OPEN

Robustness surfaces of complex networks

Marc Manzano[1], Faryad Sahneh[2], Caterina Scoglio[2], Eusebi Calle[1] & Jose Luis Marzo[1,2]

[1]Department of Architecture and Computers Technology, University of Girona, Spain, [2]Department of Electrical and Computer Engineering, Kansas State University, USA.

Despite the robustness of complex networks has been extensively studied in the last decade, there still lacks a unifying framework able to embrace all the proposed metrics. In the literature there are two open issues related to this gap: (a) how to dimension several metrics to allow their summation and (b) how to weight each of the metrics. In this work we propose a solution for the two aforementioned problems by defining the R^*-value and introducing the concept of *robustness surface* (Ω). The rationale of our proposal is to make use of Principal Component Analysis (PCA). We firstly adjust to 1 the initial robustness of a network. Secondly, we find the most informative robustness metric under a specific failure scenario. Then, we repeat the process for several percentage of failures and different realizations of the failure process. Lastly, we join these values to form the robustness surface, which allows the visual assessment of network robustness variability. Results show that a network presents different robustness surfaces (i.e., dissimilar shapes) depending on the failure scenario and the set of metrics. In addition, the robustness surface allows the robustness of different networks to be compared.

The study of complex networks has attracted significant attention in the past decade. Critical infrastructures such as power grids, telecommunication networks or transportation networks, among others, are complex networks which are omnipresent and play a pivotal role in ensuring the smooth functioning of modern day living. These networks have to constantly deal with failures of their components, hence, any disruption of the service provided might have a considerable impact upon sizable proportions of the world's inhabitants. Thus, understanding not only the structure, but also the dynamics of such networks is of paramount importance.

Failures can be classified as being either random (i.e., accidental) or intentional (also referred to as targeted or deliberated)[1,2]. Accidental failures occur as a result of random actions on network elements (e.g., human-made errors or natural disasters). In contrast, in intentional attacks components are chosen according to some criterion in order to maximize the impact of the failures (e.g., a Denial-of-Service (DoS) attack). We define a *failure scenario* as the pair given by a specific type of failure (e.g., node or link) and a given attack strategy (e.g., random or intentional).

For network engineers and operators it is crucial to quantify the tolerance of a network to a given failure scenario. Robustness is defined as the ability of a network to maintain its total throughput under node or link removal[3,4].

Robustness metrics have been evolving since the advent of network science. Initially, several works studied the robustness of complex networks by considering a single graph metric: efficiency[5], average shortest-path length[6,7], diameter[8], clustering coefficient[6,9], node and link connectivity[10], heterogeneity[11], two-terminal reliability[12], assortativity[13], betweenness centrality[14], among others. Later on, new metrics were proposed in order to capture advanced characteristics (i.e., by means of spectral graph theory): symmetry ratio[15], algebraic connectivity[16] or spectral radius[17]. Furthermore, other works presented more contemporary metrics which were based on classical graph features. For instance, the authors of[18] studied the robustness in terms of flow diversity, a metric based on the shortest-path length. More recently, generic procedures to capture the robustness of a network for the whole spectrum of possible failures have been presented. Metrics such as elasticity[3] or endurance[2] quantify the robustness of a network according to a single throughput parameter. Trajanovski et al., have proposed a framework to evaluate the robustness of complex networks, which is based on the generic metric *R*-value[19]. From now on, we will use the conventions defined in Table 1. According to[20], the *R*-value is denoted by:

$$R = \sum_{k=1}^{n} s_k t_k \qquad (1)$$

where *s* and *t* are $n \times 1$ weight and graph metric vectors, respectively, and *n* is the number of robustness metrics. Thus, the *R*-value includes several graph metrics characterizing network robustness. However, there are two open issues related to the normalization of the *t* metrics:

Table 1 | Definition of the variables

Variable	Meaning
n	number of robustness metrics
R	R-value [20]
s	vector of weights (size $n \times 1$)
t	vector of metrics (size $n \times 1$)
m	failure configurations, i.e., different realizations of the failure process
R^*	R-value computed via Principal Components (PC)
t^0	vector of metrics without failures (size $n \times 1$)
R^*_{init}	initial R^*-value (without failures)
v	eigenvector PC (size $n \times 1$)
$\hat{v}$	normalized eigenvector PC (size $n \times 1$)
P	set of percentage of failures
p	percentage of failures ($p \in P$)
t^p	vector of metrics when $p\%$ of elements fail
R^*_p	R-value when $p\%$ of elements fail
A_p	$m \times n$ matrix, i.e., m values for each of the n metrics when $p\%$ of elements fail
ω_p	vector of R^*_p values (size $m \times 1$)
ω'_p	vector ω_p sorted in decreasing order
C_P	covariance matrix of A_p (size $n \times n$)
$\bar{C}$	average of the $|P|$ covariance matrices (size $n \times n$)
V	matrix containing n eigenvectors v
D	diagonal matrix with eigenvalues (size $n \times n$)
l	number of most relevant eigenvector
Ω	robustness surface, i.e., $|P|$ vectors ω'_p

1. How to unify the dimensionality of each robustness metric of vector t in order to legitimate their summation.
2. How to define the weight of each metric to optimally extract the most significant information.

In this work we propose a solution for the two aforementioned problems by defining the R^*-value and introducing the concept of *robustness surface* (Ω). The former extracts the most informative robustness metric for a failure scenario, while the latter allows network robustness variations of different networks to be visually assessed, regardless of the failure scenario.

Results

R^*-value. The rationale of our proposal is to make use of Principal Component Analysis (PCA) (see *Methods*). Given a set of robustness metrics t, we first define the initial robustness as follows:

$$R^*_{init} = \sum_{k=1}^{n} \hat{v}_k t^0_k = 1 \qquad (2)$$

where t^0 is the set of metrics when no failures occur, and $\hat{v}$ is a normalized eigenvector or Principal Component (PC). We obtain $\hat{v}$ from the procedure that computes the robustness surface (see following subsection and Eq. 4). The fact that $\hat{v}$ is normalized makes R^*_{init} equal to 1. Additionally, R^* can be computed when $p\%$ of elements fail as denoted next:

$$R^*_p = \sum_{k=1}^{n} \hat{v}_k t^p_k \qquad (3)$$

where t^p_k is the set of metrics computed when $p\%$ of failures occur. R^*_p takes values in the interval $[0, +\infty)$.

The difference between R^* and R (Eq. 1) is that in our proposal the principal component $\hat{v}$ gives dimension and non-arbitrary weights to each of the metrics. In addition, besides finding the most informative robustness metric, we adjust the initial robustness to 1, thus simplifying the comparison of network robustness variations when failures occur.

Robustness surface (Ω). The robustness surface allows the network performance variability for a given failure scenario to be visually assessed.

In fact, Ω is a matrix where the rows are the percentage of failures (P) and the columns are the distinct failure configurations (m). The list of percentage of failures P (e.g., $P = \{1\%, 2\%..100\%\}$) denotes the range of failures for which the robustness is evaluated. A *failure configuration* represents a realization of the failure process. The different failure configurations m depict the different subsets of elements that fail for a given percentage of failures, with each subset being distinct from one another. The robustness value in $\Omega[p][i]$, where $p \in \{1\%..|P|\%\}$ and $i \in \{1..m\}$, is given by R^*_p (Eq. 3).

To obtain the robustness surface of a network given a failure scenario (e.g., node and random), we define the following procedure:

1. Let A_p be an $m \times n$ matrix where $p \in \{1..|P|\%\}$ is the percentage of failure. The goal is to transform A_p into a smaller data set, i.e., a vector ω_p of size m, while preserving the most significant information. Therefore, we define ω_p as a vector of size $m \times 1$. ω_p contains the set of m values R^*_p computed when $p\%$ of elements fail.

2. To do so, we first compute the covariance matrix C_p of each matrix A_p. Then, we average the $|P|$ covariance matrices to obtain a unique matrix $\bar{C}$. This allows us to obtain a PC independent of p.

3. We calculate the eigenvectors V and the eigenvalues D of $\bar{C}$. At this point, the l most relevant eigenvectors of V are taken as the principal components for each matrix A_p (see *Methods* for further details). Hereafter we assume that $l = 1$, i.e., v is the eigenvector PC.

4. Then, we obtain $\hat{v}$ by normalizing v:

$$\hat{v}_j = \frac{v_j}{\sum_{k=1}^{n} t^0_k v_k} \qquad j \in \{1..n\} \qquad (4)$$

5. By multiplying the principal component $\hat{v}$ by each row of A_p we obtain a vector ω_p of size m. Each value of ω_p is, indeed, R^*_p. Next, by iterating this procedure for all matrices A_p, we obtain a set of $|P|$ vectors ω_p. Finally, we define ω'_p as a vector ω_p sorted in decreasing order. Consequently, the robustness surface is given by the following expression $\Omega = \{\omega'_{1\%}, \ldots, \omega'_{|P|\%}\}$.

Although different failure scenarios (e.g, link random and link by betweenness centrality) provide different $\hat{v}$, each of them satisfies Eq. 2 because $\hat{v}$ is normalized (as shown in Eq. 4).

Case study. Here, we illustrate the suitability of our proposal for evaluating the robustness when considering several metrics. To do so, we study two real critical infrastructures: the Spanish railway network (*sprailway*)[21] and the European power grid network (*europg*)[22].

We consider incremental and irreversible random and targeted attacks (e.g., betweenness centrality (BC) or node degree). Link and node failures are considered for the *sprailway* and *europg*, respectively, to show that the robustness surface allows us to compare network robustness independently from the failure scenario. Link failures are caused randomly and by link BC, whereas node failures are caused randomly, by node degree, the clustering coefficient and node BC. In both cases, $|P|$ is set to 70, i.e., from 1% to 70% of failures. The presented results are obtained for 500 and 100 runs (m) for random and targeted attacks, respectively. For each of the runs, a different realization of the failure process is considered, i.e., a distinct subset of elements that fail according to the failure scenario.

We consider the following metrics: the largest connected component (LCC), the degree of fragmentation as a function of the number of connected components (only applicable to link failures), the average nodal degree, the two-terminal reliability, the average clustering

2

Table 2 | Main network characteristics. The table displays, from left to right, topology name, number of nodes (N), number of links (L), average node degree ± *standard deviation* (StDev) (⟨k⟩), maximum degree (k_{max}), average shortest-path length ± StDev (⟨l⟩) and assortativity (r)

topology	N	L	⟨k⟩ ± StDev	k_{max}	⟨l⟩ ± StDev	r
sprailway	169	190	2.24 ± 1.09	8	10.49 ± 4.64	−0.269
europg	1,494	2,154	2.88 ± 1.75	13	18.88 ± 8.73	−0.119

coefficient, the average shortest-path length, the diameter, the average node BC, the average link BC and the algebraic connectivity. Therefore, link failures have $n = 10$ while node failures have $n = 9$.

Table 2 presents the main characteristics of the two networks considered. Although both networks have a different number of nodes and links, they show a similar average node degree ⟨k⟩, which is between 2 and 3. However, *europg* has a higher node maximum degree, what means that such a network is more vulnerable to targeted attacks. The average shortest-path length ⟨l⟩ depicts that *europg* is about two times wider than *sprailway*. Finally, both networks have a negative value of assortativity (r), which means that nodes of dissimilar degrees are connected to each other.

Numerical results. The results of our work are presented in Figures 1 and 2. The x-axes show the different failure configurations for which the *n* metrics have been computed. The y-axes depict the range of percentage of failures (from 1% to 70%). At each coordinate (x,y), i.e., for each percentage of failures and for each subset of elements that fail, the R_p^*-value is shown. In each figure the range of colors expresses variability, with dark blue and dark red being the two extremes of each failure scenario intervals. Since $R_{init}^* = 1$, i.e., the initial robustness is set to 1 regardless of the set of *n* chosen metrics, our results allow a visual assessment of the robustness variation with respect to the initial conditions. The further the value of R_p^* is with respect to R_{init}^*, the lower the performance of the network is. When R_p^* is close to 0, the performance is considered to be totally deteriorated. Moreover, it is possible to observe $R_p^* = 1$ when $p \geq 1\%$, and the LCC of the network has similar properties to the initial network (without failures).

Figure 1 presents the robustness surface Ω of *sprailway* in the case of random (a) and link BC failures (b). Interestingly, the random case provides a smooth surface, while the targeted case presents abrupt slopes. The latter is worth noting, because the presence of abrupt

slopes in the robustness surface means that there are network elements (in this case, links) that could be protected in order to improve the overall network robustness.

In the case of *europg*, Fig. 2 depicts four robustness surfaces under different node failure scenarios. Similar to *sprailway*, the random surface in Fig. 2a depicts a regular behavior. In addition, the targeted-based cases depict rough surfaces. While Figs. 2b and 2c depict that *europg* is not robust under node degree or node BC failure scenarios, Fig. 2d shows that the network keeps the initial robustness until more than 30% of the nodes have failed. This implies that *europg* is significantly more robust under failures by the clustering coefficient than by other targeted strategies.

For some failure configurations, it is worth noting that R_p^* might increase at some percentage of failures with respect to R_{init}^*, as observed in 10% or 20% of failures in Fig. 1a and in 20% or 30% in Fig. 2a, as well as in the targeted-based surfaces. This result should not be misleading because it totally depends on the set of metrics that are being considered for the study. For instance, while some metrics might decrease as the percentage of failures increases, others might alternate increments and decrements because they depend on the number and size of largest connected components (i.e., average shortest-path length, diameter, algebraic connectivity, etc.). Therefore, the suitability of the robustness surface remains intact, because the variability of the robustness can be assessed in any case.

Finally, to compare the robustness surfaces of both networks, and considering the different failure scenarios, we average the values of each ω'_p of Ω. Thus, for each network and failure scenario, we obtain $|P|$ $\bar{R}_p^*$-values. Fig. 3 depicts a summary of the results. Fig. 3a shows the curves of $\bar{R}_p^*$ of both networks from 1% to 70% of failures. To complement the results in Fig. 3a, the variance is presented in Fig. 3b. For instance, it can be observed that both random failure scenarios show similar behaviors, although for *europg* the top of the curve is around 24% of failures. Therefore, our approach allows us to com-
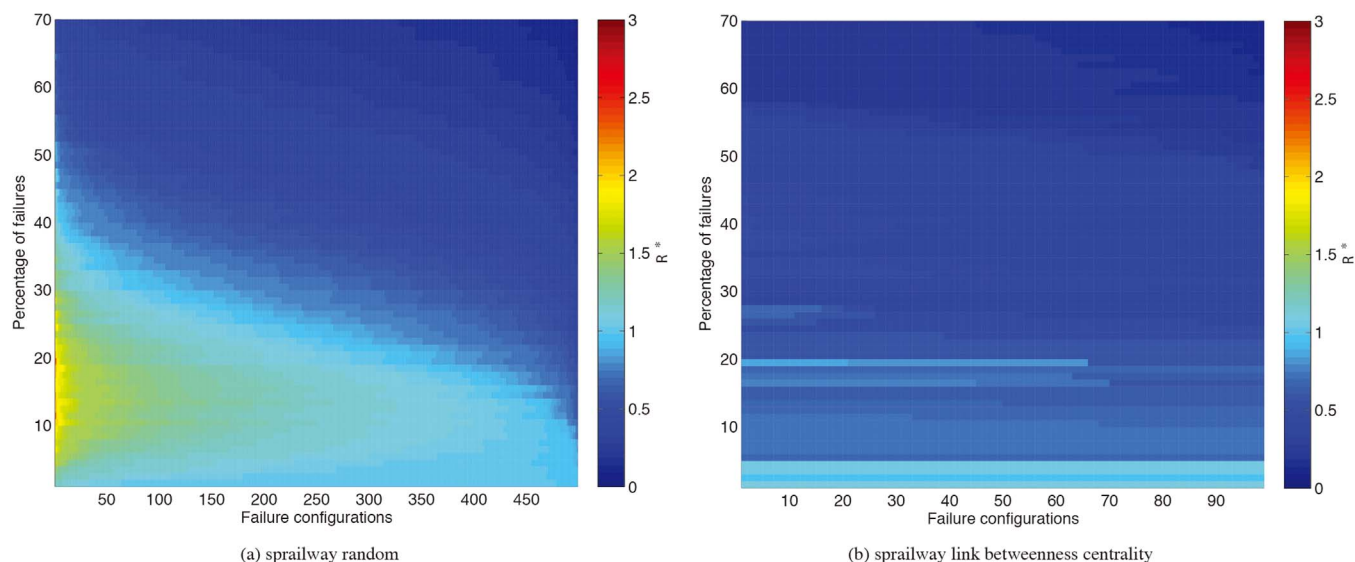

(a) sprailway random

(b) sprailway link betweenness centrality

Figure 1 | Robustness surface Ω of *sprailway* when causing links to fail randomly and by link BC.

(a) europg random

(b) europg node betweenness centrality

(c) europg node degree

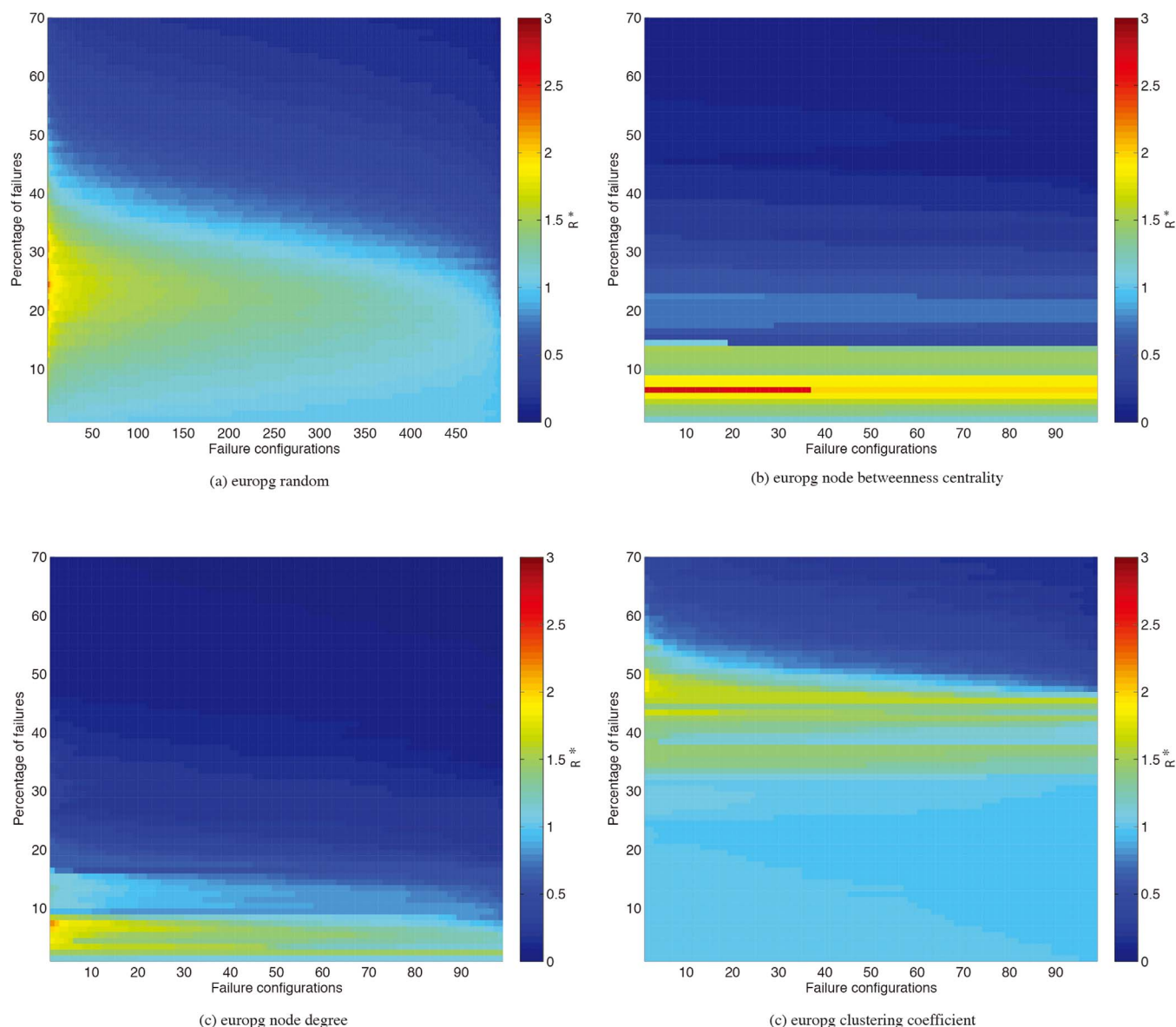

(c) europg clustering coefficient

Figure 2 | Robustness surface Ω of *europg* when causing nodes to fail randomly, by node BC, by node degree and by the clustering coefficient.

pare different networks, regardless of the failure scenarios. This comparison could be done numerically, for instance, by comparing the areas below the curves.

Discussion

In this work we present the R^*-value and the concept of *robustness surface* (Ω). The rationale of our proposal is to make use of Principal Component Analysis (PCA).

The R^*-value solves two open issues in the robustness of complex networks field. Our proposal extracts the most significant information from a set of robustness metrics. R^* is the first generic metric able to characterize the robustness of complex networks with a single value, while taking into account several robustness metrics.

The robustness surface Ω provides a framework to visually assess the network robustness variability. Moreover, it allows for the comparison of the robustness between different networks under distinct failure scenarios. To the best of our knowledge, it is the first method of its kind to allow the visual evaluation of the network robustness for a specific failure scenario, while at the same time considering several robustness metrics.

Robustness surfaces are designed as a visual monitoring tool. First, our approach is applicable to real-time monitoring of a network through a single value, when it is otherwise implemented according to multiplicity of correlated metrics with possible inherent redundancy. Second, Ω can be a pivotal part of a network robustness refinement process:

Step 1:   If the robustness surface presents abrupt slopes, then there are network elements (nodes or links) which are weaker than the rest, for a given failure scenario. These elements could be identified by means of traditional robustness metrics such as the betweenness centrality.

Step 2:   Enhance or protect the weak elements, for instance, by adding new links or applying immunization techniques.

Step 3:   Re-evaluate the robustness of a network and, instead of comparing a large number of robustness metrics, detect through visual inspection if the network robustness has been improved.

We believe that the contributions presented in this work will lay a firm foundation for future research on the robustness of complex networks.
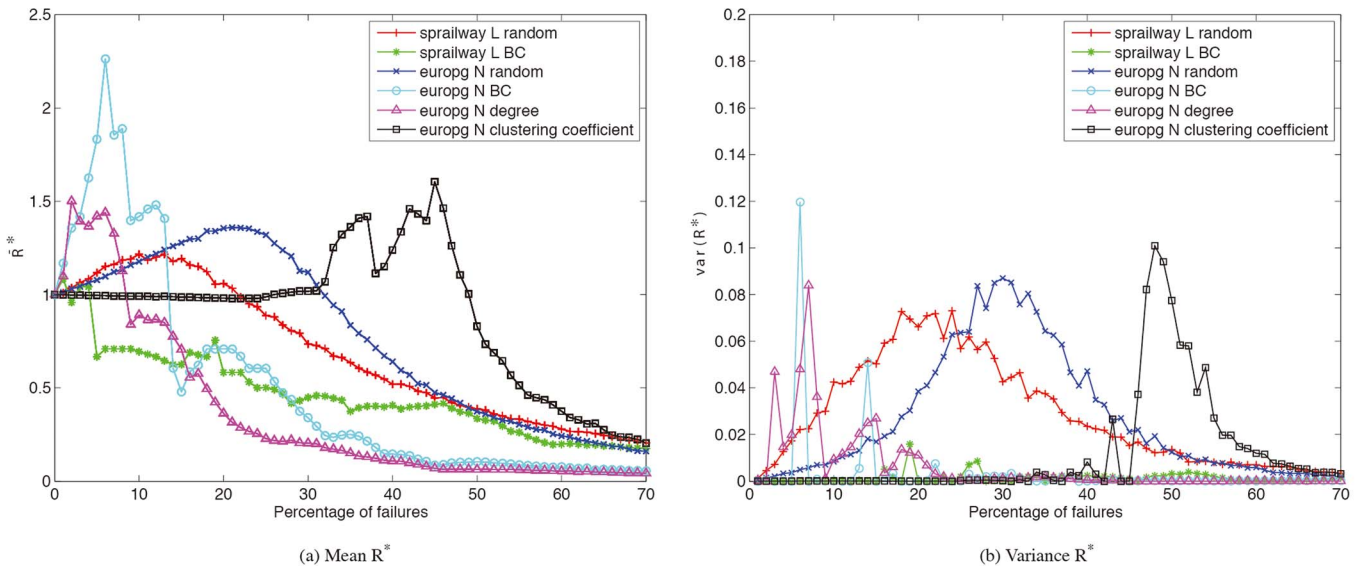
(a) Mean R*

(b) Variance R*

Figure 3 | R^* mean and variance of *sprailway* and *europg* under the different failure scenarios. As to the legend, L refers to link failures, whereas N refers to nodes.

To conclude, the R^*-value shows that there is no single and universal robustness metric for a network. Instead, the robustness varies according to the failure scenario and the metrics that are used to quantify the performance of the network.

For future work, we plan to study the stability of the robustness surfaces with respect to network size scaling.

Methods

Principal Component Analysis (PCA). PCA is a powerful tool to identify the most significant information in a data table representing observations described by several dependent variables, which can be inherently correlated. The goals of the PCA are to: (a) extract the main information of a data set and express it by means of new orthogonal variables called principal components; and (b) compress the size of the data set while preserving the most important information[23].

Let A be a data set of m observations of a vector-valued variable, i.e., $A \in R^{m \times n}$. We define C as the covariance matrix of A, which is denoted by:

$$C^{n \times n} = \left(c_{i,j}, c_{i,j} = cov\left(col_{A_i}, col_{A_j} \right) \right) \qquad (5)$$

where $i, j \in \{1..n\}$, and $cov\left(col_{A_i}, col_{A_j} \right)$ is the covariance function evaluating column i and column j.

PCA works with the spectrum of C. Let $v_i \in R^{n \times 1}\{i \in 1..n\}$ and $\lambda_i \in R$ be the eigenvectors and corresponding eigenvalues of the covariance matrix C, respectively. The matrix V with all v_i as columns represents the principal components, and provides an orthogonal transformation to the PC space. Furthermore, we denote D as a matrix with the eigenvalues in its diagonal.

Let $\tilde{V}$ be $n \times l$ matrix, which only contains the top l of the most important principal components (see *Methods: Most relevant principal components of A* for further details). Therefore, we can obtain the transformed data $\omega = A\tilde{V}$.

In our problem, each failure has a covariance matrix C_p, where p is the percentage of failure. We perform the PCA on $\bar{C} = C_p \delta p$, in order to obtain the PC independent of p.

Most relevant principal components of A. In order to choose the l most relevant principal components, matrices V (eigenvectors) and D (eigenvalues) must be column-sorted in decreasing order, according to the eigenvalues in the diagonal of D. The importance of each eigenvector is characterized by its energy quantum g. The eigenvalues represent the distribution of the energy of A among each of the eigenvectors. The energy quantum for the j_{th} eigenvector is the sum of the energy quantum across all eigenvalues from 1 to j:

$$g[j] = \sum_{k=i}^{j} D[k][k] \quad j = 1..n \quad . \qquad (6)$$

Let $\tilde{V}$ be an $n \times l$, where $l \leq n$ matrix that contains the most relevant eigenvectors. Then, the objective is to choose an l value as low as possible while preserving a reasonable high value of g on a percentage basis. For instance, we have chosen l so that g is above a certain threshold α:

$$min\left\{ l \in [1..n] : \frac{g[l]}{g[n]} \geq \alpha \right\} \qquad (7)$$

In this work we have considered $\alpha = 0.9$, from which we have obtained $l = 1$.

Simulation details. The computation of each metric has been done with PHISON[24]. The simulations were performed on a Linux system with a 16-core 64-bit Intel Xeon processor of 2 Ghz and 64 GB of RAM. The presented results are the average of 500 and 100 differently seeded simulation runs for random and targeted failures, respectively. The figures have been plotted by means of the *pcolor* function of MATLAB. In addition, the PCA has also been done with MATLAB.

1. Albert, R., Jeong, H. & Barabasi, A. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
2. Manzano, M., Calle, E., Torres-Padrosa, V., Segovia, J. & Harle, D. Endurance: A new robustness measure for complex networks under multiple failure scenarios. *Computer Netw.* **57**, 3641–3653 (2013).
3. Sydney, A., Scoglio, C., Youssef, M. & Schumm, P. Characterising the robustness of complex networks. *Int. J. Internet Technol. Secur. Syst.* **2**, 291–320 (2010).
4. Manzano, M., Calle, E. & Harle, D. Quantitative and qualitative network robustness analysis under different multiple failure scenarios. In *Proceedings of the 3rd International Workshop on Reliable Networks Design and Modeling (RNDM)*, 1–7 (2011).
5. Latora, V. & Marchiori, M. Efficient behavior of small-world networks. *Phys. Rev. Lett.* **87**, 198701 (2001).
6. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
7. Shannon, C. & Moore, D. The Spread of the Witty Worm. *IEEE Secur. Priv.* **2**, 46–50 (2004).
8. Albert, R., Jeong, H. & Barabási, A. Internet: Diameter of the world-wide web. *Nature* **401**, 130–131 (1999).
9. Bollobás, B. Mathematical results on scale-free random graphs. In *Handbook of Graphs and Networks*, 1–34 (Wiley-VCH, 2003).
10. Dekker, A. H. & Colbert, B. D. Network robustness and graph topology. In *Proceedings of the 27th Australasian conference on Computer science - Volume 26, ACSC*, Australian Computer Society, 359–368 (2004).
11. Dong, J. & Horvath, S. Understanding network concepts in modules. *BMC Syst. Biol.* **1**, 1–24 (2007).
12. Neumayer, S. & Modiano, E. Network reliability with geographically correlated failures. In *Proceedings of the 29th conference on Information Communications (INFOCOM)*, 1658–1666 (2010).
13. Mahadevan, P. *et al.* The internet AS-level topology: three data sources and one definitive metric. *SIGCOMM Comput. Commun. Rev.* **36**, 17–26 (2006).
14. Freeman, L. C. A set of measures of centrality based upon betweenness. *Sociometry* **40**, 35–41 (1977).
15. Dekker, A. H. & Colbert, B. D. The symmetry ratio of a network. In *Proceedings of the 2005 Australasian symposium on Theory of computing - Volume 41, CATS*, Australian Computer Society, 13–20 (2005).

16. Jamakovic, A. & Mieghem, P. V. On the Robustness of Complex Networks by Using the Algebraic Connectivity. In *Proceedings of Networking*, vol. 4982, 183–194 (2008).
17. Van Mieghem, P., Omic, J. & Kooij, R. Virus Spread in Networks. *IEEE/ACM Trans. Netw.* **17**, 1–14 (2009).
18. Rohrer, J. P. & Sterbenz, J. P. G. Predicting Topology Survivability using Path Diversity. In *Proceedings of the 3rd International Workshop on Reliable Networks Design and Modeling (RNDM)*, 1–7 (2011).
19. Trajanovski, S., Martín-Hernández, J., Winterbach, W. & Van Mieghem, P. Robustness envelopes of networks. *J. Complex Netw.* **1**, 44–62 (2013).
20. Van Mieghem, P. *et al.* A framework for computing topological network robustness. *Technical Report 20101218, Networks Architectures and Services, Delft University of Technology* (2010).
21. Roanes-Lozano, E. *et al.* Evolution of railway network flexibility: The Spanish broad gauge case. *Math. Comput. Simul.* **79**, 2317–2332 (2009).
22. Hutcheon, N. & Bialek, J. W. Updated and validated power flow model of the main continental european transmission network. In *Proceedings of the IEEE PowerTech 2013* (2013).
23. Abdi, H. & Williams, L. J. Principal component analysis. *Computation. Stat.* **2**, 433–459 (2010).
24. Manzano, M., Segovia, J., Calle, E. & Marzo, J. L. PHISON: Playground for High-level Simulations On Networks. In *Proceedings of the 2012 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)* (2012).

Author contributions

M.M. carried out the numerical experiments. M.M. and F.S. defined the two contributions of the manuscript. M.M., C.S. and J.L.M. prepared the figures. M.M., F.S., C.S., E.C. and J.L.M. wrote and reviewed the manuscript.

Additional information