Robustness and modular structure in networks

James P. Bagrow
j.bagrow@neu.edu
Center for Complex Network Research,
Northeastern University, Boston MA, USA
Center for Cancer Systems Biology,
Dana-Farber Cancer Institute, Boston MA, USA

Sune Lehmann
DTU Informatics,
Technical University of Denmark, Kgs Lyngby, Denmark
College of Computer and Information Science,
Northeastern University, Boston, MA, USA

Yong-Yeol Ahn
yongyeol@gmail.com
Center for Complex Network Research,
Northeastern University, Boston MA, USA
Center for Cancer Systems Biology,
Dana-Farber Cancer Institute, Boston MA, USA

February 24, 2011

Abstract

Many complex systems, from power grids and the internet, to the brain and society, can be modeled using modular networks. Modules, densely interconnected groups of elements, often overlap due to elements that belong to multiple modules. The elements and modules of these networks perform individual and collective tasks such as generating and consuming electrical load, transmitting data, or executing parallelized computations. We study the robustness of these systems to the failure of random elements. We show that it is possible for the modules themselves to become isolated or uncoupled (non-overlapping) well before the network falls apart. When modular organization is critical to overall functionality, networks may be far more vulnerable than expected.

Complex networks have recently attracted much interest due to their prevalence in nature and our daily lives [3, 21]. A critical network property is its resilience or robustness to random breakdown and failure [4, 11, 9, 12], typically studied as a percolation problem [27, 1, 10, 24], or cascading failures [16, 8, 23]. Meanwhile, most networks are modular [14, 20], comprised of small, densely connected groups of nodes. The modules often overlap, with elements belonging to multiple modules [22, 2]. Existing work on robustness has not considered the role of modular structure.

Consider a system of interacting elements representing computers, power generators, neurons, etc. These elements perform tasks sufficiently complex that they must work together in densely interconnected modules. These tasks may be parallelized computations, protein biosynthesis, or higher-order neurological functions such as visual processing or speech production. Elements are required to communicate between modules, so that modules are coupled or overlapping, and the system functions properly only when modules can communicate. We ask how these networks respond when a random fraction of elements fail: do the modules become uncoupled before the network loses global connectivity? Random failures provide a toy model of, e.g., a traumatic brain injury or degenerative disease. If enough elements fail, the modules can no longer communicate (higher brain functions are lost) even though the network may remain connected (simpler autonomic responses persist). Likewise, an individual module may fail if too many of its member elements cease to function.

Modular structure can be represented as a bipartite network (Fig. 1a) [18, 19] characterized by two degree distributions, r_m and s_n, governing the fraction of elements that belong to m modules and the fraction of modules that contain n elements, respectively. The average number of modules per element is $\mu \equiv \sum_m m r_m$ and the average number of elements per module is $\nu \equiv \sum_n n s_n$. We derive two networks from the bipartite graph by projecting onto either the elements or the modules: One is the network between elements, while the other is a network where each node represents a module and two modules are linked if they share at least one element. The giant component in the element network disappears when the network loses global connectivity; in the module network it vanishes when the modules become uncoupled (non-overlapping). Before projection elements fail with probability $1-p$ and are removed from the network. Meanwhile, a module is unable to complete its collective task if fewer than a critical fraction f_c of its original elements remain. These failed modules are removed from the module network but any surviving member elements are not removed from the element network. See Fig. 1b.

We wish to determine $S(p)$, the fraction of remaining nodes within the giant component as a function of p, for both the element and module networks. We define four generating functions [18, 19]:

$$\begin{aligned} f_0(z) &= \sum_{m=0}^{\infty} r_m z^m, \quad f_1(z) = \frac{1}{\mu}\sum_{m=0}^{\infty} m r_m z^{m-1}, \\ g_0(z) &= \sum_{n=0}^{\infty} s_n z^n, \quad\ \ g_1(z) = \frac{1}{\nu}\sum_{n=0}^{\infty} n s_n z^{n-1}. \end{aligned} \tag{1}$$

These functions generate the probabilities for (f_0) a randomly chosen element to belong

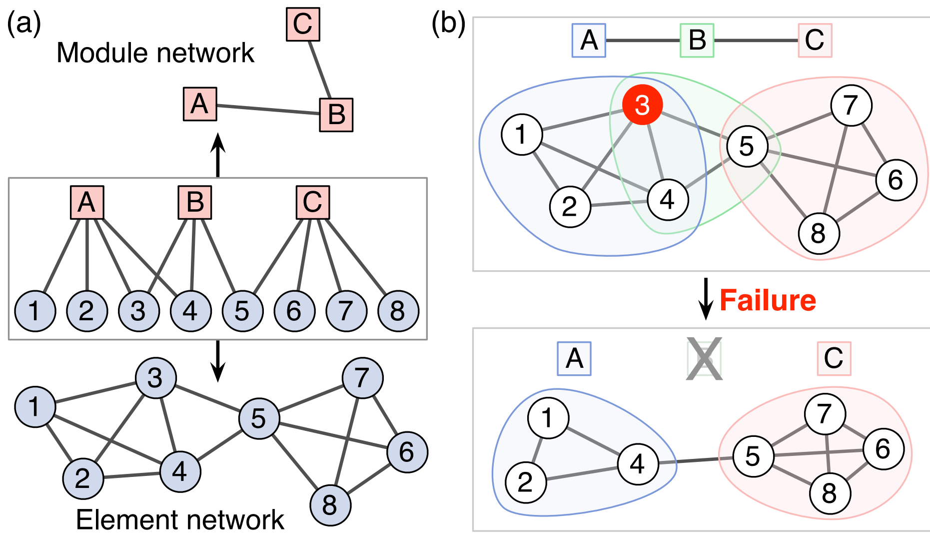

Figure 1: The modular network representation [18, 19]. (a) We obtain two networks by projecting onto elements or modules. (b) The failure of element 3 induces the failure of module B, uncoupling the remaining modules, even though the network itself remains connected.

to m modules, (f_1) a random element within a randomly chosen module to belong to m other modules, (g_0) a random module to contain n elements, and (g_1) a random module of a randomly chosen element to contain n other elements.

1 Element network

Consider a randomly chosen element A that belongs to a group of size n. Let $P(k|n)$ be the probability that A still belongs to a connected cluster of k nodes (including itself) in this group after failures occur:

$$P(k|n) = \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k}. \tag{2}$$

The generating function for the number of other elements connected to A within this group is

$$h_n(z) = \sum_{k=1}^{n} P(k|n) z^{k-1} = (zp + 1 - p)^{n-1}. \tag{3}$$

Averaging over module size:

$$h(z) = \frac{1}{\nu} \sum_{n=0}^{\infty} n s_n h_n(z) = g_1(zp + 1 - p). \tag{4}$$

The total number of elements that A is connected to, from all modules it belongs to, is then generated by

$$G_0(z) = f_0(h(z)). \tag{5}$$

Likewise, the total number of elements that a randomly chosen neighbor of A is connected to is generated by

$$G_1(z) = f_1(h(z)). \tag{6}$$

Before determining S, we first identify the critical point p_c where the giant component emerges. This happens when the expected number of elements two steps away from a random element exceeds the number one step away, or

$$\partial_z G_0(G_1(z))\big|_{z=1} - \partial_z G_0(z)\big|_{z=1} > 0. \tag{7}$$

Substituting Eqs. (5) and (6) gives $f_0'(1)h'(1)[f_1'(1)h'(1) - 1] > 0$ or $f_1'(1)h'(1) > 1$. Finally, the condition for a giant component to exist, since $h'(1) = pg_1'(1)$, is

$$pf_1'(1)g_1'(1) > 1. \tag{8}$$

For the uniform case, $r_m = \delta(m, \mu)$ and $s_n = \delta(n, \nu)$, this gives $p(\mu - 1)(\nu - 1) > 1$. If $\mu = 3$ and $\nu = 3$, then the transition occurs at $p_c = 1/4$.

To find S, consider the probability u for element A to not belong to the giant component. A is not a member of the giant component only if all of A's neighbors are also not members, so u satisfies the self-consistency condition $u = G_1(u)$. The size of the giant component is then $S = 1 - G_0(u)$.

2 Module network

Consider a random module C and then a random member element A. Let $Q(\ell|m)$ be the probability that C is connected to ℓ modules, including itself, through element A, who was originally connected to m modules including C:

$$Q(\ell|m) = \binom{m-1}{\ell-1} q_1^{\ell-1} (1-q_1)^{m-\ell}, \tag{9}$$

where

$$q_1 = \frac{1}{\nu}\sum_{n=0}^{\infty} n s_n \sum_{i=x}^{n} \binom{n-1}{i-1} p^{i-1}(1-p)^{n-i}. \tag{10}$$

(Notice that $q_1 = 1$ when $x(n) \equiv \lceil n f_c \rceil = 1$ for all n.) The generating function j_m for the number of modules that C is connected to, including itself, through A is

$$j_m(z) = \sum_{\ell=1}^{m} Q(\ell|m) z^{\ell-1} = (zq_1 + 1 - q_1)^{m-1}. \tag{11}$$

Once again, averaging j_m over memberships gives

$$j(z) = \frac{1}{\mu}\sum_{m=0}^{\infty} m r_m j_m(z) = f_1(zq_1 + 1 - q_1). \tag{12}$$

The total number of modules that C is connected to is *not* generated by $g_0(j(z))$ but by $\tilde{g}_0(j(z))$, where the $\tilde{g}_i$ are the generating functions for module size after elements fail:

$$\tilde{g}_0(z) = \sum_{n=0}^{\infty} \tilde{s}_n z^n, \qquad \tilde{g}_1(z) = \frac{\sum_{n=0}^{\infty} n\tilde{s}_n z^{n-1}}{\sum_{n=0}^{\infty} n\tilde{s}_n}. \tag{13}$$

The probability $\tilde{s}_k$ to have k member elements remaining in a module after percolation is given by

$$\tilde{s}_k = \frac{\sum_n \binom{n}{k} p^k (1-p)^{n-k} s_n}{\sum_n \sum_{k'=x}^{n} \binom{n}{k'} p^{k'} (1-p)^{n-k'} s_n}. \tag{14}$$

The denominator is necessary for normalization since we cannot observe modules with fewer than $\lceil n f_\mathrm{c} \rceil$ members. Notice that $\tilde{s}_n = s_n$ when $s_n = \delta(n, \nu)$ and $\lceil n f_\mathrm{c} \rceil = n = \nu$.

Finally, the total number of modules connected to C through any member elements is generated by $F_0(z) = \tilde{g}_0(j(z))$ and the total number of modules connected to a random neighbor of C is generated by $F_1(z) = \tilde{g}_1(j(z))$. As before, the module network has a giant component when $\partial_z F_0(F_1(z))|_{z=1} - \partial_z F_0(z)|_{z=1} > 0$ and $S = 1 - F_0(u) = 1 - \tilde{g}_0(j(u))$, where u satisfies $u = F_1(u) = \tilde{g}_1(j(u))$.

For the uniform case with $\mu = 3$, $\nu = 3$, and $f_\mathrm{c} > 2/3$, the critical point for the module network is $p_\mathrm{c} = 1/2$, a considerably higher threshold than for the element network ($p_\mathrm{c} = 1/4$). In Fig. 2 we show S for $\mu = 3$ and $\nu = 6$. The "robustness gap" between the element and module networks widens as the module failure cutoff increases, covering a significant range of p for the larger values of f_c.

Of particular interest are scale-free networks [7, 28, 21]. Here we take $r_m = \delta(m, \mu)$ as before, but now $s_n \sim n^{-\lambda}$, with $\lambda \geq 2$ [1]. It is known that scale-free networks are robust to random failures when $2 < \lambda < 3$ (meaning that $p_\mathrm{c} \to 0$). However, this result also requires that the maximum value K of the degree distribution be large ($K \gg 1$) [11]. Indeed, as we lower λ, we discover that, while we increase the robustness of the elements, we actually *decrease* the robustness of the modules (Fig. 3). For modular networks, it may not be feasible to build extremely large modules. Interestingly, enforcing on s_n a maximum module size cutoff $N = \max\{n \mid s_n > 0\}$ only improves element robustness.

3 Empirical results

We study failures in multiple social, biological, and informational real-world datasets (see App. A). Unlike the model, we do not know the modules in advance, so we estimate them with an overlapping community algorithm [2] (a second method [22] displays similar behavior). These networks tend to be smaller than those previously discussed, introducing finite-size effects that mask the behavior of S. To overcome this, we instead use S', the fraction of *original* nodes that remain in the giant component (see App. B). As shown in Fig. 4, the

[1] The degree distribution after projection remains scale-free (with the same exponent), although the maximum degree may increase.

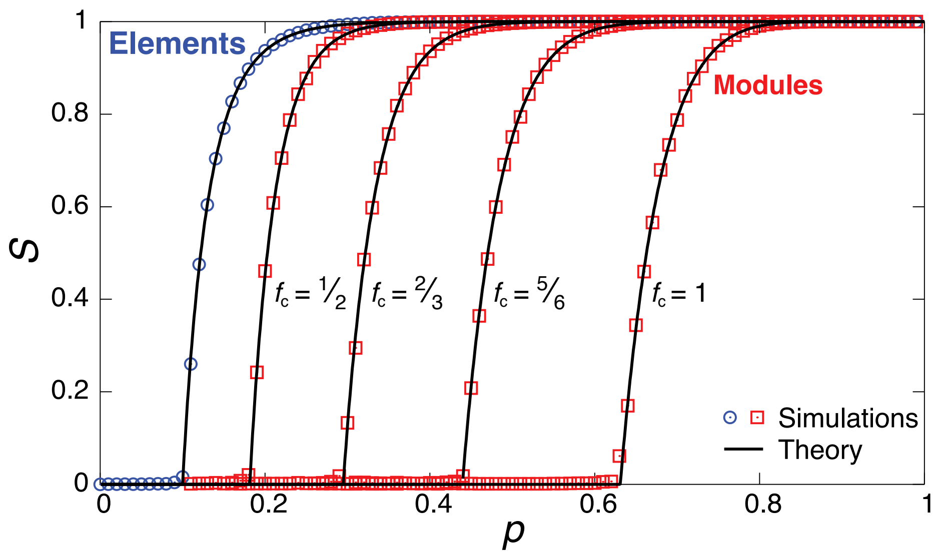

Figure 2: The size of the giant component S for $r_m = \delta(m, \mu), s_n = \delta(n, \nu)$, with $\mu = 3$ and $\nu = 6$. Theory and simulations confirm that the network undergoes a transition from coupled to non-overlapping modules well before it loses global connectivity. Symbols represent the element (□) and module (○) networks.

modules fall apart more easily than the elements, qualitatively matching our model across a broad range of networks.

4 Conclusions

There are a number of interesting avenues for further work. We considered the simplest case of random failures but extensions to purposeful attacks (failure proportional to n or m) are also important. Likewise, the model we use assumes that all links exist within modules, but links between modules are certainly possible. These additional links can only enhance the robustness of the element network, but will not improve the module network, so that the robustness gap may be significantly increased. Beyond structural characteristics of these modular networks it is important to understand the effect of failures and modular structure on critical phenomena such as synchronization [6, 5], contact processes [15, 26], cascades [16, 8, 23] or other dynamics [13].

Finally, this work can also help us to understand how empirical networks are affected by *missing data*, of critical importance when studying communities. Here p is the probability that a network element is successfully captured by an experiment, such as a high-throughput biological assay or web crawler. The robustness gap can explain how non-overlapping community methods may succeed in networks where overlap is expected: the network is sampled down to the intermediate regime where nodes are connected but modules are uncoupled.

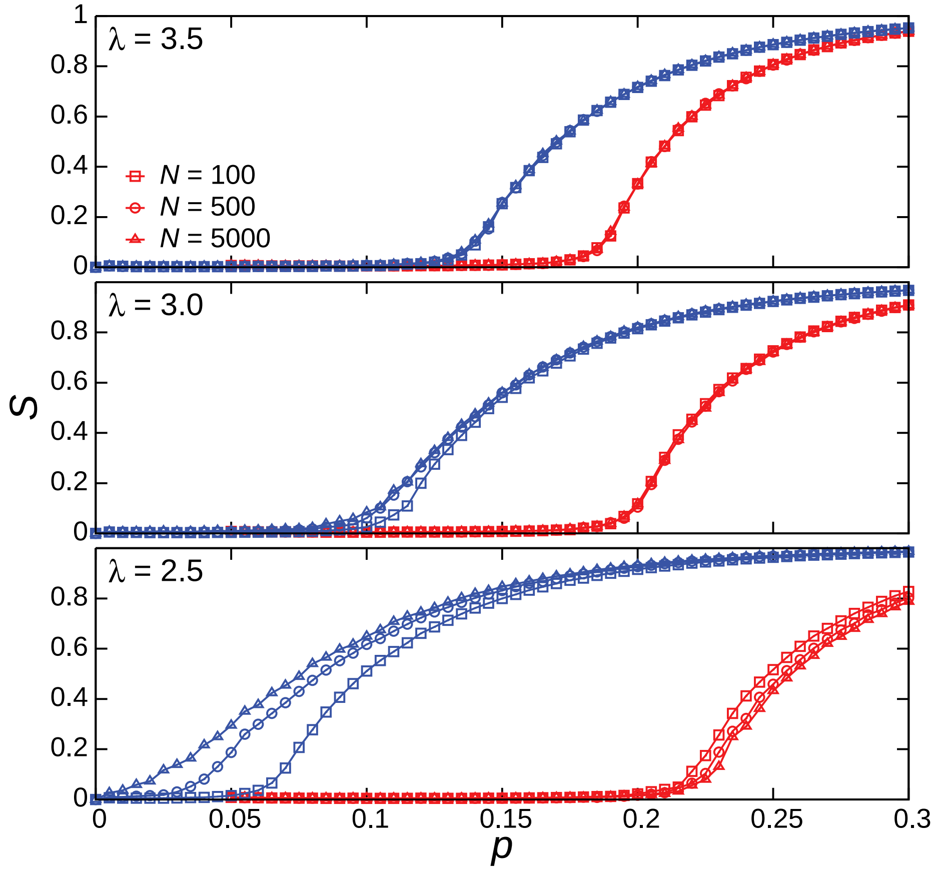

Figure 3: Robustness of scale-free networks. Here $r_m = \delta(m, 3)$, $s_n \sim n^{-\lambda}$, $f_c = 1/2$, and $N \equiv \max\{n \mid s_n > 0\}$. Increasing N and decreasing λ, measures known to improve the robustness of scale-free networks, actually magnifies the robustness gap. Surprisingly, this also increases the fragility of the module network, indicating that optimizing against structural failure may worsen the network's functional resilience.

Acknowledgments

We thank H. Rozenfeld, F. Simini, Y.-R. Lin, D. Wang, D. ben-Avraham, and A.-L. Barabási for many useful discussions, and Hartwig Siebner and Kristoffer Madsen at Hvidovre Hospital's Danish Research Centre for Magnetic Resonance for normal-patient fMRI data. The authors acknowledge the Center for Complex Network Research, supported by the James S. McDonnell Foundation, the NSF, NIH, US ONR and ARL, DTRA, and NKTH NAP.

A Datasets

In this work, we study six empirical networks. The Word Association, Metabolic, and Protein-Protein Interaction (all) networks were previously used in [2]; details are available

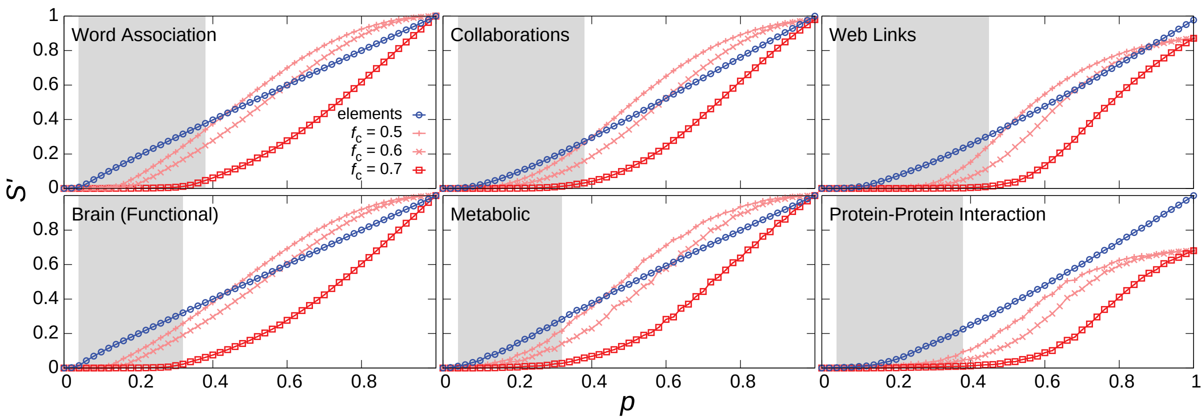

Figure 4: We simulate failures in a number of real networks, from functional brain networks to WWW hyperlinks and collaborative social networks. Many of these networks are robust to random failures (the element networks exhibit very small p_c), but all networks The behavior of the giant component for all the empirical networks qualitatively matches that of the model, as the identified modules uncouple faster than the network itself. Shaded regions provide a guide to the eye for the robustness gap ($f_c = 0.7$). For full dataset details, see App. A.

there. The Web Links network is constructed from a web crawl made available by Google; see http://google.com/programming-contest/. The Collaborations network is constructed between authors who share at least one publication on the arXiv:cond-mat system [17]. The Brain network was derived using normal patient fMRI data where each node is a "voxel" dividing the brain spatially and links exist between voxels whose respective BOLD time series are correlated (measured using Normalized Mutual Information). We begin with the top 200k most correlated links. A single voxel had very high degree, $k = 0.73N$ (the next highest degree is $k = 0.096N$) so we first remove it. This leaves 5038 nodes and 196311 links. We further preprocess this dense network by extracting its multiscale backbone [25] ($\alpha = 0.37$), giving a final network of 5038 nodes and 77680 links. For all networks, link communities were extracted at the level of maximum partition density [2], providing the estimated modules.

B Finite-size effects

For the empirical networks analyzed in Fig. 4, we modified our definition of the quantity S due to finite-size effects. There are two sources for these effects: (i) the number of modules is often much smaller than the number of elements, so that a small network of a few thousand elements may only have a few hundred modules; and (ii) the rate at which elements fail may be slower than the rate at which modules fail (the former is simply given by p but the latter also depends on s_n and f_c). We suppress these effects by choosing S with a well-behaved denominator as $p \to 0$. Specifically, our options are $S(p) = N_{\mathrm{gcc}}(p)/N(p)$ or $S' = N_{\mathrm{gcc}}(p)/N(1)$, where $N_{\mathrm{gcc}}(p)$ is the number of nodes (either elements or modules) within

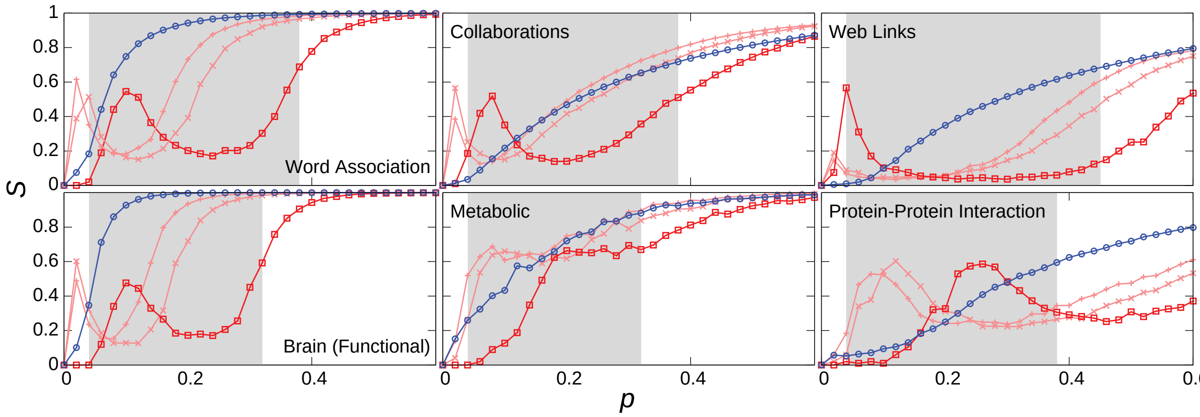

Figure 5: For the empirical datasets in Fig. 4, we present here the same data for the original definition of S, the fraction of remaining nodes within the giant component. For some of the networks the transition points are more dramatic in this representation, however for many it is difficult to determine their location due to strong finite-size effects.

the largest component at percolation probability p and $N(p)$ is the total number of nodes at percolation probability p. The quantity S' has better behavior under the above conditions, although the transition appears less dramatic than it does for S. In Fig. 5 we present the same as Fig. 4 using the original definition of S.

References

[1] D. Achlioptas, R. M. D'Souza, and J. Spencer. Explosive Percolation in Random Networks. *Science*, 323(5920):1453–1455, 2009.

[2] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307):761–764, 2010.

[3] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47–97, Jan 2002.

[4] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[5] J. A. Almendral, I. Leyva, D. Li, I. Sendiña Nadal, S. Havlin, and S. Boccaletti. Dynamics of overlapping structures in modular networks. *Phys. Rev. E*, 82(1):016115, Jul 2010.

[6] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou. Synchronization in complex networks. *Physics Reports*, 469(3):93–153, 2008.

[7] A.-L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.

[8] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, April 2010.

[9] 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.*, 85(25):5468–5471, Dec 2000.

[10] Y. S. Cho, J. S. Kim, J. Park, B. Kahng, and D. Kim. Percolation transitions in scale-free networks under the achlioptas process. *Phys. Rev. Lett.*, 103(13):135702, Sep 2009.

[11] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.*, 85(21):4626–4628, Nov 2000.

[12] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin. Breakdown of the internet under intentional attack. *Phys. Rev. Lett.*, 86(16):3682–3685, Apr 2001.

[13] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Critical phenomena in complex networks. *Rev. Mod. Phys.*, 80(4):1275–1335, Oct 2008.

[14] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc Natl Acad Sci USA*, 99(12):7821–7826, 2002.

[15] T. M. Liggett. *Interacting Particle Systems (Classics in Mathematics)*. Springer, 2004.

[16] A. E. Motter. Cascade control and defense in complex networks. *Phys. Rev. Lett.*, 93(9):098701, Aug 2004.

[17] M. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2):404, 2001.

[18] M. Newman. Properties of highly clustered networks. *Phys. Rev. E*, 68(2):026121, 2003.

[19] M. Newman and J. Park. Why social networks are different from other types of networks. *Phys. Rev. E*, 68(3):036122, 2003.

[20] M. E. J. Newman. Modularity and community structure in networks. *Proc Natl Acad Sci USA*, 103(23):8577–8582, 2006.

[21] M. E. J. Newman. *Networks: An Introduction.* Oxford University Press, USA, 2010.

[22] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.

[23] R. Parshani, S. V. Buldyrev, and S. Havlin. Interdependent networks: Reducing the coupling strength leads to a change from a first to second order percolation transition. *Phys. Rev. Lett.*, 105(4):048701, Jul 2010.

[24] F. Radicchi and S. Fortunato. Explosive percolation in scale-free networks. *Phys. Rev. Lett.*, 103(16):168701, Oct 2009.

[25] M. Serrano, M. Boguñá, and A. Vespignani. Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences*, 106(16):6483, 2009.

[26] V. Sood and S. Redner. Voter model on heterogeneous graphs. *Phys. Rev. Lett.*, 94(17):178701, May 2005.

[27] D. Stauffer and A. Aharony. *Introduction to Percolation Theory.* Taylor & Francis, 1994.

[28] S. H. Strogatz. Exploring complex networks. *Nature*, 410:268–276, 2001.