Supporting Information

Parshani et al. 10.1073/pnas.1008404108

SI Text

General Model for Dependency Groups Larger than 2. Our manuscript presents an analytical analysis of a network containing both connectivity links and dependency groups of size $s = 2$. However, our results describing the fundamental increase in the network vulnerability due to the existence of cascading failures that lead to a first-order phase transition, is a general property that occurs in networks containing dependency groups. Moreover, for networks with larger dependency groups the fundamental change in the network behavior becomes even more extreme. Fig. S1 presents simulation results of the giant cluster size, α_∞, as a function of p for increasing sizes of dependency groups. For larger dependency groups, the first-order transition occurs for larger values of p indicating that the network is more vulnerable.

Formalism for General Case ($0 \le q \le 1$). The formalism for the general case of $0 \le q \le 1$ is significantly more involved. After an initial removal of a fraction $1-p$ of the network nodes the remaining part of the network is $\beta_0 = p$. The initial removal will cause additional nodes to disconnect from the giant cluster (percolation process). The remaining functional part of the network after the percolation process is $\alpha_1 = g(\beta_0)\beta_0$. Next, all nodes that depend on the fraction, $1-\alpha_1$, of nonfunctional nodes, will also fail (dependency process). Because only a fraction q of the pairs cause each other to fail, the fraction of new nodes that will fail due to dependencies is $\delta_1 = q(1-\alpha_1)\alpha_1$ (we multiply by α_1 because we are interested in the fraction of new nodes that fail, i.e., the fraction of nodes that fail from α_1), which is equivalent to a new random removal of $r_1 = (1-\beta_0) + q(1-\alpha_1)\beta_0$ from the original network. To understand why, note that when reducing the problem at step n to an equivalent random removal of a fraction, r_n, from the original network we are assuming a percolation process has not yet occurred. Thus, the fraction of nodes depending on the $(1-\alpha_n)$ failing nodes must not be calculated from the remaining part α_n after percolation but rather from the remaining part prior to the percolation process β_{n-1}. To that we must add $1-\beta_{n-1}$ which is the fraction of nodes removed in the previous random removal.

The remaining fraction of nodes after the new random removal is $\beta_1 = 1 - r_1 = p^2 q g(\beta_0) + p(1-q)$. After performing again the percolation phase, the remaining functional part of the giant component is $\alpha_2 = \beta_1 g(\beta_1)$. The next step involves calculating the fraction of nodes that failed in the percolation process and will cause the nodes that depend on them to fail. However, compared to the original network, this fraction $(\alpha_1 - \delta_1) - \alpha_2$ includes a higher density of nodes that cannot cause other nodes to fail (their pairs were removed on the previous stage). To overcome this problem we present an equivalent representation which also defines, $\delta_1' = (1-\alpha_1)\alpha_1$, a quantity that takes into account not only $\delta_1 = q(1-\alpha_1)\alpha_1$, the fraction of new nodes that fail due to dependencies, but also the fraction $(1-q)(1-\alpha_1)\alpha_1$ that do not have pairs. Accordingly, we define $r_1' = (1-\beta_0) + (1-\alpha_1)\beta_0$, $\beta_1' = p^2 g(\beta_0)$, and $\alpha_2' = \beta_1' g(\beta_1)$, respectively. Now, instead of $(\alpha_1 - \delta_1) - \alpha_2$, that includes a higher density of nodes that cannot cause other nodes to fail, we define $[\alpha_1 - (1-\alpha_1)\alpha_1] - \alpha_2'$ which has the original fraction q of dependent pairs. Accordingly, the fraction of nodes that are disconnected due to dependencies is $\delta_2 = q(\alpha_2'/\alpha_1^2)(\alpha_1^2 - \alpha_2') = q[1 - (\alpha_2'/\alpha_1^2)]\alpha_2'$, which is equivalent to an initial removal of $r_2 = (1-\beta_1) + q[1 - (\alpha_2'/\alpha_1^2)]\beta_1'$ from the original network. The remaining fraction of nodes is therefore $\beta_2 = 1 - r_2 = \beta_0^2 g(\beta_1) + \beta_0(1-q) = p^2 q g(\beta_1) + p(1-q)$. Following this approach, we obtain the sequence

$$\beta_0 = p.$$
$$\beta_1 = qp^2 g(\beta_0) + p(1-q).$$
$$\beta_2 = qp^2 g(\beta_1) + p(1-q)\ldots$$
$$\beta_m = qp^2 g(\beta_{m-1}) + p(1-q).$$

Note that when substituting $q = 1$, we obtain $\beta_m = p^2 g(\beta_{m-1})$, the specific solution presented in the manuscript for $q = 1$.

Number of Stages in the Iterative Process of Cascading Failures. We show that at the first-order region, above the transition point p^I ($p > p^I$) the average number of stages $\langle n \rangle$, in the iterative process of cascading failures scales as $\langle n \rangle \sim \ln N/\sqrt{p - p^I}$, below p^I ($p < p^I$), $\langle n \rangle \sim 1/\sqrt{p^I - p}$ and at $p = p^I$, $\langle n \rangle \sim N^{1/4}$. Thus as the distance from the transition is increased, the number of stages sharply drops providing a useful method for identifying the critical value of p^I at the first-order transition.

For $p > p^I$, near the first-order transition, Eq. **1** (in the main paper) has two roots produced by the intersection of the curved line which can be approximated by a parabola $y = a(p)x^2 + b(p)x + c(p)$ and a straight line $y = x$. The solution can also be obtained by solving a quadratic equation

$$a(p)x^2 + [b(p) - 1]x + c(p) = 0. \qquad \textbf{[S1]}$$

The value $p = p^I$ is given by the discriminant of this equation equal to zero: $d(p^I) \equiv (b(p^I) - 1)^2 - 4a(p^I)b(p^I) = 0$. In the general case, all three parameters, $a(p)$, $b(p)$, and $c(p)$, have nonzero derivatives at $p = p^I$. Therefore, in the general case $d(p)$ has also a nonzero derivative at $p = p^I$, and hence the difference between the roots scales as $\sqrt{p - p^I}$. Thus, the derivative of the curve at the largest root, which corresponds to the limit of the iterative process scales as $f' = 1 - \alpha\sqrt{p - p^I}$, where α is some positive constant. For Eq. **1** (in the main paper) the iterations converge to the root as $f_n' = \exp(-\alpha\sqrt{p - p^I}n)$. In a real network, they will stop when the difference between two successive iterations will be smaller than one node, which yields a condition $\exp(-\alpha\sqrt{p - p^I}n) \sim 1/N$. Hence indeed $\langle n \rangle \sim \ln N/\sqrt{p - p^I}$.

For $p < p^I$, the solution does not exist and the curve misses the line with the distance proportional to the negative discriminant. As the curve comes close to the line, the steps are proportional to $(x - x_c)^2 + d$, where $d \sim p^I - p$ is the minimal distance between the curve and the line. The number of such steps per dx is $dx/[(x - x_c)^2 + d]$. The total number of steps are thus the integral of this quantity between $x = p$ and $x = 0$, which in the limit $d \to 0$ gives $\langle n \rangle = \pi/\sqrt{d} \sim 1/\sqrt{p - p^I}$.

Exactly at the critical point $p = p^I$ the straight line touches the curve at a single point and the sequence of iterations converges as

$$x_{n+1} - x_c = x_n - x_c - a(x_n - x_c)^2. \qquad \textbf{[S2]}$$

These iterations converge to x_c as $1/n$ which can be seen by plugging into this equation $x_n - x_c = C/n^\beta + o(n^{-\beta})$ where C and β are some unknown constants. Expanding $(n+1)^{-\beta}$ in Taylor series and equating coefficients for equal powers, one can see that $\beta = 1$. However, for a finite network, the critical fraction of remaining nodes μ_a has a Gaussian spread and thus the distance from criticality scales as $(\mu_a - p^I) \sim 1/\sqrt{N}$. Therefore the distributions of the number of stages in the cascade has an exponential

tail $\exp[-\alpha n\sqrt{\mu_a - p^I}]$, in which $(\mu_a - p_c^I)$ must be replaced by its typical value $1/\sqrt{N}$. Accordingly, the distribution of $P(n)$ will have an exponential tail $P(n) \sim \exp[-\alpha' n/N^{1/4}]$, where α' is some positive constant. Thus at criticality, we expect that $\langle n \rangle \sim N^{1/4}$ as supported by our simulations (Fig. S3).

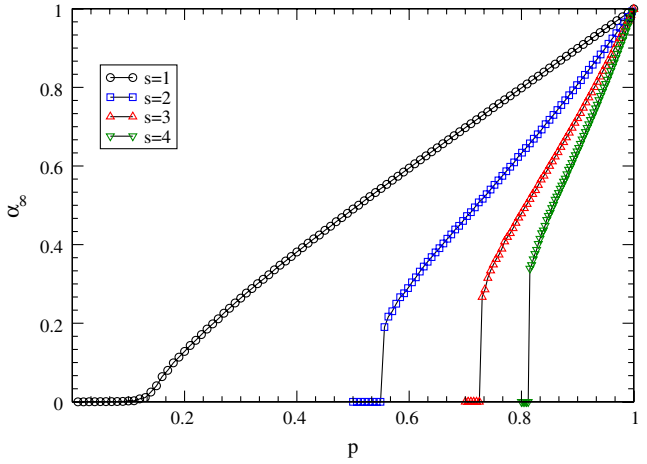

Fig. S1. Simulation results of the giant cluster size, α_∞, as a function of p for increasing sizes of dependency groups. For larger dependency groups, the first-order transition occurs for larger values of p indicating that the network is more vulnerable.

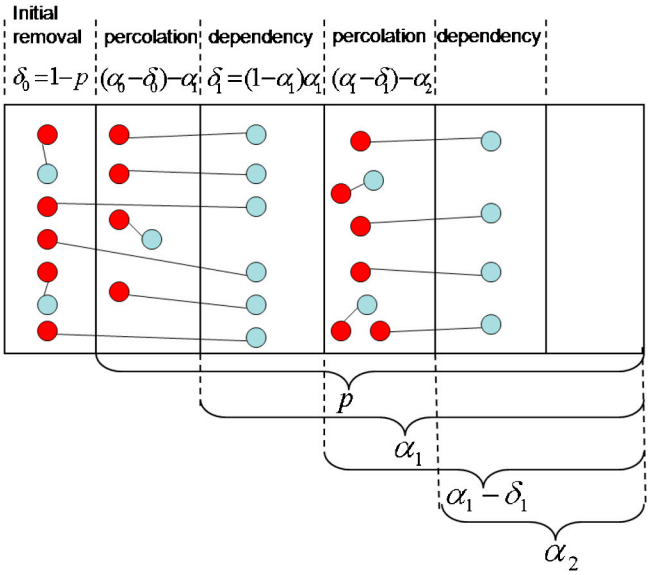

Fig. S2. A schematic illustration of the iterative cascade process for the case of $q = 1$. Each step is composed of two phases: (*i*) the percolation phase on which nodes fail due to the percolation process, and (*ii*) the dependency phase on which nodes that depend on nodes that have already failed also fail. One node from each pair of nodes connected by a dependency link is marked in red and the other in gray. At each stage, α_n is the remaining fraction after the percolation process ($\alpha_0 = 1$), and δ_n is the fraction removed due to dependencies (see manuscript for more details). The fraction of nodes failing on each step are shown above the graphical description of the system and the fraction of remaining nodes after each step is presented below the graphical description.

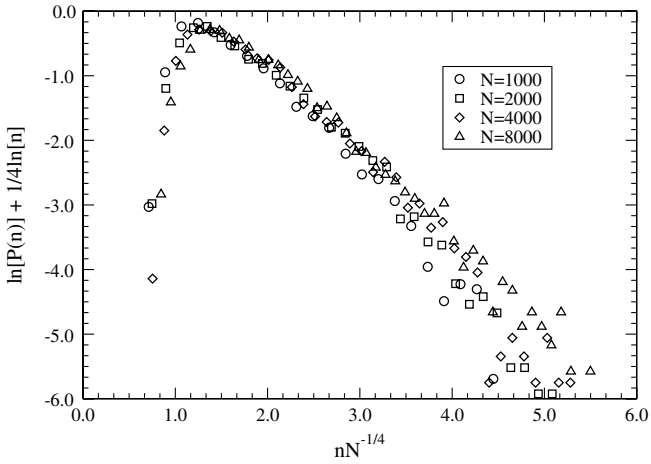

Fig. S3. Scaled distribution of the number of stages in the cascade failures for Erdős–Rényi network with $\langle k \rangle = 3$ at criticality, for different values of N.