A survey on the Randic index," *Commun. Math. Comput. Chem.*, vol. 59, no. 1, pp. 127–156, 2008.
[15] J. Leskovec and C. Faloutsos, "Sampling from large graphs," in *Proc. ACM Int. Conf. Knowl. Disc. Data Min. (SIGKDD)*, Philadelphia, PA, USA, 2006, pp. 631–636.
[16] M. Gjoka, M. Kurant, C. Butts, and A. Markopoulou, "Walking in Facebook: A case study of unbiased sampling of OSNs," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, San Diego, CA, USA, 2010, pp. 2498–2506.
[17] N. K. Ahmed, J. Neville, and R. Kompella, "Network sampling: From static to streaming graphs," *ACM Trans. Knowl. Disc. From Data*, vol. 8, no. 2, pp. 1–56, 2013.
[18] S. A. Catanese, P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, "Crawling Facebook for social network analysis purposes," in *Proc. Int. Conf. Web Intell. Min. Semantics (WIMS)*, Sogndal, Norway, 2011, p. 52.
[19] D. Wang *et al.*, "RSH: A link-addition strategy for capacity enhancement in scale-free networks," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2110–2113, Dec. 2015.
[20] R. S. Burt, *Structural Holes: The Social Structure of Competition*. Cambridge, MA, USA: Harvard Univ. Press, 2009.
[21] M. Ouellet, M. Bouchard, and M. Hart, "Criminal collaboration and risk: The drivers of Al Qaeda's network structure before and after 9/11," *Soc. Netw.*, vol. 51, pp. 171–177, Oct. 2017.
[22] S. Agreste, S. Catanese, P. De Meo, E. Ferrara, and G. Fiumara, "Network structure and resilience of mafia syndicates," *Inf. Sci.*, vol. 351, pp. 30–47, Jul. 2016.

[23] I. Eusgeld, W. Kröger, G. Sansavini, M. Schlapfer, and E. Zio, "The role of network theory and object-oriented modeling within a framework for the vulnerability analysis of critical infrastructures," *Rel. Eng. Syst. Safety*, vol. 94, no. 5, pp. 954–963, 2009.

[24] A. H. Dekker and B. D. B. Colbert, "Network robustness and graph topology," in *Proc. Aust. Conf. Comput. Sci.*, Dunedin, New Zealand, 2004, pp. 359–368.

[25] F. Xie and D. Levinson, "Topological evolution of surface transportation networks," *Comput. Environ. Urban Syst.*, vol. 33, no. 3, pp. 211–223, 2009.

[26] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," *ACM Trans. Inf. Syst. Security*, vol. 10, no. 4, p. 1, 2008.

[27] Y.-Q. Zhang, X. Li, and A. V. Vasilakos, "Spectral analysis of epidemic thresholds of temporal networks," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2017.2743003.

[28] H. Tong *et al.*, "On the vulnerability of large graphs," in *Proc. IEEE Int. Conf. Data Min. (ICDM)*, Sydney, NSW, Australia, 2010, pp. 1091–1096.

[29] T. L. Long, T. Eliassi-Rad, and H. Tong, "MET: A fast algorithm for minimizing propagation in large graphs with small eigen-gaps," in *Proc. SIAM Int. Conf. Data Min. (SDM)*, Vancouver, BC, Canada, 2015, pp. 694–702.

[30] S. Saha, A. Adiga, B. A. Prakash, A. Vullikanti, and A. K. S. Kumar, "Approximation algorithms for reducing the spectral radius to control epidemic spread," in *Proc. SIAM Int. Conf. Data Min. (SDM)*, Vancouver, BC, Canada, 2015, pp. 568–576.

[31] E. Estrada, "Network robustness to targeted attacks. the interplay of expansibility and degree distribution," *Eur. Phys. J. B Condens. Matter Complex Syst.*, vol. 52, no. 4, pp. 563–574, 2006.

[32] F. D. Malliaros, V. Megalooikonomou, and C. Faloutsos, "Estimating robustness in large social graphs," *Knowl. Inf. Syst.*, vol. 45, no. 3, pp. 645–678, 2015.

[33] J. Wu, M. Barahona, Y.-J. Tan, and H.-Z. H. Deng, "Spectral measure of structural robustness in complex networks," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 6, pp. 1244–1252, Nov. 2011.

[34] H. Chan, L. Akoglu, and H. Tong, "Make it or break it: Manipulating robustness in large networks," in *Proc. SIAM Int. Conf. Data Min. (SDM)*, Philadelphia, PA, USA, 2014, pp. 325–333.

[35] H. Chan, S. Han, and L. Akoglu, "Where graph topology matters: The robust subgraph problem," in *Proc. SIAM Int. Conf. Data Min. (SDM)*, Vancouver, BC, Canada, 2015, pp. 10–18.

[36] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Math. J.*, vol. 23, no. 2, pp. 298–305, 1973.

[37] A. Jamakovic and P. Van Mieghem, "On the robustness of complex networks by using the algebraic connectivity," in *Proc. Int. Conf. Res. Netw.*, Prague, Czech Republic, 2008, pp. 183–194.

[38] D. J. Klein and M. Randić, "Resistance distance," *J. Math. Chem.*, vol. 12, no. 1, pp. 81–95, 1993.

[39] W. Ellens, F. M. Spieksma, P. Van Mieghem, A. Jamakovic, and R. E. Kooij, "Effective graph resistance," *Linear Algebra Appl.*, vol. 435, no. 10, pp. 2491–2506, 2011.

[40] I. Gutman and B. Mohar, "The quasi-Wiener and the Kirchhoff indices coincide," *J. Chem. Inf. Comput. Sci.*, vol. 36, no. 5, pp. 982–985, 1996.

[41] J.-B. Liu and J. Cao, "The resistance distances of electrical networks based on Laplacian generalized inverse," *Neurocomputing*, vol. 167, pp. 306–313, Nov. 2015.

[42] J.-B. Liu, X.-F. Pan, L. Yu, and D. Li, "Complete characterization of bicyclic graphs with minimal Kirchhoff index," *Discr. Appl. Math.*, vol. 200, pp. 95–107, Feb. 2016.

[43] J.-B. Liu, W.-R. Wang, Y.-M. Zhang, and X.-F. Pan, "On degree resistance distance of cacti," *Discr. Appl. Math.*, vol. 203, pp. 217–225, Apr. 2016.

[44] Q. Liu, J.-B. Liu, and J. Cao, "Further results on resistance distance and Kirchhoff index in electric networks," *Discr. Dyn. Nat. Soc.*, vol. 2016, p. 9, Jan. 2016, doi: 10.1155/2016/4682527.

[45] Q. Liu, J.-B. Liu, and J. Cao, "The Laplacian polynomial and Kirchhoff index of graphs based on R-graphs," *Neurocomputing*, vol. 177, pp. 441–446, Feb. 2016.

[46] H. Chan and L. Akoglu, "Optimizing network robustness by edge rewiring: A general framework," *Data Min. Knowl. Disc.*, vol. 30, no. 5, pp. 1395–1425, 2016.

[47] F. Arrigo and M. Benzi, "Updating and downdating techniques for optimizing network communicability," *SIAM J. Sci. Comput.*, vol. 38, no. 1, pp. B25–B49, 2016.

[48] S. Butler, *Eigenvalues and Structures of Graphs*, ProQuest, Ann Arbor, MI, USA, 2008.

[49] J.-B. Liu, J. Cao, T. Hayat, and F. E. Alsaadi, "Asymptotic behavior of Laplacian-energy-like invariant of the 3.6.24 lattice with various boundary conditions," *SpringerPlus*, vol. 5, no. 1, p. 1415, 2016.

[50] G. Yu, M. Ye, G. Cai, and J. Cao, "Signless Laplacian spectral conditions for Hamiltonicity of graphs," *J. Appl. Math.*, vol. 2014, p. 6, Jun. 2014, doi: 10.1155/2014/282053.

[51] J.-B. Liu, J. Cao, and J. Xie, "On the incidence energy of some toroidal lattices," in *Abstract and Applied Analysis*, vol. 2014. Akron, OH, USA: Hindawi Corporat., 2014, p. 6, doi: 10.1155/2014/568153.

[52] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[53] E. Cem, M. Toza, and K. Sarac, "Impact of sampling design in estimation of graph characteristics," in *Proc. Int. Perform. Comput. Commun. Conf. (IPCCC)*, San Diego, CA, USA, 2013, pp. 1–10.

[54] S. Datta and H. Kargupta, "Uniform data sampling from a peer-to-peer network," in *Proc. IEEE Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Toronto, ON, Canada, 2007, p. 50.

[55] C. Gkantsidis, M. Mihail, and A. Saberi, "Random walks in peer-to-peer networks," in *Proc. IEEE Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, Hong Kong, 2004, p. 130.

[56] V. Krishnamurthy *et al.*, "Sampling large Internet topologies for simulation purposes," *Comput. Netw.*, vol. 51, no. 15, pp. 4284–4302, 2007.

[57] C. Hübler, H. P. Kriegel, K. Borgwardt, and Z. Ghahramani, "Metropolis algorithms for representative subgraph sampling," in *Proc. 8th IEEE Int. Conf. Data Min. (ICDM)*, Pisa, Italy, 2008, pp. 283–292.

[58] J. Echeverria and S. Zhou, "Discovery, retrieval, and analysis of 'Star Wars' botnet in Twitter," *arXiv preprint arXiv:1701.02405*, 2017. [Online]. Available: https://arxiv.org/abs/1701.02405v3

[59] B. Krishnamurthy, P. Gill, and M. Arlitt, "A few chirps about Twitter," in *Proc. Int. Workshop Online Soc. Netw.*, Seattle, WA, USA, 2008, pp. 19–24.

[60] D. Stutzbach, R. Rejaie, N. Duffield, S. Sen, and W. Willinger, "On unbiased sampling for unstructured peer-to-peer networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 2, pp. 377–390, Apr. 2009.

[61] A. Rajadesingan, R. Zafarani, and H. Liu, "Sarcasm detection on Twitter: A behavioral modeling approach," in *Proc. Int. Conf. Web Search Data Min. (WSDM)*, Shanghai, China, 2015, pp. 97–106.

[62] M. Gabielkov and A. Legout, "The complete picture of the Twitter social graph," in *Proc. Student Workshop Colocated Int. Conf. Emerg. Netw. Exp. Technol. (CoNEXT)*, Nice, France, 2012, pp. 19–20.

[63] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, nos. 3–5, pp. 75–174, 2010.

[64] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 76, no. 3, 2007, Art. no. 036106.

[65] M. E. J. Newman, "Modularity and community structure in networks," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 23, pp. 8577–8582, 2006.

[66] U. Brandes *et al.*, *On Modularity-NP-Completeness and Beyond*. Karlsruhe, Germany: Univ. Karlsruhe, 2006.

[67] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 29, no. 4, pp. 251–262, 1999.

[68] L. A. Adamic and B. A. Huberman, "Power-law distribution of the World Wide Web," *Science*, vol. 287, no. 5461, p. 2115. 2000.

[69] M. E. J. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemp. phys.*, vol. 46, no. 5, pp. 323–351, 2005.

[70] Z. Li, J. Liu, and K. Wu, "A multiobjective evolutionary algorithm based on structural and attribute similarities for community detection in attributed networks," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2017.2720180.

Pasquale De Meo received the Ph.D. degree in systems engineering and computer science from the University of Calabria, Rende, Italy.

He is an Associate Professor of computer science with the Department of Ancient and Modern Civilizations with the University of Messina, Messina, Italy. He is a fellow with the Data Science Institute, Bournemouth University, Poole, U.K. He has been the Marie Curie Fellow with Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. His current research interests include social networks, recommender systems, and user profiling.

Fabrizio Messina received the Ph.D. degree in computer science from the Department of Mathematics and Informatics, University of Catania, Catania, Italy.

He is currently a Research Fellow with the Department of Mathematics and Informatics, University of Catania. He has authored over 60 papers in international conferences and journals. His current research interests include cloud and grid computing, trust and recommender systems, complex systems, and simulation systems.

Domenico Rosaci received the Ph.D. degree in electronic engineering from the University of Reggio Calabria, Reggio Calabria, Italy, in 1999.

He is an Associate Professor of computer engineering with the Mediterranea University of Reggio Calabria. He has published over 120 papers in outstanding international journals and conference proceedings. His current research interests include distributed artificial intelligence, multiagent systems, and trust and reputation in social communities.

Dr. Rosaci is an Associate Editor of the *Journal of Universal Computer Science* (Springer). He is a member of a number of conference PCs.

Giuseppe M. L. Sarné received the M.Sc. degree in engineering from the University of Calabria, Rende, Italy.

He is an Assistant Professor of computer science with the Mediterranea University of Reggio Calabria, Reggio Calabria, Italy. He has published over 110 papers in outstanding international journals and conference proceedings. His current research interests include distributed artificial intelligence, multiagent systems, and trust and reputation in social communities.

Mr. Sarné is an Associate Editor of *E-Commerce Research and Applications* (Elsevier). He is a member of a number of conference PCs.

Athanasios V. Vasilakos received the Ph.D. degree in computer science and engineering from the University of Patras, Patras, Greece.

He is a Professor with Innopolis University, Innopolis, Russia.

Dr. Vasilakos served or is serving as an Editor for several technical journals, such as the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, the IEEE TRANSACTIONS ON CLOUD COMPUTING, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON NANOBIOSCIENCE the IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, the *ACM Transactions on Autonomous and Adaptive Systems*, the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He is also the General Chair of the European Alliances for Innovation.