



# CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

## Stability of a giant connected component in a complex network

Maksim Kitsak, Alexander A. Ganin, Daniel A. Eisenberg, Pavel L. Krapivsky, Dmitri Krioukov, David L. Alderson, and Igor Linkov

Phys. Rev. E **97**, 012309 — Published 22 January 2018

DOI: [10.1103/PhysRevE.97.012309](https://doi.org/10.1103/PhysRevE.97.012309)

# Stability of a Giant Connected Component in a Complex Network

Maksim Kitsak,<sup>1</sup> Alexander A. Ganin,<sup>2,3</sup> Daniel A. Eisenberg,<sup>4</sup> Pavel L. Krapivsky,<sup>5</sup> Dmitri Krioukov,<sup>6</sup> David L. Alderson,<sup>7</sup> and Igor Linkov<sup>8</sup>

<sup>1</sup>*Department of Physics, Northeastern University, 110 Forsyth Street,  
111 Dana Research Center, Boston, MA, 02115, USA.*

<sup>2</sup>*University of Virginia, Department of Systems and Information Engineering, Charlottesville, VA, 22904, USA*

<sup>3</sup>*U.S. Army Engineer Research and Development Center, Contractor, Concord, MA, 01742, USA*

<sup>4</sup>*School of Sustainable Engineering and the Built Environment,  
Arizona State University, Tempe, AZ, 85281, USA*

<sup>5</sup>*Department of Physics, Boston University, Boston, MA, 02215, USA*

<sup>6</sup>*Departments of Physics, Mathematics, and Electrical and Computer Engineering,  
Northeastern University, Boston, MA, 02115, USA*

<sup>7</sup>*Operations Research Department, Naval Postgraduate School, Monterey, CA, 93943, USA*

<sup>8</sup>*U.S. Army Engineer Research and Development Center,  
Environmental Laboratory, Concord, MA, 01742, USA*

(Dated: January 8, 2018)

We analyze the stability of the network’s giant connected component under impact of adverse events, which we model through the link percolation. Specifically, we quantify the extent to which the largest connected component of a network consists of the same nodes, regardless of the specific set of deactivated links. Our results are intuitive in the case of single-layered systems: the presence of large degree nodes in a single-layered network ensures both its robustness and stability. In contrast, we find that interdependent networks that are robust to adverse events have unstable connected components. Our results bring novel insights to the design of resilient network topologies and the reinforcement of existing networked systems.

Robustness and resilience of networked systems under the impact of adverse events have been extensively studied in network science for two decades, but the research has been primarily focused on computing the mean-field properties, such as the expected size of the system’s giant connected component (GCC) [1–3]. The inherent assumption is that only connected subnetworks retain their functionality, with the largest of these being most relevant to the overall performance of the system.

While the mean-field analysis of a system’s behavior is undoubtedly an important first step toward understanding its robustness, in most practical situations it is insufficient to know the expected size of the so-called ‘functional component.’ Rather, the location of the functional component within the network itself is important. It is especially true in the case of resilience where critical system function and its recovery is of prime importance [4, 5]. For example, in the case of a major natural disaster, such as flood or an earthquake, one needs to know infrastructure units and transportation routes that are likely to remain functional. The efficiency of immunization strategies depends on our knowledge of low and high-risk groups in social networks. Likewise, the success of a marketing campaign depends on the knowledge of the target audience.

Yet, apart from a handful of recent works aiming to understand individual node properties in percolation and epidemic processes [6–10], organization patterns of individual network’s functional components are poorly studied. One reason is that the random nature of adverse events coupled with the complexity of relevant networked systems often makes the prediction of functional sub-

graphs impossible. Indeed, consider two toy networks of the same size and average degree, which we repeatedly subject to adverse events of equal magnitude, modeled by link percolation, i.e., by deactivating a fraction of links selected uniformly at random, Fig. 1(a). Even though the expected GCC sizes of both networks are similar, the GCC of the first network is unstable and strongly depends on the set of deactivated links. In contrast, the GCC of the second network seems to be centered at the two largest degree nodes and only weakly depends on the set of deactivated links. Clearly, nodes of the first network are topologically identical and, thus, are equally likely to enter the GCC. The two central nodes of the second network, on the other hand, have much higher probability to enter the GCC compared to the remaining nodes, serving as anchors for its GCC.

Our work is motivated by two recent results in percolation theory [9, 10] analyzing heterogeneous network responses to different percolation realizations. In particular, Ref. [9] demonstrates a considerable variation of network’s connected component sizes and individual node probabilities to appear within the GCC. Concurrently, Ref. [10] introduces a framework to quantify GCC fluctuations by analyzing network responses to two random but possibly correlated percolation realizations. In our work we ask a related question: How stable is network’s GCC? In other words, we aim to quantify the extent to which the giant connected component of a network consists of the same nodes, regardless of the specific set of deactivated links during an adverse event. We analyze both single and multi-layer networks, finding that large GCC size does not necessarily imply GCC stability.

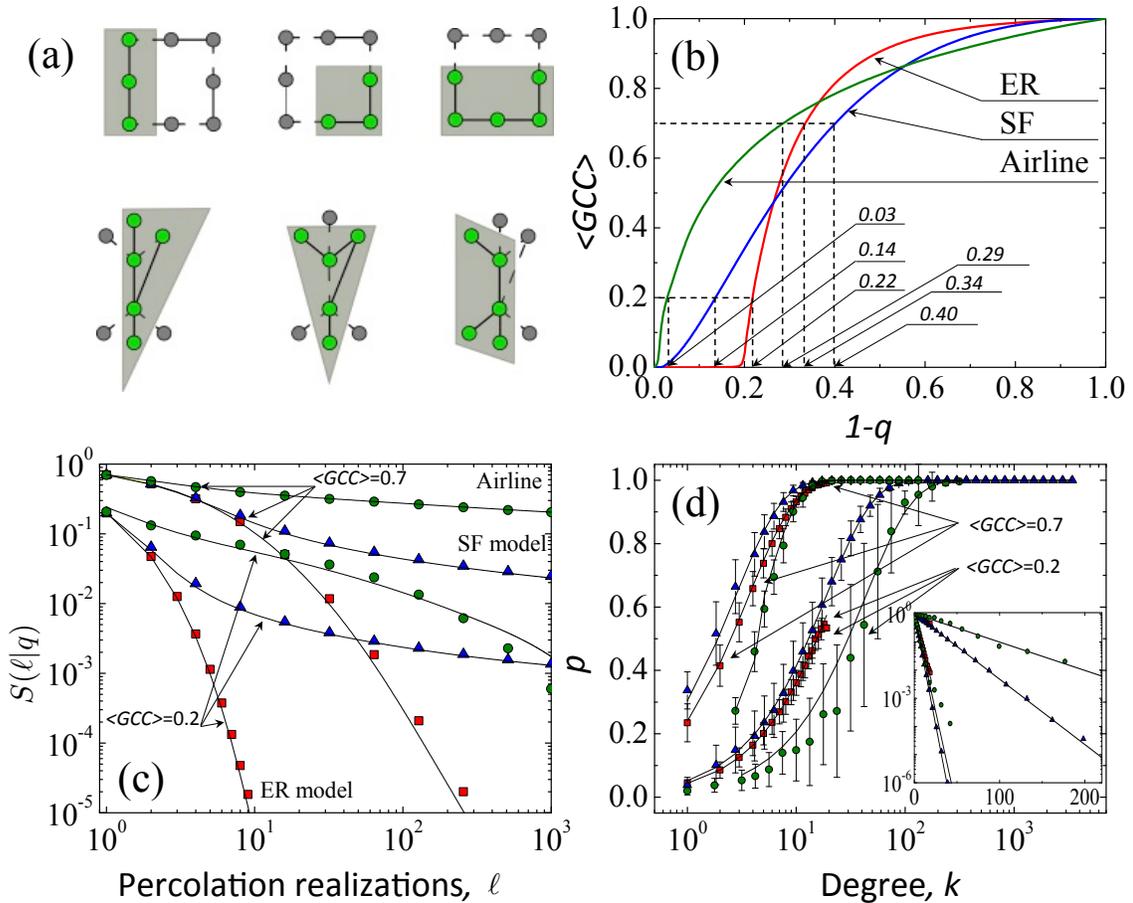


FIG. 1: (color online) (a) GCCs shown for two toy networks at various links percolation realizations. Note that the the GCC of the top network varies significantly depending on the set of removed links, while the GCC of the second network is more stable, and in all realizations includes the two central nodes. (b) The expected GCC size for an ER and a SF models, as well as the airline network as a function of the fraction of deactivated links  $q$ . Both model networks are of the same size,  $N = 2 \times 10^5$  and the same average degree  $\langle k \rangle = 5.1$ . The SF network is characterized by  $\gamma = 2.5$  and minimum degree  $k_{\min} = 2$ . Dashed lines show considered values of  $q$  that are chosen such that  $\langle GCC \rangle = 0.2$  and  $\langle GCC \rangle = 0.7$ . (c) Stability  $S(\ell|q)$  as a function of  $\ell$  for the (red squares) ER model, (blue triangles) SF model, and (green circles) the Airline network described in Appendix A. Solid lines correspond to analytical predictions by Eq. (4). (d) Probability for a node to appear within the GCC of a network as a function of its degree. The color scheme is that of panel (c). Data is binned logarithmically, and error bars display the standard deviation values. The inset shows the probability of the node *not* to appear within the GCC in the log-linear format with solid lines corresponding to analytical solutions given by Eqs. (2) and (3).

To simplify the exposition we model adverse events by link percolation [1, 11] and limit our consideration to random network models with prescribed degree sequences. We also assume that there exists at most one GCC per network. At the same time we note that our approach is extendable to other types of percolation and more complex network topologies, including the situations with multiple GCCs per network [12, 13]. To quantify GCC stability we subject the network of interest to  $\ell$  independent link percolation scenarios; each deactivating a random fraction  $q$  of network links. We then compute the sets of nodes  $\Omega_t(q)$  for each percolation realization  $t \in \{1, \dots, \ell\}$  and use them to define the GCC stability function as the fraction of nodes inside all sets  $\Omega_t(q)$ ,

namely

$$S(\ell|q) \equiv \frac{1}{N} \left\| \bigcap_{t=1}^{\ell} \Omega_t(q) \right\|, \quad (1)$$

where  $N$  is the network size and  $\|\Omega\|$  is the cardinality of set  $\Omega$ . In the case of GCCs consisting of random sets of nodes  $S(\ell|q)$  decreases exponentially as a function of  $\ell$ , while in the case of stable GCCs  $S(\ell|q)$  is expected to decrease slowly or remain constant. We first focus on GCC stability of the Erdős Rényi (ER) [14] and scale-free (SF) random network models. The former is fully prescribed by the number of nodes  $N$  and the constant connection probability  $p$  for every node pair. The latter is characterized by the scale-free (SF) degree distributions,  $p(k) \sim k^{-\gamma}$  and is generated by the Molloy-Reed algorithm [15]. Our results depicted in Fig. 1(b,c) indicate

that GCC stability in ER is significantly smaller than that in SF networks. Indeed,  $S(\ell|q)$  in the case of SF network models decays at significantly slower rates implying the existence of a stable sub-component, shared by all  $\Omega_t(q)$  sets. At the same time  $S(\ell|q)$  in ER network models seems to decrease exponentially, albeit at rates slower than expected for GCCs consisting of randomly selected nodes.

Our results for  $S(\ell|q)$  are consistent with the observation that individual node probabilities to belong to the GCC are strongly correlated with node's degree. As seen from Fig. 1(d), the probability for a node to belong to the GCC increases as a function of its degree and approaches 1 for nodes of sufficiently large degree. These nodes belong to the GCC with high probability and serve as its anchors in a network, effectively stabilizing its GCCs. In contrast, the absence of such large degree anchor nodes in ER networks leads to unstable GCCs, Fig. 1(d).

To quantify the observed GCC stability phenomena we employ the generating function formalism [16]. The starting point of our analysis is the mean-field solution for the individual node probability to enter the GCC [17]:

$$p = 1 - [u(q)]^k, \quad (2)$$

where  $u(q)$  can be regarded as the mean-field probability that a given end of randomly chosen link leads to a connected component of finite size [18]:

$$u = q + (1 - q)G_1(u), \quad (3)$$

and  $G_1(x)$  is the generating function for the outgoing edge degree distribution,  $G_1(x) \equiv \sum_{k=0}^{\infty} p_e(k)x^k$ , where  $p_e(k) = (k+1)p(k+1)/\langle k \rangle$ , and  $p(k)$  is the degree distribution [19].

Then, the probability that a node of degree  $k$  is present in the GCC independently in  $\ell$  percolation realizations is  $(1 - [u(q)]^k)^\ell$  and the expected stability of the GCC in  $\ell$  independent percolation realizations is given by

$$S(\ell|q) = \sum_k p(k) \left(1 - [u(q)]^k\right)^\ell. \quad (4)$$

To validate our theoretical results we solved Eq. (3) numerically for each combination of  $p(k)$  and percolation parameter  $q$  used in Fig. 1(b). We then used the obtained  $u$  values to superimpose the mean-field behaviors of individual node probabilities to enter the GCC and stability  $S(\ell|q)$  with the numerical results, arriving at the excellent agreement between the two, Fig. 1(c,d).

The differences between  $S(\ell|q)$  in SF and ER networks become more pronounced as expected GCC size decreases. Indeed, close to the criticality  $S(\ell|q)$  in ER decays exponentially:

$$\begin{aligned} S(\ell|q) &\approx \sum_{m=0}^{\ell} \binom{\ell}{m} (-1)^m \exp[-\langle k \rangle (1 - e^{-\epsilon \ell})] \\ &\sim [1 - u^{\langle k \rangle}]^\ell, \end{aligned} \quad (5)$$

where  $\epsilon \equiv 1 - u \ll 1$  and  $\langle k \rangle$  is the average degree in the ER network. In contrast,  $S(\ell|q)$  in SF networks

$$S(\ell|q) \propto (\epsilon k_0)^{\gamma-1} \int_{\epsilon k_0}^{\infty} [1 - e^{-x}]^\ell \frac{dx}{x^\gamma} \sim \left(\frac{\epsilon k_0}{\ln \ell}\right)^{\gamma-1} \quad (6)$$

for  $\ell \gg 1$ . Here  $\gamma > 2$  is the degree distribution exponent,  $p(k) \sim k^{-\gamma}$ , see Fig. 3 and Appendix B for derivation.

In the case of interdependent networks the deactivation of nodes or links in one layer leads to the deactivation of nodes in other layers that depend on it [20]. In this case, the probability of a node to belong to the GCC depends not only on its own connections but also on the connections of supporter nodes in other layers. In the following, we consider a simple case of an interdependent network, consisting of equally sized layers  $A$  and  $B$  with one-to-one node interdependence; every layer representing a random network with prescribed degree distribution [21]. Assuming no correlations in interdependencies between the two layers, the probability of a node to belong to the GCC is

$$p = \left(1 - [u_A(\mathbf{q})]^{k_A}\right) \left(1 - [u_B(\mathbf{q})]^{k_B}\right), \quad (7)$$

where  $k_A$  and  $k_B$  are the degrees of the two interdependent nodes in layers  $A$  and  $B$ , respectively,  $\mathbf{q} \equiv \{q_A, q_B\}$ ,  $q_A$  and  $q_B$  are the initial fractions of deactivated links in layers  $A$  and  $B$  respectively; while parameter  $u_A$  ( $u_B$ ) is the effective probability that a given end of a randomly chosen link in layer  $A$  ( $B$ ) leads to a mutual connected component of finite size. The mean-field parameters  $u_A$  and  $u_B$  can be found from the recursive approach of Ref. [21] and are given by the system of equations

$$1 - u_A = (1 - q_A) (1 - G_1^A(u_A)) (1 - G_0^B(u_B)), \quad (8)$$

$$1 - u_B = (1 - q_B) (1 - G_1^B(u_B)) (1 - G_0^A(u_A)), \quad (9)$$

where  $G_{\{0,1\}}^{\{A,B\}}(x)$  are the generation functions for the regular and the outgoing degree distributions of layers  $A$  and  $B$ , see Fig. 2(a,b) and Appendix C for derivation. We note that this result is in good agreement with numerical results, which we obtain for ER-ER, ER-SF and SF-SF models as well as the network of protein-protein interactions (PPI), and depict for SF-SF network in Fig. 2(a).

Since interdependence links between the layers are established randomly, the joint degree distribution of the network is  $p(k_A, k_B) = p_A(k_A)p_B(k_B)$ , where  $p_{A,B}(k)$  are degree distributions of layers  $A$  and  $B$  respectively. Then, the stability of the mutual connected component is simply the product of the contributions from the two layers:

$$S_{AB}(\ell|q) = S_A(\ell|q)S_B(\ell|q), \quad (10)$$

where  $S_A$  and  $S_B$  are the stability contributions of domains  $A$  and  $B$  respectively:  $S_A \equiv \sum_k p_A(k) \left(1 - [u_A(\mathbf{q})]^{k_A}\right)^\ell$ , and  $S_B$  is defined similarly. This result together with our results in Fig. 1

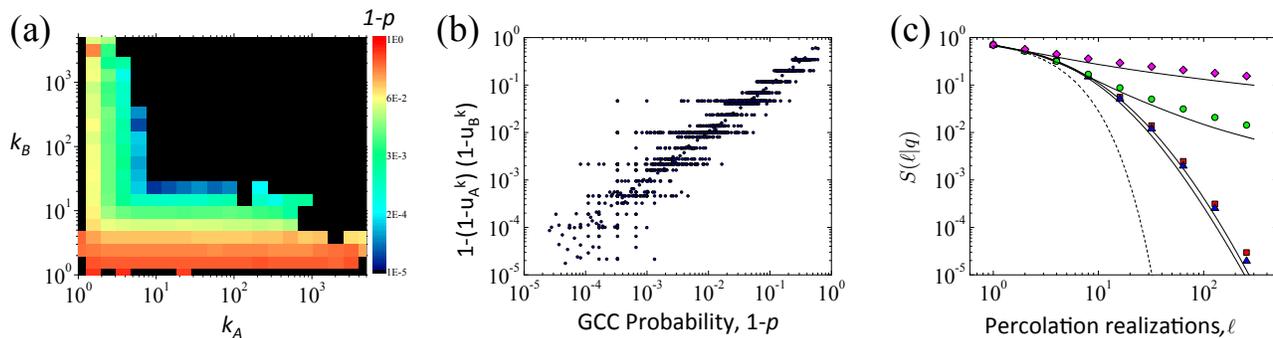


FIG. 2: (color online) (a) Individual node probability to appear within the mutual GCC of the interconnected network as a function of its degrees in the network layers. The interdependent network consists of two SF layers with random interdependence links. Both SF layers are generated using parameters reported in Fig. 1. Link percolation thresholds are chosen as  $q_A = 0.49$  and  $q_B = 0$ , corresponding to  $\langle GCC \rangle = 0.7$ . (b) Individual node probabilities *not* to appear within the GCC compared to the analytical estimate of Eq. (7). (c) Stability of the mutual GCC as a function of number of independent realizations  $\ell$  calculated for (green circles) SF-SF, (blue triangles) ER-SF and (red squares) ER-ER interdependent network models as well as (pink rhombi) the protein protein interaction network described in Appendix A. Link percolation thresholds are, respectively,  $q_A = 0.56$ ,  $q_A = 0.63$ ,  $q_A = 0.64$ , and  $q_A = 0.68$ ,  $q_B = 0$ , all corresponding to  $\langle GCC \rangle = 0.7$ . The dashed line corresponds to  $S(\ell|q)$  of mutual GCCs consisting of random sets of nodes. Solid lines correspond to analytic solutions of Eqs (8-10)

for single layered networks explains stable mutual GCCs in SF-SF networks and unstable GCCs in ER-ER and ER-SF networks that we observe in Fig. 2(c).

Taken together, our results indicate that networks with broader degree distributions are characterized by stable GCCs. Large degree nodes in these networks are part of the GCC with probabilities close to 1, and can be regarded as anchors keeping the GCC in place. At the same time, it is important to note that network robustness and GCC stability do not in general imply each other. We find that interdependent networks with broad degree distribution of their layers (e.g., SF-SF networks) have stable mutual GCCs. At the same time, however, SF-SF networks are known to be vulnerable to cascading failures [21]. Indeed, large degree nodes in interdependent networks are likely to depend on small degree nodes in other layers. As a result, failures of these small degree nodes immediately lead to failures of the large degree nodes that depend on them. In contrast, interdependent networks with narrow degree distributions, e.g., ER-ER networks, are less vulnerable due to a more uniform mixing of node degrees in their layers. The stability of connected components in interdependent networks, on the other hand, exclusively relies on the existence of anchor nodes, which according to Eq. (7) are large degree nodes dependent on other large degree nodes. While these anchor nodes are not frequent in SF-SF networks, they are non-existent in ER-ER networks, explaining why mutual GCCs are not stable in the latter.

Our findings open new avenues toward the design of efficient network reinforcement strategies. Indeed, building upon our results one can formulate the subgraph reinforcement problem as an optimization. Similar to the GCC stability, one can define the stability of any sub-

graph  $\Omega$  of interest:

$$S_{\Omega}(\ell|q) \equiv \left\| \bigcap_{t=1}^{\ell} \Omega_t(q) \right\| = \sum_{i \in \Omega} \left( 1 - [u(q)]^{k_i} \right)^{\ell}, \quad (11)$$

where  $u$  is the mean-field probability that a given end of a randomly chosen link leads to a cluster of finite size and is given by Eq. (3). Then the reinforcement strategy is defined in using extra resources, e.g. additional nodes or links to maximize the  $S_{\Omega}(\ell|q)$  objective function. It is also immediately clear that the optimal reinforcement strategies depend on the number of expected adverse events  $\ell$ : if the system is designed to withstand a single adverse event, the reinforcement strategy should be aimed at maximizing the expected GCC size. On the other hand, if the system is designed to withstand multiple adverse events, the investments should be made to further reinforce or create the anchor nodes.

The mean-field analysis offered here has important limitations. The generating function approach works well for locally tree-like networks, while real systems contain significant amount of short loops. Likewise, it is now understood that heterogeneous distributions in particular can have considerable diversity in their topologies [22] and that the hubs in these networks are not always centrally located. As a result, we do notice significant deviations for individual node probabilities in real networks, Fig. 1(d), that can't be explained by the mean-field approach.

Finally, within the engineering sciences there is a general understanding that the organization of many real-world networks is governed by processes that are not sufficiently characterized by stylized models of random networks and that these differences can have important implications for understanding robustness and vulnerability [23, 24]. There is a growing body of work in large-

scale optimization and game theory that leverages the specialized structure of these networks to discover specific vulnerabilities and to identify limited defensive investments that can maximally increase robustness and resilience (e.g., Ref. [4, 25]). Bridging the gap between the specific recommendations of these highly detailed models and the insights from more general models described here is an important goal for ongoing research.

We thank Shlomo Havlin and Pim van der Hoorn for

useful discussions and suggestions. This work was supported by the U.S. Defense Threat Reduction Agency. MK and DK were additionally supported by NSF grants CCF-1212778, CCF-1212778, and IIS-1741355, and ARO grants W911NF-16-1-0391 and W911NF-17-1-0491. The views and opinions expressed in this article are those of the individual authors and not those of the U.S. Army, or other sponsor organizations.

- 
- [1] D. Stauffer and A. Aharony, *Introduction to percolation theory* (CRC press, 1994).
- [2] S. N. Dorogovtsev, *Lectures on complex networks*, Vol. 24 (Oxford University Press Oxford, 2010).
- [3] R. Cohen and S. Havlin, *Complex networks: structure, robustness and function* (Cambridge university press, 2010).
- [4] A. A. Ganin, E. Massaro, A. Gutfraind, N. Steen, J. M. Keisler, A. Kott, R. Mangoubi, and I. Linkov, *Scientific Reports* **6**, 19540 (2016).
- [5] I. Linkov and J. M. Palma-Oliveira, eds., *Resilience and Risk*, NATO Science for Peace and Security Series C: Environmental Security (Springer Netherlands, Dordrecht, 2017).
- [6] K. E. Hamilton and L. P. Pryadko, *Phys. Rev. Lett.* **113**, 208701 (2014).
- [7] F. Morone and H. A. Makse, *Nature* **524**, 65 (2015).
- [8] T. Rogers, *Europhysics Letters* **109**, 28005 (2015).
- [9] R. Kuehn and T. Rogers, *Europhysics Letters* **118**, 68003 (2017).
- [10] G. Bianconi, *Phys. Rev. E* **96**, 012302 (2017).
- [11] B. Bollobás and O. Riordan, *Percolation* (Cambridge University Press, 2006).
- [12] E. Ben-Naim and P. L. Krapivsky, *Phys. Rev. E* **77**, 061132 (2008).
- [13] W. Chen and R. M. DSouza, *Phys. Rev. Lett.* **106**, 115701 (2011).
- [14] P. Erdős and A. Rényi, in *Publ. Math. Inst. Hung. Acad. Sci.* (Citeseer, 1960).
- [15] M. Molloy and B. Reed, *Random structures & algorithms* **6**, 161 (1995).
- [16] H. S. Wilf, *generatingfunctionology* (Elsevier, 2013).
- [17] M. Newman, *Networks: An Introduction* (Oxford University Press, 2010).
- [18] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000).
- [19] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
- [20] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, *Physics Reports* **544**, 1 (2014).
- [21] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, *Nature* **464**, 1025 (2010).
- [22] D. L. Alderson and L. Li, *Phys. Rev. E* **75**, 046102 (2007).
- [23] J. C. Doyle, D. L. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger, *Proc. Nat. Acad. Sci. USA* **102**, 14497 (2005).
- [24] M.-V. Florin and I. Linkov, eds., *IRGC resource guide on resilience* (EPFL International Risk Governance Center (IRGC), Lausanne, Switzerland, 2016).
- [25] D. L. Alderson, G. G. Brown, and W. M. Carlyle, *Risk Analysis* **35**, 562 (2015).
- [26] A. Chatr-aryamontri, B.-J. Breitzkreutz, R. Oughtred, L. Boucher, S. Heinicke, D. Chen, C. Stark, A. Breitzkreutz, N. Kolas, L. O'Donnell, *et al.*, *Nucleic acids research*, gku1204 (2014).

## Appendix A: Real networks

We test our GCC stability results on two real networks. The first real network is the complete US airport network in 2010, where nodes are US airports and an undirected link is established between two airports if there is a commercial flight between them. The network consists of  $N = 1,574$  nodes and  $E = 28,236$  links, and is publicly available from <https://openflights.org/> and <https://toreopsahl.com/datasets/>.

The second is the network of human protein-protein interactions (PPI) that we derived from the BioGRID database [26]. We represent the PPI dataset as a two-layer interdependent network, where layer  $A$  is formed by direct protein interactions, and links in layer  $B$  correspond to physical associations between proteins. In our analysis we focus on the mutual connected component of the PPI network that contains  $N = 11,365$  nodes in each layer with the average degrees of  $\langle k_A \rangle = 11.00$  and  $\langle k_B \rangle = 25.38$  respectively.

## Appendix B: GCC stability of single layer networks close to the criticality

Here we consider the asymptotic behavior of  $S(\ell|q)$  for three types of network models: regular, ER, and SF network models.

### 1. Regular network models

All nodes of a regular network model have the same degree  $k_0$ , resulting in  $p(k) = \delta(k, k_0)$ , where  $\delta(m, n)$  is the Kronecker delta. Then, GCC stability of a regular network is trivially  $S(\ell|q) = (1 - u(q)^{k_0})^\ell$  for all  $u(q)$  values.

## 2. ER network models

ER network models are characterized by the Poisson degree distribution,  $p(k) = e^{-\langle k \rangle} \langle k \rangle^k / k!$ , where  $\langle k \rangle = p(N-1)$  is the average degree. In order to obtain the asymptotic expression for  $S(\ell|q)$  we first expand  $(1 - u(q)^k)^\ell$  into the binomial sum. Then by reversing the summation order in Eq. (4) we obtain

$$S(\ell|q) = \sum_{m=0}^{\ell} \binom{\ell}{m} (-1)^m \exp(-\langle k \rangle (1 - u(q)^m)). \quad (\text{B1})$$

Close to the criticality  $\epsilon(q) \equiv 1 - u(q) \ll 1$  and therefore,  $1 - u(q)^m \simeq \epsilon(q)m$ . Plugging this into Eq. (B1) results in

$$S(\ell|q) \approx \left[1 - e^{-\epsilon(q)\langle k \rangle}\right]^\ell \approx \left[1 - u(q)^{\langle k \rangle}\right]^\ell. \quad (\text{B2})$$

## 3. SF network models

SF network models are characterized by power-law degree distributions  $p(k) \approx (\gamma-1)k_0^{\gamma-1}k^{-\gamma}$ , where  $k_0$  is the smallest degree and  $\gamma > 2$ .

To deduce the asymptotic behavior of  $S(\ell|q)$  we first approximate  $u(q)^k$  in Eq. (4) as  $e^{-k\epsilon(q)}$ , where  $\epsilon(q) \equiv 1 - u(q) \ll 1$ . Then, by approximating the summation in Eq. (4) with an integral we obtain:

$$S(\ell|q) \approx (\gamma-1) [\epsilon(q)k_0]^{\gamma-1} I_\ell(\gamma), \quad (\text{B3})$$

$$I_\ell(\gamma) \equiv \int_{\epsilon(q)k_0}^{\infty} x^{-\gamma} [1 - e^{-x}]^\ell dx. \quad (\text{B4})$$

Further, to deduce the asymptotic behavior of (B4) we note that the function  $[1 - e^{-x}]^\ell$  is monotonically increasing from 0 to 1 on the interval  $0 < x < \infty$ . It is useful to think about  $[1 - e^{-x}]^\ell$  as a smoothed version of the step function  $\theta(x - \ln \ell)$ . To appreciate the shift we center the function  $[1 - e^{-x}]^\ell$  at the position  $x_0$  where it equals  $e^{-1}$ . From  $[1 - e^{-x_0}]^\ell = e^{-1}$  we obtain  $e^{-x_0} = 1 - e^{-1/\ell}$  leading, when  $\ell \gg 1$ , to

$$x_0 = \ln \ell \quad (\text{B5})$$

and thereby explaining the  $\ln \ell$  shift. To appreciate the width of the quasi-step function  $[1 - e^{-x}]^\ell$  we define the left boundary through  $[1 - e^{-x_-}]^\ell = e^{-M}$  and the right boundary through  $[1 - e^{-x_+}]^\ell = 1 - e^{-M}$ . These definitions are reasonable if we choose  $M \gg 1$ . From these definitions

$$x_- = \ln \ell - \ln M, \quad x_+ = \ln \ell + M \quad (\text{B6})$$

Thus the quasi-step function  $[1 - e^{-x}]^\ell$  has the following properties:

1. It is essentially vanishes when  $0 < x < x_-$ .

2. It monotonically increases from 0 to 1 in the interval  $x_- < x < x_+$ .

3. It is essentially equal to 1 in the interval  $x > x_+$ .

We next establish the upper and lower bounds for the integral in (B4). The upper bound is obtained by replacing  $[1 - e^{-x}]^\ell$  by unity when  $x > x_-$  and  $e^{-M}$  when  $k_0\epsilon(q) < x < x_-$ ; the lower bound is obtained by replacing  $[1 - e^{-x}]^\ell$  by  $1 - e^{-M}$  when  $x > x_+$  and zero when  $0 < x < x_+$ . Since  $\int_{x_*}^{\infty} \frac{dx}{x^\gamma} = \frac{1}{\gamma-1} \frac{1}{x_*^{\gamma-1}}$ , the bounds are

$$\frac{1 - e^{-M}}{x_+^{\gamma-1}} \leq (\gamma-1)I_\ell(\gamma) \leq \left( \frac{1}{x_-^{\gamma-1}} + e^{-M} \frac{1}{[\epsilon(q)k_0]^{\gamma-1}} \right) \quad (\text{B7})$$

Since both  $x_\pm \rightarrow \ln \ell$  in the  $\{\ell, M\} \rightarrow \infty$  limit, such that  $1 \ll M \ll \ell$ ,  $I_\ell(\gamma) \sim \ln \ell^{1-\gamma}$  and

$$P(\ell|q) \sim \left[ \frac{k_0\epsilon(q)}{\ln \ell} \right]^{1-\gamma}. \quad (\text{B8})$$

## Appendix C: GCC stability of interdependent networks

Link percolation results for interdependent networks can be recovered from Ref. [21] by setting the fraction of active nodes to 1 and replacing degree generating functions with those for link percolated networks. Indeed, let us introduce the generating functions for degree distributions of layers  $A$  and  $B$  as

$$G_0^{\{A,B\}}(x) \equiv \sum_k p_{\{A,B\}}(k)x^k, \quad (\text{C1})$$

$$G_1^{\{A,B\}}(x) \equiv \sum_k \frac{(k+1)}{\langle k_{\{A,B\}} \rangle} p_{\{A,B\}}(k+1)x^k, \quad (\text{C2})$$

where  $p_A(k)$  and  $p_B(k)$  are the degree distributions of domains  $A$  and  $B$  respectively, while  $\langle k_A \rangle$  and  $\langle k_B \rangle$  are the corresponding average degree values.

Then the expected size of the mutual connected component  $\mu_\infty$  is given by

$$\mu_\infty = xy, \quad (\text{C3})$$

where the mean-field parameters  $x$  and  $y$  can be regarded as the effective fractions of nodes belonging to the GCC of layers  $A$  and  $B$  respectively, and are given by the system of four equations:

$$x = 1 - G_0^A(1 - y(1 - f_A)), \quad (\text{C4})$$

$$y = 1 - G_0^B(1 - x(1 - f_B)), \quad (\text{C5})$$

$$f_A = G_1^A(1 - y(1 - f_A)), \quad (\text{C6})$$

$$f_B = G_1^B(1 - x(1 - f_B)). \quad (\text{C7})$$

Here parameters  $f_A$  and  $f_B$  in analogy with the single layer case can be regarded as the effective probabilities

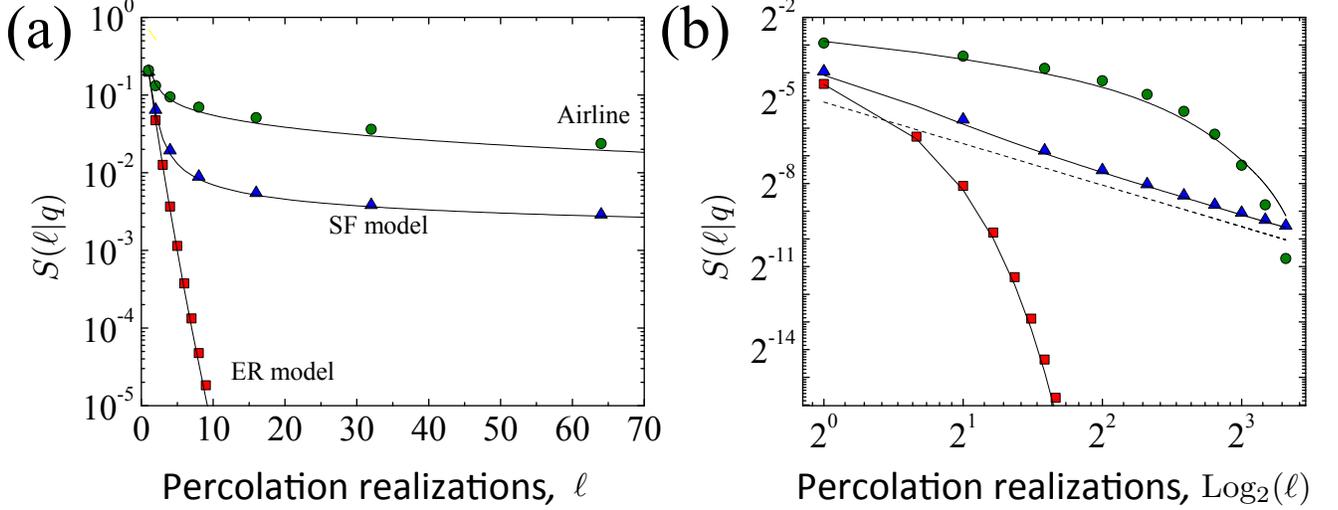


FIG. 3: (color online) Stability  $S(\ell|q)$  as a function of (a)  $\ell$  and (b)  $\text{Log}_2(\ell)$  shown at  $\langle GCC \rangle = 0.2$  for the (red squares) ER model, (blue triangles) SF model, and (green circles) the Airline network. All parameters are those of Fig. 1(b,c) in the main text. The dashed line in panel (b) corresponds to the asymptotic regime of  $S(\ell|q) \sim [\text{Ln } \ell]^{-1.5}$  prescribed by Eq. (6).

that a given end of a randomly chosen link in layer  $A$  or  $B$ , leads to a connected component of a finite size.

By making use of Eqs. (C4) and (C5) one can rewrite the expression for the mutual connected component as

$$\mu_\infty = (1 - G_0^A(1 - y(1 - f_A))) (1 - G_0^B(1 - x(1 - f_B))), \quad (\text{C8})$$

from which one can immediately extract the individual node probability to appear within the mutual GCC:

$$p = (1 - u_A^{k_A}) (1 - u_B^{k_B}), \quad (\text{C9})$$

where  $k_A$  and  $k_B$  are the degrees of the two interdependent nodes in layers  $A$  and  $B$ , and the mean-field parameters  $u_A$  and  $u_B$  are given by

$$u_A = 1 - y(1 - f_A), \quad (\text{C10})$$

$$u_B = 1 - x(1 - f_B). \quad (\text{C11})$$

The same equations hold in the link percolation case, except the original generating functions  $G_{\{0,1\}}^{\{A,B\}}(x)$  need to be replaced with those for percolated layers:

$$\tilde{G}_{\{0,1\}}^A(x) = G_{\{0,1\}}^A(1 - (1 - q_A)(1 - x)), \quad (\text{C12})$$

$$\tilde{G}_{\{0,1\}}^B(x) = G_{\{0,1\}}^B(1 - (1 - q_B)(1 - x)), \quad (\text{C13})$$

where  $q_A$  and  $q_B$  are the fractions of initially deactivated links in layers  $A$  and  $B$  respectively.

After a series of straightforward substitutions and simplifications the final result reads

$$p = (1 - [u_A(\mathbf{q})]^{k_A}) (1 - [u_B(\mathbf{q})]^{k_B}), \quad (\text{C14})$$

where

$$1 - u_A = (1 - q_A) (1 - G_1^A(u_A)) (1 - G_0^B(u_B)),$$

$$1 - u_B = (1 - q_B) (1 - G_1^B(u_B)) (1 - G_0^A(u_A)),$$

reproducing Eqs. (8) and (9) in the main text.