

Resilience of modular complex networks

Saray Shai,¹ Dror Y. Kenett,² Yoed N. Kenett,³

Miriam Faust,^{3,4} Simon Dobson,¹ and Shlomo Havlin⁵

¹*School of Computer Science, University of St Andrews,
St Andrews, Fife KY16 9SX, Scotland, UK*

²*Center for Polymer Studies and Department of Physics,
Boston University, Boston, Massachusetts 02215, USA*

³*Gonda Brain Research Center, Bar-Ilan University, Ramat-Gan, Israel*

⁴*Department of Psychology, Bar-Ilan University, Ramat-Gan, Israel*

⁵*Department of Physics, Bar-Ilan University, 52900 Ramat-Gan, Israel*

(Dated: April 21, 2014)

Complex networks often have a modular structure, where a number of tightly-connected groups of nodes (modules) have relatively few interconnections. Modularity had been shown to have an important effect on the evolution and stability of biological networks [1], on the scalability and efficiency of large-scale infrastructure [2, 3], and the development of economic and social systems [4, 5]. An analytical framework for understanding modularity and its effects on network vulnerability is still missing. Through recent advances in the understanding of multilayer networks [6–8], however, it is now possible to develop a theoretical framework to systematically study this critical issue. Here we study, analytically and numerically, the resilience of modular networks under attacks on interconnected nodes, which exhibit high betweenness values [9, 10] and are often more exposed to failure [3, 11–13]. Our model provides new understandings into the feedback between structure and function in real world systems, and consequently has important implications as diverse as developing efficient immunization strategies, designing robust large-scale infrastructure, and understanding brain function.

Network science has become a leading approach to the study of emergent collective phenomena in complex systems, with a wide range of applications to fundamental real world systems [14]. Many real world systems have been shown to exhibit a modular structure, in which smaller clusters of nodes are connected more to each other than to the network at large, which is key to their behavior and functioning [15–18]. For example, recent studies of biological networks show that the deletion of nodes connecting between modules can have a deleterious effect on the network integrity [19], efficiency [20], and stability [21]. Here we provide an analytical framework for studying the robustness of modular networks in the presence of attacks on interconnected nodes. We study a percolation process on networks consisting of a varying number of modules, m , and a varying number of interconnected nodes. The analytical solution reveals two percolation regimes separated by a critical number of modules m^* : for $m < m^*$ one needs to remove all interconnected nodes to break the system, while the modules are almost unaffected internally. In contrast, for $m > m^*$ one needs to remove only a fraction of the interconnected nodes, before the system collapses. This is due to the fact that for $m > m^*$ the number of interconnected nodes is high and partial removal of these breaks the modules internally, which helps to bring about the rapid collapse of the whole system. Our approach can also be used to study analytically attacks on high betweenness centrality nodes, which in modular structures, correspond to interconnected nodes. Such attacks, which have only been studied numerically so far, are considered to be among the most harmful attack strategies [9, 10].

We consider a modular network with N nodes divided into m equal sized modules. Similarly to [16], we define p_{in} as the probability to connect nodes in the same module and p_{out} as the probability to connect nodes in different modules. Thus, the total number of intra-module (inter-module) links is given by the probability for a link p_{in} (p_{out}) multiplied by the number of possible links yielding

$$M_{\text{in}} = p_{\text{in}} \frac{N(\frac{N}{m} - 1)}{2}, \quad (1)$$

$$M_{\text{out}} = p_{\text{out}} \frac{N(m - 1)\frac{N}{m}}{2}. \quad (2)$$

We define α to be the ratio between the probabilities for an intra- and inter-module link

$$\alpha = \frac{p_{\text{in}}}{p_{\text{out}}}. \quad (3)$$

In Fig. 1(a)-(c) we present an example of modular networks generated with different values of α , and visualized using force-directed layout, which has been shown to demonstrate network modularity [22]. Note that the ratio between the number of inter-modules links and intra-module links depends not only on α , but also on the number of modules

$$\frac{M_{\text{in}}}{M_{\text{out}}} = \frac{p_{\text{in}}(\frac{N}{m} - 1)}{p_{\text{out}}(m - 1)\frac{N}{m}} \sim \frac{p_{\text{in}}}{p_{\text{out}}(m - 1)} = \frac{\alpha}{m - 1}. \quad (4)$$

Thus, our model is taking into consideration that systems comprised of more modules have more inter-links, as illustrated in Fig. 1(d). See also Supplementary Fig. S1, where we show the increase of the mean inter-degree as a function of m .

Given the model described above for generating random modular networks, we proceed to study percolation properties for such networks. We consider a modular Erdős-Rényi (ER) network [23, 24] where both the intra- and inter-connectivity are Poisson distributed with means k_{intra} and k_{inter} respectively. Using the generating function approach presented in [8] (see full details in the Supplementary Information), we find that in the presence of random nodes failure, the giant component emerges when the following is satisfied

$$(1 - k_{\text{intra}})(1 - k_{\text{intra}} - \frac{(m - 2)k_{\text{inter}}}{m - 1}) - \frac{k_{\text{inter}}^2}{m - 1} = 0. \quad (5)$$

This condition yields $k = k_{\text{intra}} + k_{\text{inter}} = 1$ for every m , recovering the standard result for single networks without communities. Thus, in the case of random node failure the percolation threshold only depends on the mean degree, k .

However, in real systems the interconnected nodes are often more exposed to failure than other nodes. For example, it has been shown that aging and schizophrenia could result in damage to the interconnected nodes in brain networks [12, 13]. In addition, it is often the case that interconnected nodes are considered to be important; for example, the New York City and London airports, which provide an attractive target for attacks [3]. Therefore, in the following, we consider an attack on modular ER networks where the interconnected nodes are randomly removed. Let $r_i(k_1, k_2, \dots, k_m)$ be the occupation probability of a node from module i with k_1 links in module 1, k_2 links in module 2 and etc. When the interconnected nodes are randomly removed, this probability is given by

$$r_i(k_1, k_2, \dots, k_m) = \begin{cases} 1, & \text{if } \sum_{\substack{j=1 \\ j \neq i}}^m k_j = 0 \\ q, & \text{otherwise} \end{cases}, \quad (6)$$

where q is the probability that a randomly chosen interconnected node is occupied. Let p be the general occupation probability, i.e. the probability that a randomly chosen node is occupied. Since the probability for a node to be interconnected is $1 - e^{-k_{\text{inter}}}$, i.e. one minus a Poisson distribution with mean k_{inter} at 0, we obtain

$$q = \frac{p - e^{-k_{\text{inter}}}}{1 - e^{-k_{\text{inter}}}}. \quad (7)$$

We extend Callaway *et al.*'s approach [25] for studying the robustness of networks to intentional attacks, from single-module networks to modular networks in a similar approach as in [8] (see full details in the Methods section), and solve for the occupation probability given in (6), obtaining two possible solutions for the critical occupation probability of interconnected nodes

$$q_c = 0 \quad (8)$$

$$q_c = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad (9)$$

$$\begin{aligned} \text{where} \quad a &= k_{\text{intra}} k_{\text{inter}} e^{-k_{\text{inter}}} \\ b &= k_{\text{intra}} + k_{\text{inter}} - k_{\text{intra}} e^{-k_{\text{inter}}} - k_{\text{intra}} k_{\text{inter}} e^{-k_{\text{inter}}} \\ c &= k_{\text{intra}} e^{-k_{\text{inter}}} - 1. \end{aligned}$$

From these solutions, we obtain the critical occupation probability p_c , using Eq. (7).

Due to symmetry, once the giant component emerges ($p > p_c$) the fraction of nodes of module i in the giant component equals to the total fraction of nodes in the giant component, $S_i = S$, and one obtains,

$$S = e^{-k_{\text{inter}}(1-q)}(1 - e^{-k_{\text{intra}}S}) + q(1 - e^{-(k_{\text{intra}}S + k_{\text{inter}}S)}). \quad (10)$$

For $k_{\text{intra}} = 0$, only a fraction q of the nodes in the network are connected, and one obtains $S = q(1 - e^{-kS})$, recovering the standard result for percolation in single networks [23, 24].

In Fig. 2, we confirm our analytical solution (Eqs. (8)-(10)) by extensive numerical simulation of ER modular networks of size $N = 600\,000$. First, we show the percolation threshold as a function of the number of modules m where the mean degree is kept fixed $k = 4$ and $\alpha = 100$, see Fig. 2(a). Similar results are shown for $\alpha = 10$, and 1000 in Supplementary Fig. S2. Let m^* be the transition point where the two analytical solutions cross each other.

In the regime where $m < m^*$ the attack on interconnected nodes mainly breaks the connectivity between the modules leaving their internal structure intact. Thus, only the removal of all the interconnected nodes ($q_c = 0$) breaks down the giant component.

In order to illustrate this effect, in Fig. 3 we visualize the giant component at $S = 0.1$ (close to total collapse) with interconnected nodes shown in black and all other nodes colored according to the module they belong to. For a network with $m = 4 < m^*$, random node failure destroys the internal structure of the modules evenly, see Fig. 3(a). In this random failure case, all the modules always appear in the giant component (i.e. there is always at least one node from each module in the giant component) as shown in Fig. 3(e), and the size of modules is very narrowly distributed, see Fig. 3(g). In contrast to random failure, when attacking the interconnected nodes (at $S = 0.1$), see Fig. 3(b), not all the modules remain in the giant component (for example, in Fig. 3(b) there are only two of them). However, the modules that do remain, are almost intact, containing 14.6% of their initial nodes, significantly more than in the random case. This point is demonstrated also in Supplementary Figs. S5-S6 where we analyze the modules structure in the second largest cluster.

In contrast, for $m > m^*$, the interconnected nodes play an important role also in the internal structure of modules and therefore there is no need to remove all of them in order to break down the giant component. Nevertheless, the attack still leaves them more complete than in the case of random removal, see Fig. 3(c)-(d). Furthermore, in the case of attack, usually *all* modules appear in the giant component (see Fig. 3(f)), and thus their relative size is smaller compared to the $m < m^*$ case (see Fig. 3(g)). As m increases, the difference between attack and random case becomes smaller, and as a result the percolation threshold converges to the one obtained for random failure, see Supplementary Fig. S3.

For the case of $m < m^*$, the attack of interconnected nodes has a weak effect on the internal structure of the modules, and the removal of inter-module nodes results in an abrupt decrease in the size of the giant component, see Fig. 2(b). In addition, while for $m = 100 > m^*$ (Fig. 2(c)) we observe a regular second order percolation transition characterized by the continuous decrease of S and the sharp peak in S_{second} , the case of $m < m^*$ demonstrates an abrupt, first order transitions. The reason is that the second largest cluster contains large connected subgraphs corresponding to modules who “dropped” from the giant component, see Fig. 3(e). Therefore, with the emergence of the giant component, these modules become

part of it, leading to a sudden drop in the size of the second largest cluster.

In Fig. 2(d), we show the critical number of modules, m^* , as a function of α for networks with mean degree $k = 4$. It is seen that m^* is increasing with α , and the percolation threshold at this point is $p_c^* \approx 0.3417$ independent of α , meaning the transition takes place at a fixed inter-module average degree $k_{\text{inter}}^* = -\ln(p_c^*) \approx 1.0738$. We show how the critical percolation threshold p_c^* and the critical mean inter-degree k_{inter}^* are changing with k in Supplementary Fig. S4.

In order to further demonstrate the transition in the p_c behavior, in Fig. 2(e) we show the percolation threshold as a function of k_{inter} for networks with mean intra-degree $k_{\text{intra}} = 2$ and number of modules $m = 10$ fixed. Here we see a similar transition in p_c as before, but the critical point is now a function of the concentration of interconnected nodes. At a critical $k_{\text{inter}}^* = k_{\text{inter}} \approx 0.693$, p_c changes from Eq. (8) behavior to Eq. (9).

Finally, in modular structures the interconnected nodes have high betweenness centrality (see Fig. S11), and thus, our framework also provides an analytical tool of studying attacks on high betweenness centrality nodes, where only numerical simulations currently exist that suggest such an attack is one of the most harmful attack strategies [9, 10]. Figure S11 compares the betweenness centrality of nodes with inter-module connections (called inter-nodes) and nodes with only intra-module connections (called intra-nodes) for networks of size $N = 100\,000$ with $m = 10$ modules. First, we show that the average betweenness centrality of interconnected nodes is significantly higher than for nodes without interconnections in networks with mean intra-degree $k_{\text{intra}} = 2$ and a varying number of interconnections, see Fig. S11(a). Then, for $k_{\text{inter}} = 2$, we show that the betweenness centrality distribution of interconnected nodes has a broader tail, meaning that interconnected nodes are much more likely to have high betweenness centrality. In Supplementary Fig. S7 we obtain similar results in networks with $k = 4$ and different values of α . Thus, our analytical results of attack on interconnected nodes can be regarded as a theory for attacking high betweenness nodes.

Our analytical and numerical investigation of the effect of modularity on network stability has important implications for real world networks, such as cognitive and neural brain networks. The modular architecture of neural structural and functional networks is considered a fundamental principle of the brain [26]. This non-random modular architecture is crucial for the brain's functional demands of segregation and integration of information [27]. In fact,

disrupted brain modular organization is related to neuropathology, such as schizophrenia [28], autism [29], Alzheimer's [30] and impulsivity [31]. Nevertheless, research investigating any possible negative aspects of modular organization in brain networks is lacking. At the cognitive level (the level of information processing in the brain), network analysis is mainly focused on language and memory networks [32]. Yet, knowledge on modular effect and importance in cognitive network organization is limited. Recently, the semantic memory organization of persons with Asperger syndrome was compared to that of neurotypical controls using network analysis [33]. This research found that the semantic memory network of persons with Asperger syndrome is more modular than that of neurotypical matched controls. The authors suggest that this "hyper-modularity" is related to the Asperger syndrome rigidity of thought, e.g. difficulty in comprehending high level aspects of language. Thus, modular organization can have a negative effect on real world networks by leading to rigidity of the network which might hinder proper network function.

Finally, our study offers an efficient immunization approach in modular networks, where epidemic spreading can be prevented at a lower cost by immunizing interconnected nodes. For both regimes, below and above m^* , the percolation threshold obtained from attacking the interconnected nodes is higher than the case of random failure and therefore immunization of these nodes is more effective. For the regime $m < m^*$, this can be done at a very low cost as the percolation threshold is very high. Thus, in geographically distant social networks, it is worth vaccinating people that link between different communities such as businessmen traveling a lot between countries.

METHODS

We give a brief derivation of our analytical solution (Eqs. (8)-(10)). We extend Callaway *et al.*'s approach [25] for studying the robustness of networks to intentional attacks, from single-module networks to modular networks in a similar manner that was done in [8]. We define the generating functions for the degree and excess degree distributions of occupied nodes

$$F_i(x_1, x_2, \dots, x_m) = \sum_{k_1, k_2, \dots, k_m=0}^{\infty} p_{k_1, k_2, \dots, k_m}^i r_{k_1, k_2, \dots, k_m}^i x_1^{k_1} x_2^{k_2} \dots x_m^{k_m} \quad (11)$$

$$F_{ij}(x_1, x_2, \dots, x_m) = \sum_{k_1, k_2, \dots, k_m=0}^{\infty} q_{k_1, k_2, \dots, k_m}^{ij} r_{k_1, k_2, \dots, k_m}^i x_1^{k_1} x_2^{k_2} \dots x_m^{k_m} \quad (12)$$

where $p_{k_1, k_2, \dots, k_m}^i$ is the probability that a node from module i has degree (k_1, k_2, \dots, k_m) , $q_{k_1, k_2, \dots, k_m}^{ij}$ is the probability of following a randomly chosen ij -edge to a node with excess degree (k_1, k_2, \dots, k_m) , and $r_{k_1, k_2, \dots, k_m}^i$ is the occupation probability of a node with degree (k_1, k_2, \dots, k_m) defined in (6). By substituting (6) into (11)-(12), in the case of modular ER networks with average intra- and inter-degree k_{intra} , k_{inter} respectively, we obtain

$$F_i(x) = e^{k_{\text{intra}}(x_i-1)-k_{\text{inter}}}(1-q) + qG_i(x) \quad (13)$$

$$F_{ij}(x) = \begin{cases} \frac{\partial F_i(x)}{\partial x_j} \frac{m-1}{k_{\text{inter}}} = qG_i(x), & \text{if } i \neq j \\ \frac{\partial F_i(x)}{\partial x_i} \frac{1}{k_{\text{intra}}} = F_i(x), & \text{otherwise} \end{cases} \quad (14)$$

where $x = (x_1, x_2, \dots, x_m)$ and $G_i(x)$ is the generating function of the degree distribution, see Supplementary Information. We define the generating function for the distribution of the number of occupied nodes in the component reachable by following a randomly chosen ij -edge to a i -node and then following its additional outgoing links

$$J_{ij}(x) = 1 - F_{ij}(1) + x_i F_{ij}[J_{1i}, J_{2i}, \dots, J_{mi}]. \quad (15)$$

And similarly, the distribution of the number of nodes reachable from a randomly chosen i -node (rather than ij -edge) is generated by

$$J_i(x) = 1 - F_i(1) + x_i F_i[J_{1i}, J_{2i}, \dots, J_{mi}]. \quad (16)$$

Then, the average number of occupant j -nodes in the component of a randomly chosen i -node, is given by

$$\begin{aligned} \langle s_i \rangle_j &= \frac{\partial J_i}{\partial x_j}(x)|_{x=1} \\ &= \delta_{ij} F_i(1) + \frac{\partial F_i}{\partial x_j}(1) \frac{\partial J_{ii}}{\partial x_j}(1) + \sum_{\substack{l=1 \\ l \neq i}}^m \frac{\partial F_i}{\partial x_l}(1) \frac{\partial J_{li}}{\partial x_j}(1) \\ &= \delta_{ij} F_i(1) + k_{\text{intra}} F_i(1) \frac{\partial J_{ii}}{\partial x_j}(1) + q \frac{k_{\text{inter}}}{m-1} \sum_{\substack{l=1 \\ l \neq i}}^m \frac{\partial J_{li}}{\partial x_j}(1). \end{aligned} \quad (17)$$

Solving the system (17), see full details in the Supplementary Information, we obtain the critical occupation probability of interconnected nodes in which the average component size

diverges, given in Eqs. (8)-(9). Finally, once the giant component emerges ($p > p_c$), the fraction of i -nodes belonging to the giant component, S_i , is given by

$$S_i = 1 - J_i(1) = F_i(1) - F_i(u_{1i}, u_{2i}, \dots, u_{mi}) \quad (18)$$

where $u_{ji} = 1 - S_j$, yielding Eq. (10).

Acknowledgments

SH and DYK thank DTRA, ONR, BSF, the LINC (No. 289447) and the Multiplex (No. 317532) EU projects, the DFG, and the Israel Science Foundation for support. SS is supported by a scholarship from the Scottish Informatics and Computer Science Alliance.

Author contributions

SS, DYK, YNK, SD, MF, and SH performed the research and wrote the paper.

Competing financial interests

The authors declare no competing financial interests.

-
- [1] E. Bullmore and O. Sporns, Nat Rev Neurosci **13**, 336 (2012).
 - [2] K. A. Eriksen, I. Simonsen, S. Maslov, and K. Sneppen, Phys. Rev. Lett. **90**, 148701 (2003).
 - [3] R. Guimerà, S. Mossa, A. Turtshi, and L. A. N. Amaral, Proc. Natl Acad. Sci. USA **102**, 7794 (2005).
 - [4] A. Garas, P. Argyrakis, and S. Havlin, Eur Phys J B **63**, 265 (2008).
 - [5] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West, Proc. Natl Acad. Sci. USA **104**, 7301 (2007).
 - [6] S. Buldyrev, R. Parshani, G. Paul, H. Stanley, and S. Havlin, Nature **464**, 1025 (2010).
 - [7] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, Nat Phys **8**, 40 (2012).
 - [8] E. A. Leicht and R. M. D’Souza, (2009), arXiv:0907.0894.
 - [9] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, Phys. Rev. E **65**, 056109 (2002).
 - [10] C. M. Schneider, A. A. Moreira, J. S. Andrade, S. Havlin, and H. J. Herrmann, Proceedings of the National Academy of Sciences **108**, 3838 (2011).

- [11] A. Bashan, R. P. Bartsch, J. W. Kantelhardt, S. Havlin, and P. C. Ivanov, *Nat Commun* **3**, 702 (2012).
- [12] D. Meunier, S. Achard, A. Morcom, and E. Bullmore, *NeuroImage* **44**, 715 (2009).
- [13] F. Shi, P.-T. Yap, W. Gao, W. Lin, J. H. Gilmore, and D. Shen, *NeuroImage* **62**, 1622 (2012).
- [14] S. Havlin, D. Kenett, E. Ben-Jacob, A. Bunde, R. Cohen, H. Hermann, J. Kantelhardt, J. Kertész, S. Kirkpatrick, J. Kurths, J. Portugali, and S. Solomon, *The European Physical Journal Special Topics* **214**, 273 (2012).
- [15] M. E. J. Newman, *Phys. Rev. E* **67**, 026126 (2003).
- [16] M. E. J. Newman and M. Girvan, *Phys. Rev. E* **69**, 026113 (2004).
- [17] J.-j. Wu, Z.-y. Gao, and H.-j. Sun, *Phys. Rev. E* **74**, 066111 (2006).
- [18] J. P. Gleeson, *Phys. Rev. E* **77** (2008).
- [19] J.-D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth, and M. Vidal, *Nature* **430**, 88 (2004).
- [20] O. Sporns, C. J. Honey, and R. Kötter, *PLoS ONE* **2**, e1049 (2007).
- [21] Y. He, J. Wang, L. Wang, Z. J. Chen, C. Yan, H. Yang, H. Tang, C. Zhu, Q. Gong, Y. Zang, and A. C. Evans, *PLoS ONE* **4**, e5226 (2009).
- [22] A. Noack, *Phys. Rev. E* **79**, 026102 (2009).
- [23] P. Erdős and A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci* **5**, 17 (1960).
- [24] B. Bollobás, *Random Graphs*, 2nd ed. (Cambridge University Press, 2001).
- [25] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000).
- [26] D. Meunier, R. Lambiotte, and E. T. Bullmore, *Frontiers in neuroscience* **4** (2010).
- [27] E. Bullmore and O. Sporns, *Nature Reviews Neuroscience* **13**, 336 (2012).
- [28] A. F. Alexander-Bloch, N. Gogtay, D. Meunier, R. Birn, L. Clasen, F. Lalonde, R. Lenroot, J. Giedd, and E. T. Bullmore, *Frontiers in systems neuroscience* **4** (2010).
- [29] P. Barttfeld, B. Wicker, S. Cukier, S. Navarta, S. Lew, R. Leiguarda, and M. Sigman, *Neuropsychologia* (2012).
- [30] E. C. van Straaten and C. J. Stam, *European Neuropsychopharmacology* (2012).
- [31] F. C. Davis, A. R. Knodt, O. Sporns, B. B. Lahey, D. H. Zald, B. D. Brigidi, and A. R. Hariri, *Cerebral Cortex* **23**, 1444 (2013).
- [32] A. Baronchelli, R. Ferrer-i Cancho, R. Pastor-Satorras, N. Chater, and M. H. Christiansen,

Trends in Cognitive Sciences **17**, 348 (2013).

[33] Y. N. Kenett, R. Gold, and M. Faust, (submitted).

[34] M. Bastian, S. Heymann, and M. Jacomy, in *International AAAI Conference on Weblogs and Social Media* (2009).

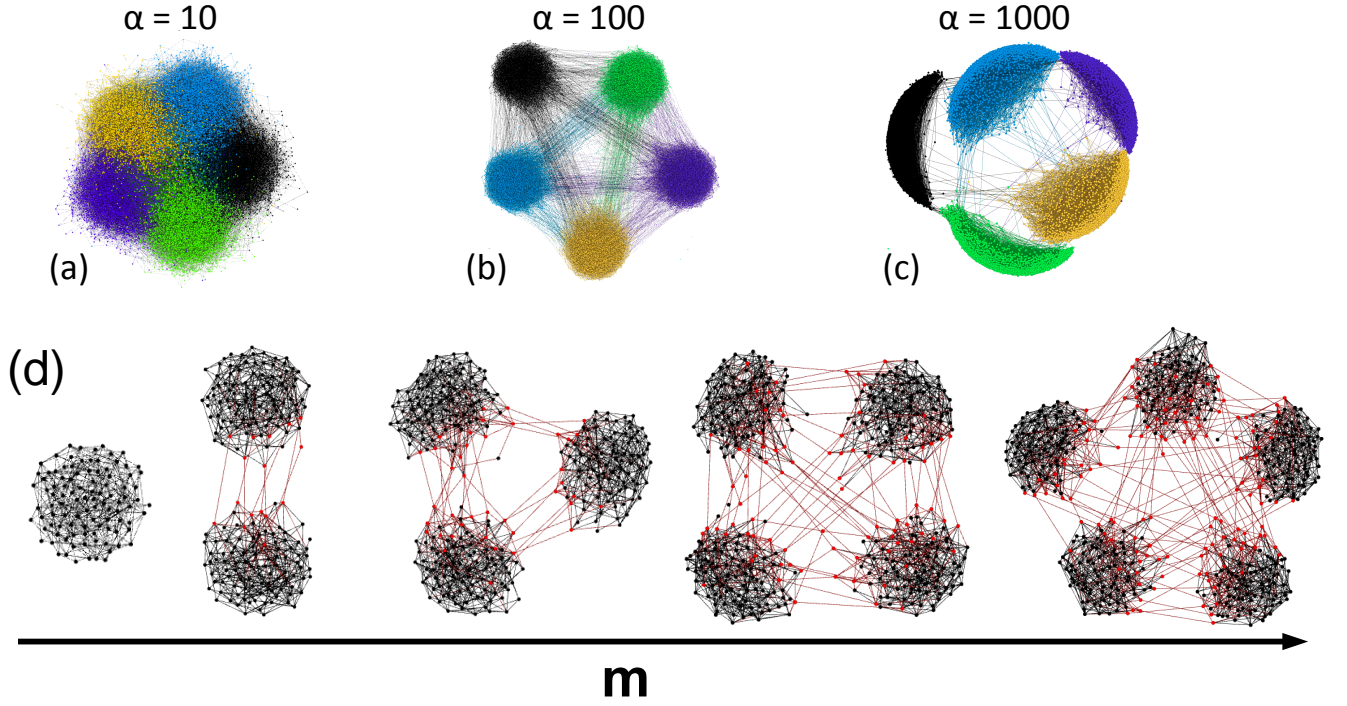


FIG. 1. **Visualization of the model for generating random modular networks.** (a)-(c) Illustration of the effect of α on the obtained modular network using Gephi with force atlas layout [22, 34], on a network of size $N = 10\,000$ with mean degree $k = 8$ divided into $m = 5$ modules. (d) Illustration of the effect of the number of modules m on the obtained network with a number of inter-module links increasing with the number of modules. Inter-connected nodes are shown in red.

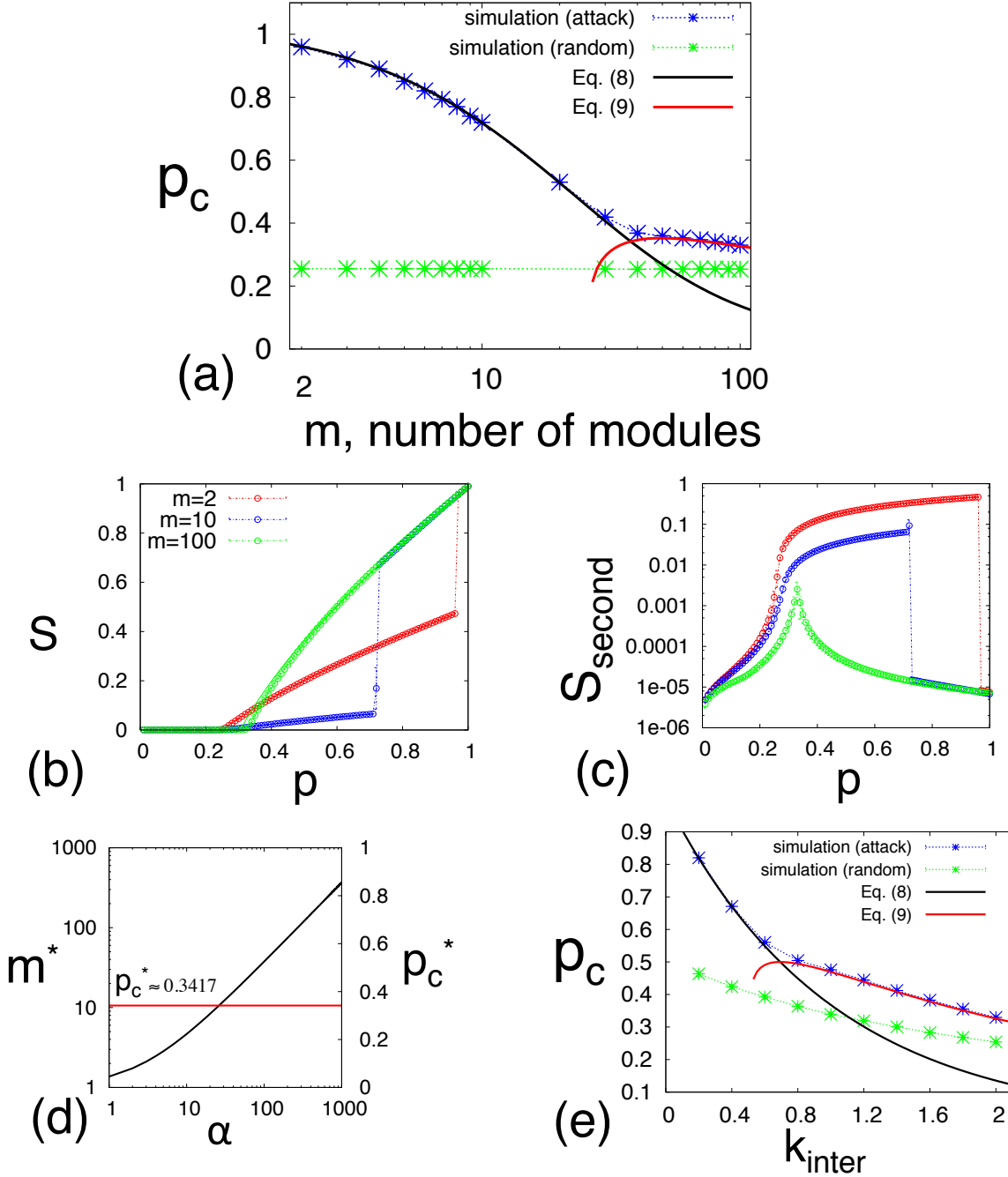


FIG. 2. **Two percolation regimes when attacking interconnected nodes.** (a) p_c as a function of m calculated for networks with $k = 4$, $\alpha = 100$. Simulation points obtained from at least 1000 simulation runs of networks of size $N = 600\,000$. Solid lines represent the analytical result obtained in (8)-(9). (b)-(c) Fraction of nodes in the largest cluster S and second largest cluster S_{second} as a function of occupation probability p . Solid lines represent the analytical result obtained in (10). (d) Critical number of modules m^* , defined as the point where the solutions from (8) and (9) cross each other, as a function of α . p_c^* is the percolation threshold at this point. (e) p_c as a function of k_{inter} calculated for networks with $m = 10$, $k_{\text{intra}} = 2$.

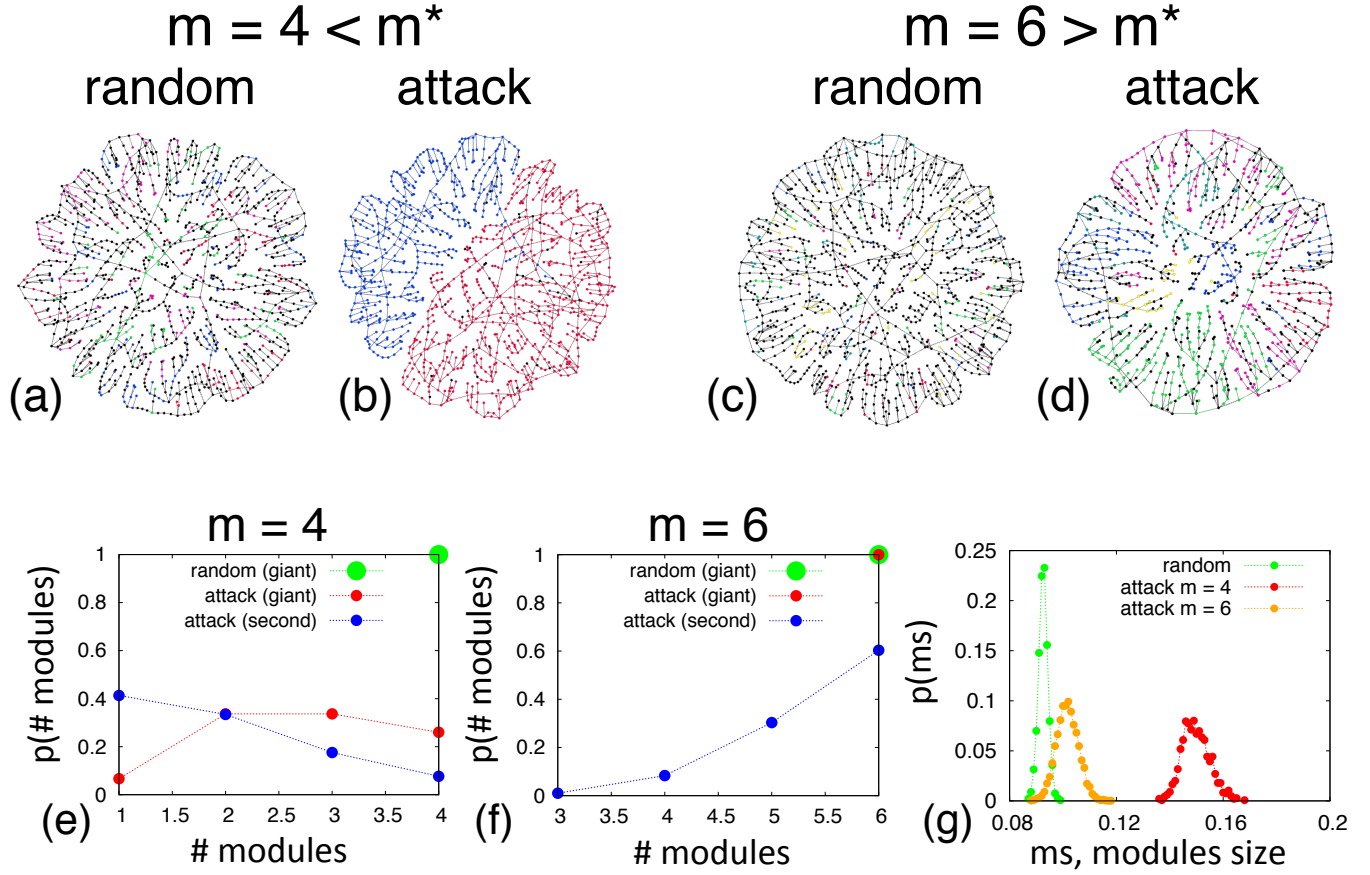


FIG. 3. **Size of modules in the giant component at $S = 0.1$.** Visualization is shown for networks of size $N = 12\,000$ with mean degree $k = 4$ and $\alpha = 10$, at the point where the giant component contains 10% of the nodes ($S = 0.1$) (a),(c) for random node removal, (b),(d) for attack on interconnected nodes. (e)-(f) Distribution of the number of modules in the giant component and second largest component at $S = 0.1$. A module is considered to be part of a component if at least one of its nodes are part of the component. (g) Distribution of the size of modules in the giant component at $S = 0.1$, normalized by the initial module size. Note that in (g), the size of modules is measured by reconstructing the graph of each module in the giant component, and counting its number of nodes in this graph. In other words, interconnected nodes that have been detached from their original module are not considered. Results obtained by at least 1000 simulation runs of networks of size $N = 600\,000$ with mean degree $k = 4$.

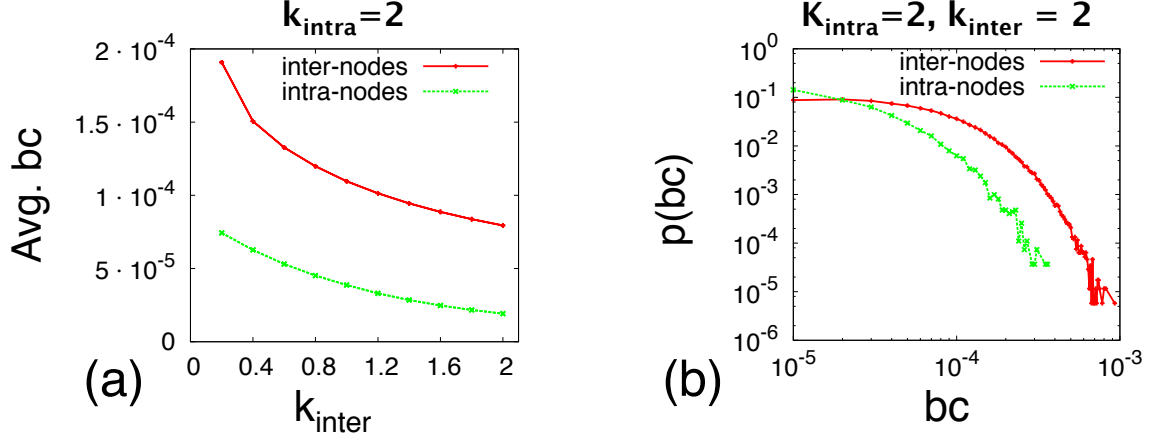


FIG. 4. **Betweenness centrality of interconnected nodes.** Betweenness centrality of inter-nodes (nodes that have at least one interconnection) and intra-nodes (nodes with only intraconnections) in networks of size $N = 100\,000$ with $m = 10$ modules. (a) Mean betweenness centrality as a function of k_{inter} in networks with $k_{\text{intra}} = 2$. (b) Distribution of betweenness centrality in networks with $k_{\text{intra}} = k_{\text{inter}} = 2$.

Resilience of modular complex networks: Supplementary Information

Saray Shai, Dror Y. Kenett, Yoed N. Kenett, Miriam Faust, Simon Dobson,
and Shlomo Havlin

I. INTRODUCTION

We present supplementary material on our paper: “Resilience of modular complex networks”. First, in section II we describe in more details the derivation of the analytical solution presented in the main text. Then, in section III, we examine the properties of our model for generating random modular networks. In section IV we further discuss the two percolation regimes found in the analytical framework and their implications. In section V we examine the internal structure of modules in the second largest cluster, and finally in section VI we discuss the betweenness centrality of interconnected nodes in modular structures.

II. ANALYTICAL SOLUTION

A. Random failure

First, we give full derivation of Eq. (5) in the main text. The generating function for the degree and excess degree distribution (see [1]) of modular ER networks with average intra- and inter-degree k_{intra} , k_{inter} respectively is given by

$$G_i(x) = e^{k_{\text{intra}}(x_i-1)} e^{\frac{k_{\text{inter}}}{m-1} \sum_{j \neq i} (x_j-1)} \quad (\text{S1})$$

$$G_{ij}(x) = \begin{cases} \frac{\frac{\partial G_i}{\partial x_j}(x)}{\frac{\partial G_i}{\partial x_j}(1)} = \frac{k_{\text{inter}}}{m-1} G_i(x) \frac{m-1}{k_{\text{inter}}} = G_i(x), & \text{if } i \neq j \\ \frac{\frac{\partial G_i}{\partial x_j}(x)}{\frac{\partial G_i}{\partial x_j}(1)} = k_{\text{intra}} G_i(x) \frac{1}{k_{\text{intra}}} = G_i(x), & \text{otherwise} \end{cases} \quad (\text{S2})$$

where $x = (x_1, x_2, \dots, x_m)$. Then from [1], the average number of j -nodes in the component of a randomly chosen i -node is given by

$$\langle s_i \rangle_j = \delta_{ij} + k_{\text{intra}} \frac{\partial H_i}{\partial x_j}(1) + \frac{k_{\text{inter}}}{m-1} \sum_{\substack{l=1 \\ l \neq i}}^m \frac{\partial H_l}{\partial x_j}(1) \quad (\text{S3})$$

where δ_{ij} denotes the Kronecker delta and

$$H_{ij}(x) = x_i G_{ij}[H_{1i}, H_{2i}, \dots, H_{mi}] \quad ; \quad H_i(x) = x_i G_i[H_{1i}, H_{2i}, \dots, H_{mi}] \quad (\text{S4})$$

For example, using the notation $h_i = \frac{\partial H_i}{\partial x_1}(1)$, the system obtained for $\langle s_1 \rangle_1$ is:

$$\begin{aligned} h_1 &= 1 + k_{\text{intra}} h_1 + \frac{k_{\text{inter}}}{m-1} (h_2 + h_3 + \dots + h_m) \\ h_2 &= k_{\text{intra}} h_2 + \frac{k_{\text{inter}}}{m-1} (h_1 + h_3 + \dots + h_m) \\ &\vdots \\ h_m &= k_{\text{intra}} h_m + \frac{k_{\text{inter}}}{m-1} (h_1 + h_2 + \dots + h_{m-1}) \end{aligned} \quad (\text{S5})$$

Summing equations for h_2, \dots, h_m , we obtain:

$$\begin{aligned} h_2 + h_3 + \dots + h_m &= k_{\text{inter}} h_1 + (k_{\text{intra}} + \frac{m-2}{m-1} k_{\text{inter}}) (h_2 + h_3 + \dots + h_m) \\ \Rightarrow h_2 + h_3 + \dots + h_m &= \frac{k_{\text{inter}} h_1}{1 - k_{\text{intra}} - \frac{m-2}{m-1} k_{\text{inter}}} \end{aligned} \quad (\text{S6})$$

Substituting into (S5), we obtain:

$$\begin{aligned} h_1 &= 1 + k_{\text{intra}} h_1 + \frac{k_{\text{inter}}}{m-1} (h_2 + h_3 + \dots + h_m) \\ &= 1 + k_{\text{intra}} h_1 + \frac{k_{\text{inter}}}{m-1} \left(\frac{k_{\text{inter}} h_1}{1 - k_{\text{intra}} - \frac{m-2}{m-1} k_{\text{inter}}} \right) \\ &\Rightarrow h_1 \left(1 - k_{\text{intra}} - \frac{\frac{k_{\text{inter}}^2}{m-1}}{1 - k_{\text{intra}} - \frac{m-2}{m-1} k_{\text{inter}}} \right) = 1 \\ &\Rightarrow h_1 = \frac{1 - k_{\text{intra}} - \frac{m-2}{m-1} k_{\text{inter}}}{(1 - k_{\text{intra}}) \left(1 - k_{\text{intra}} - \frac{m-2}{m-1} k_{\text{inter}} \right) - \frac{k_{\text{inter}}^2}{m-1}}. \end{aligned} \quad (\text{S7})$$

h_1 diverges when $(1 - k_{\text{intra}}) \left(1 - k_{\text{intra}} - \frac{(m-2)k_{\text{inter}}}{m-1} \right) - \frac{k_{\text{inter}}^2}{m-1} = 0$. This is also where all h_i diverges, yielding Eq. (5) in the main text.

B. Attack on interconnected nodes

In the case of attack on interconnected nodes, the average number of j -nodes in the component of a randomly chosen i -node is given by

$$\langle s_i \rangle_j = \delta_{ij} F_i(1) + k_{\text{intra}} F_i(1) \frac{\partial J_{ii}}{\partial x_j}(1) + q \frac{k_{\text{inter}}}{m-1} \sum_{\substack{l=1 \\ l \neq i}}^m \frac{\partial J_{li}}{\partial x_j}(1). \quad (\text{S8})$$

where

$$\begin{aligned}
F_i(x) &= e^{k_{\text{intra}}(x_i-1)-k_{\text{inter}}}(1-q) + qG_i(x) \\
F_{ij}(x) &= \begin{cases} \frac{\partial F_i}{\partial x_j}(x) \frac{m-1}{k_{\text{inter}}} = qG_i(x), & \text{if } i \neq j \\ \frac{\partial F_i}{\partial x_i}(x) \frac{1}{k_{\text{intra}}} = F_i(x), & \text{otherwise} \end{cases} \\
J_{ij}(x) &= 1 - F_{ij}(1) + x_i F_{ij}[J_{1i}, J_{2i}, \dots, J_{mi}] \\
J_i(x) &= 1 - F_i(1) + x_i F_i[J_{1i}, J_{2i}, \dots, J_{mi}]
\end{aligned} \tag{S9}$$

For example, using the notation $j_{ij} = \frac{\partial J_{ij}}{\partial x_1}(1)$, the system obtained for $\langle s_1 \rangle_1$ is:

$$\begin{aligned}
j_{11} &= F_1(1) + k_{\text{intra}} F_1(1) j_{11} + q \frac{k_{\text{inter}}}{m-1} (j_{21} + j_{31} + \dots + j_{m1}) \\
j_{12} &= q + q k_{\text{intra}} j_{11} + q \frac{k_{\text{inter}}}{m-1} (j_{21} + j_{31} + \dots + j_{m1}) \\
&\vdots \\
j_{1m} &= q + q k_{\text{intra}} j_{11} + q \frac{k_{\text{inter}}}{m-1} (j_{21} + j_{31} + \dots + j_{m1}) \\
j_{21} &= q k_{\text{intra}} j_{22} + q \frac{k_{\text{inter}}}{m-1} (j_{12} + j_{32} + \dots + j_{m2}) \\
j_{22} &= k_{\text{intra}} F_2(1) j_{22} + q \frac{k_{\text{inter}}}{m-1} (j_{12} + j_{32} + \dots + j_{m2}) \\
&\vdots \\
j_{2m} &= q k_{\text{intra}} j_{22} + q \frac{k_{\text{inter}}}{m-1} (j_{12} + j_{32} + \dots + j_{m2}) \\
&\vdots \\
j_{m1} &= q k_{\text{intra}} j_{mm} + q \frac{k_{\text{inter}}}{m-1} (j_{1m} + j_{3m} + \dots + j_{m-1m}) \\
j_{m2} &= q k_{\text{intra}} j_{mm} + q \frac{k_{\text{inter}}}{m-1} (j_{1m} + j_{3m} + \dots + j_{m-1m}) \\
&\vdots \\
j_{mm} &= k_{\text{intra}} F_m(1) j_{mm} + q \frac{k_{\text{inter}}}{m-1} (j_{1m} + j_{3m} + \dots + j_{m-1m})
\end{aligned} \tag{S10}$$

Since $F_1(1) = F_i(1) = e^{-k_{\text{inter}}}(1 - q) + q$ for all i , we can sum all equations for j_{ii} obtaining:

$$\begin{aligned} j_{11} + j_{22} + \dots + j_{mm} &= F_1(1) + k_{\text{intra}}F_1(1)(j_{11} + j_{22} + \dots + j_{mm}) + \\ &+ q \frac{k_{\text{inter}}}{m-1} (j_{12} + j_{13} + \dots + j_{1m} + \dots + j_{m1} + j_{m2} + \dots + j_{mm-1}) \end{aligned} \quad (\text{S11})$$

Summing for all equations j_{il} for $i \neq l$ we obtain

$$\begin{aligned} j_{12} + j_{13} + \dots + j_{1m} + \dots + j_{m1} + j_{m2} + \dots + j_{mm-1} &= (m-1)q + (m-1) q k_{\text{intra}}(j_{11} + \dots + j_{mm}) + \\ &+ q k_{\text{inter}}(j_{12} + j_{13} + \dots + j_{1m} + \dots + j_{m1} + j_{m2} + \dots + j_{mm-1}) \\ \Rightarrow j_{12} + j_{13} + \dots + j_{1m} + \dots + j_{m1} + j_{m2} + \dots + j_{mm-1} &= \frac{(m-1)q + (m-1) q k_{\text{intra}}(j_{11} + \dots + j_{mm})}{1 - q k_{\text{inter}}} \end{aligned} \quad (\text{S12})$$

And by substituting (S12) into (S11), we obtain

$$\begin{aligned} j_{11} + \dots + j_{mm} &= F_1(1) + k_{\text{intra}}F_1(1) + \frac{q^2 k_{\text{inter}}}{1 - q k_{\text{inter}}} + \frac{q^2 k_{\text{intra}} k_{\text{inter}}(j_{11} + \dots + j_{mm})}{1 - q k_{\text{inter}}} \\ \Rightarrow j_{11} + \dots + j_{mm} &= \frac{F_1(1)(1 - q k_{\text{inter}}) + q^2 k_{\text{inter}}}{(1 - k_{\text{intra}}F_1(1))(1 - q k_{\text{inter}}) - q^2 k_{\text{intra}} k_{\text{inter}}} \end{aligned} \quad (\text{S13})$$

leading to the critical occupation probability of interconnected nodes (in which the average component size diverges) given in Eqs. (8)-(9) in the main text.

III. MODEL FOR GENERATING RANDOM MODULAR NETWORKS

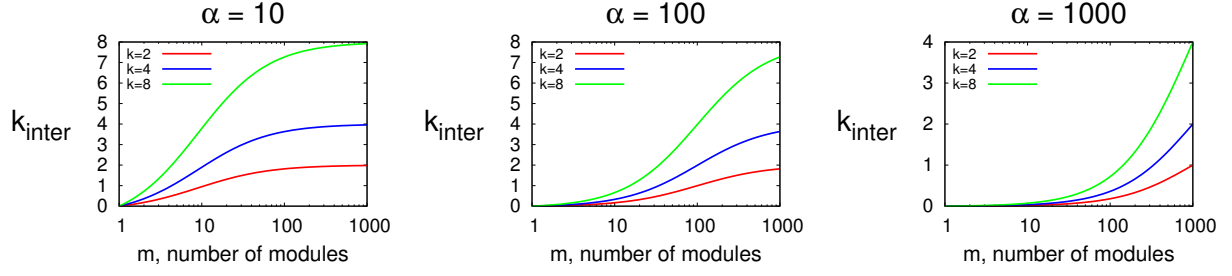


FIG. S5. **The effect of α on the convergence of k_{inter} to the mean degree.** According to the model described in the main text, k_{inter} is increasing with m taking into consideration that systems comprised of more modules have more inter-links accordingly. At the limit of large m , k_{inter} is approaching the mean degree in a rate determined by α .

IV. TWO PERCOLATION REGIMES

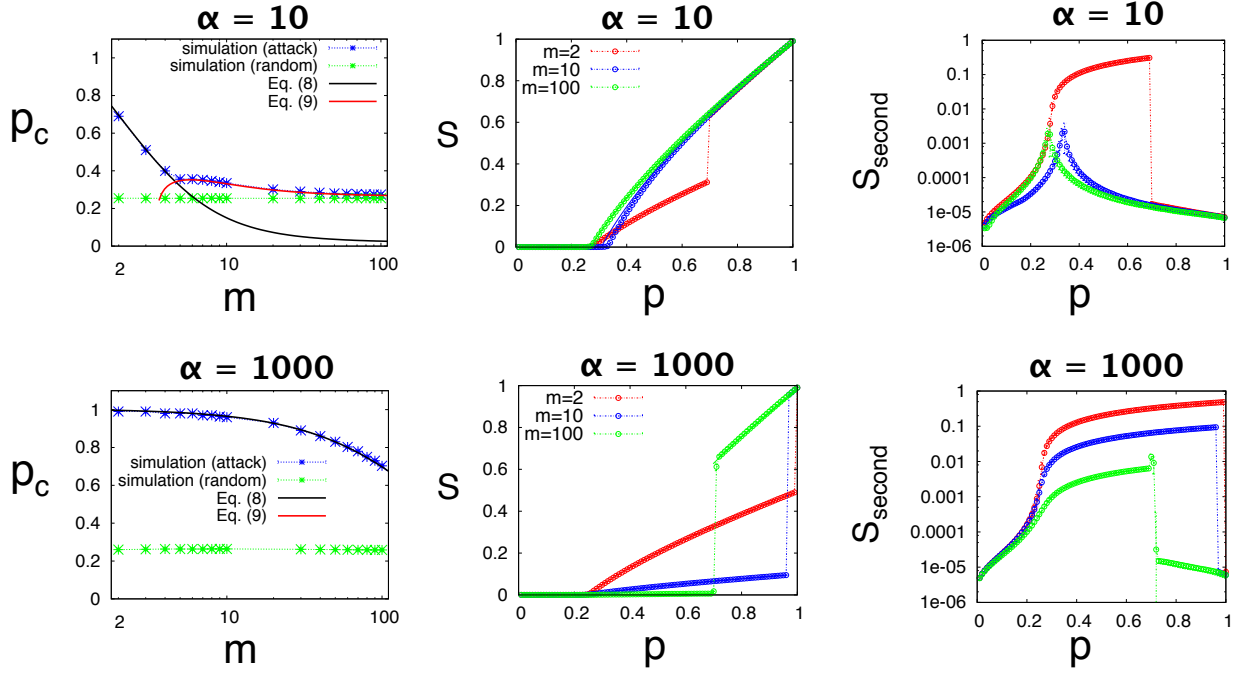


FIG. S6. **Two percolation regimes when attacking interconnected nodes.** Here we show the two percolation regimes discussed in the main text for $\alpha = 10$ and $\alpha = 1000$. Results shown are for networks with fixed mean degree $k = 4$. As discussed in the main text, the regime of $m < m^*$, corresponds to $q_c = 0$, is characterized by an abrupt first order transition, while for $m > m^*$ we observe a regular second order percolation transition characterized by the continuous decrease of S and the sharp peak in S_{second} .

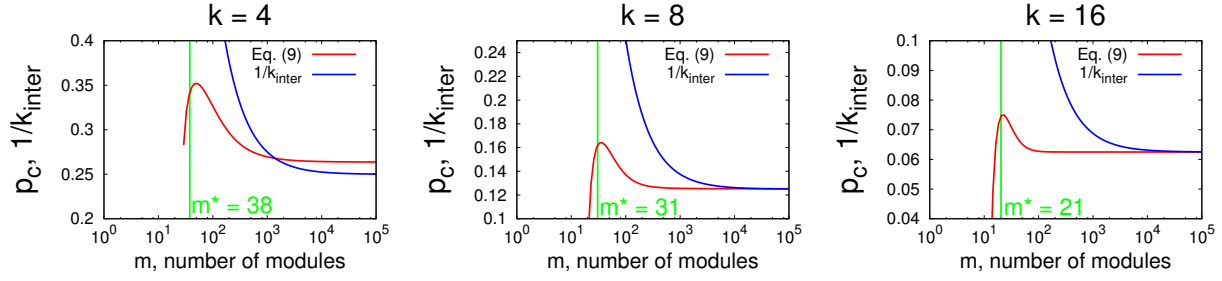


FIG. S7. **Convergence of p_c to $1/k_{\text{inter}}$.** Analytical results for $\alpha = 100$. Red lines represent p_c obtained from Eq. (9) in the main text, blue lines shows $1/k_{\text{inter}}$ and green lines shows m^* . Since unlike random node removal, nodes with no links are never removed in the attack, p_c is converging to $1/k + e^{-k}$. Here we show that indeed for higher degrees than shown in the main text, our model is converging to the percolation threshold of random removal as the number of modules increases.

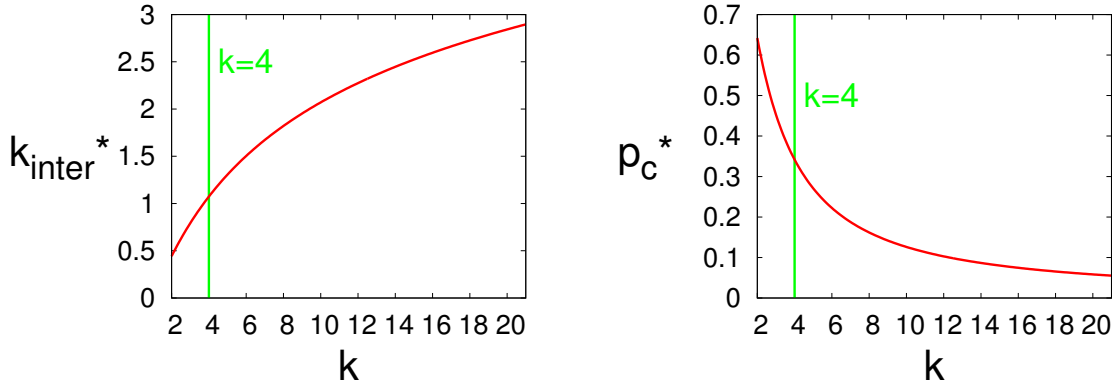


FIG. S8. **Changing of p_c^* and k_{inter}^* as the mean degree increases.** The percolation threshold p_c^* and the mean inter-degree k_{inter}^* at the transition point between the two regimes (where the two solutions for p_c cross) as a function of k . In the main text we show results for $k = 4$ where $p_c^* \approx 0.3417$ and $k_{\text{inter}}^* \approx 1.0738$ (vertical lines).

V. MODULES STRUCTURE IN THE SECOND LARGEST CLUSTER

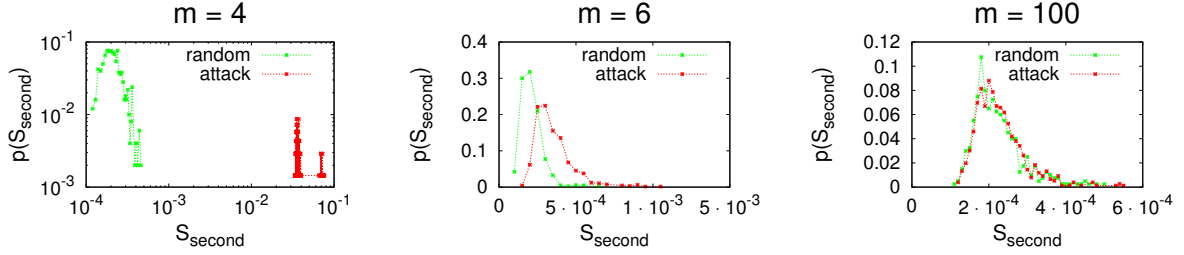


FIG. S9. **Size of the second largest cluster at $S=0.1$.** Distribution of the size of the second largest cluster at $S = 0.1$ for $\alpha = 10$. The size of the second largest cluster is significantly larger (more than two orders of magnitude) in the case of attack for $m = 4 < m^*$, in agreement with the abrupt first order transition seen in Fig. 2. As mentioned in the main text, this is caused by large modules (i.e. modules that were not much damaged by the attack) that “dropped” from the giant component. For $m = 6 > m^*$, the attack still results in larger second clusters, in comparison to the random attack, because it contains subgraphs (i.e. modules) with more complete internal structure. As m increases, the difference in sizes disappears.

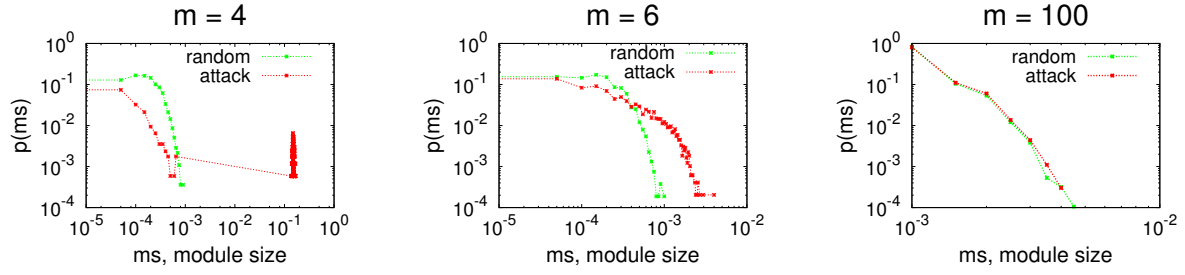


FIG. S10. **Size of modules in the second largest cluster at $S=0.1$.** Distribution of the sizes of modules in the second largest component at $S = 0.1$ for $\alpha = 10$. Modules sizes are normalized by the initial module size. Here we can see that the differences in the size of the second cluster discussed above (Fig. S5), are indeed originate at big modules that “dropped” from the largest component. These modules are getting smaller as m increases since for a large number of modules, the number of interconnected nodes is large, and removing them is damaging the internal structure of modules, just like random node failure.

VI. BETWEENNESS CENTRALITY OF INTERCONNECTED NODES IN MODULAR STRUCTURES

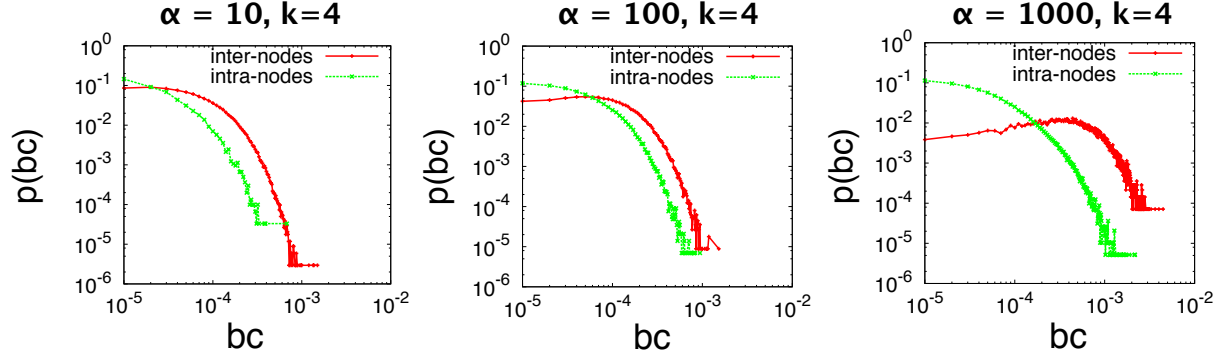


FIG. S11. **Betweenness centrality of interconnected nodes for various α .** Betweenness centrality of inter-nodes (nodes that have at least one interconnection) and intra-nodes (nodes with only intraconnections) in networks of size $N = 100\,000$ with $m = 10$ modules, mean degree $k = 4$ and $\alpha = 10, 100, 1000$ respectively. This figure further illustrates the point that in modular structures the interconnected nodes have high betweenness centrality.

[1] E. A. Leicht and R. M. D'Souza, (2009), arXiv:0907.0894.