Critical effect of dependency groups on the function of networks

Roni Parshani*1, Sergey V. Buldyrevb, and Shlomo Havlin*

*Minerva Center and Department of Physics, Bar-Ilan University, Ramat Gan 52900, Israel; and Department of Physics, Yeshiva University, 500 West 185th Street, New York, New York 10033

Edited* by H. Eugene Stanley, Boston University, Boston, MA, and approved November 16, 2010 (received for review June 28, 2010)

Current network models assume one type of links to define the relations between the network entities. However, many real networks can only be correctly described using two different types of relations. Connectivity links that enable the nodes to function cooperatively as a network and dependency links that bind the failure of one network element to the failure of other network elements. Here we present an analytical framework for studying the robustness of networks that include both connectivity and dependency links. We show that a synergy exists between the failure of connectivity and dependency links that leads to an iterative process of cascading failures that has a devastating effect on the network stability. We present exact analytical results for the dramatic change in the network behavior when introducing dependency links. For a high density of dependency links, the network disintegrates in a form of a first-order phase transition, whereas for a low density of dependency links, the network disintegrates in a second-order transition. Moreover, opposed to networks containing only connectivity links where a broader degree distribution results in a more robust network, when both types of links are present a broad degree distribution leads to higher vulnerability.

percolation | critical phenomena | complex networks | scale free networks | Erdős–Rényi networks

Many friendships between individuals in a social network, numerous business connections in a financial network, or multiple cables between Internet routers are all examples of networks with a high density of connectivity links (1–11). Such networks are regarded as very stable to attacks because, even after a failure of many nodes, the network still remains connected. In contrast, dependencies between the network nodes endanger the network stability because the failure of several nodes may lead to the immediate failure of many others. As an example, consider a financial network: Each company has trading and sales connections with other companies (connectivity links). These connections enable the companies to interact with each other and function together as a global financial market. But there are also dependencies between companies; several companies that belong to the same owner depend on one another. If one company fails, the owner might not be able to finance the other companies that will fail too. Such dependencies jeopardize the network stability and are the possible cause of many major financial crises. Another example is an online social network (Facebook or Twitter): Each individual communicates with his friends (connectivity links), thus forming a social network through which information and rumors can spread. However, many individuals will only participate in a social network if other individuals with common interests also participate (dependency links) in that social network.

The effect of failing nodes on the network stability has been studied separately for networks containing only connectivity links (12–17) and for networks containing only dependency links (18–22). The fundamental difference between connectivity and dependency links is that for dependency links when a node fails, his direct neighbors also fail (with some probability), but for connectivity links a node fails only when it (or the cluster it is in) becomes completely disconnected from the network. Percolation theory is a major tool for studying the stability of networks connected only by connectivity links. In a percolation process on a network of size $N$, a fraction $1 - p$ of the network nodes are removed. If the remaining fraction of nodes, $p$, is larger than a critical value ($p > p_c$), a spanning cluster connecting order $N$ nodes exists; if, however, $p < p_c$, the network collapses into small clusters. At $p = p_c$, the network undergoes a second-order phase transition (12–17).

Previous studies of networks containing dependencies can be divided into two categories: (i) Failures due to overloads in networks containing a flow of a physical quantity. For example, disturbances in power transmission systems or congestion instabilities in transportation networks and Internet traffic (18–21). These models show that when one node is overloaded and the traffic cannot be routed through it, choosing alternative paths will cause other nodes to also become overloaded. This process may develop into a series of cascading failures that can disable the entire network. (ii) Models that are based on local dependencies, such as decision making of interacting agents (22). In these models the state of a node depends on the state of its neighbors and therefore a failing node will cause its neighbors to also fail and so on.

Results

Here we present an analytical framework for studying the robustness of networks that include both connectivity and dependency links. When nodes fail in a network containing both types of links, two different processes occur. (i) Connectivity links are disconnected, causing other nodes and clusters to disconnect from the network (percolation process). (ii) Failing nodes cause other nodes that depend on them to also fail even though they are still connected to the network via connectivity links (dependency process). We show that the synergy between the percolation process and the dependency process leads to a cascade of failures that can fragment the entire network (Fig. 1). We find that the density of dependency links, $q$, plays a key role in determining the robustness of such networks. For networks containing connectivity links and a high density of dependency links, an initial failure of even a small fraction of the network nodes disintegrate the network in a form of a first-order phase transition. If, however, the fraction of dependency links is reduced below a certain threshold, $q_c$, the network disintegrates in a form of a second-order phase transition.

The cascading process leading to a first-order transition exists for a wide range of topologies including lattices, Erdős–Rényi (ER), and scale free (SF) networks, indicating it is a general property of many networks (Fig. 2A). Comparing networks with both connectivity and dependency links but with different topologies,

*Author contributions: R.P. and S.H. designed research; R.P. and S.V.B. performed research; R.P. analyzed data; and R.P., S.V.B., and S.H. wrote the paper.

The authors declare no conflict of interest.

This Direct Submission article had a prearranged editor.

To whom correspondence should be addressed. E-mail: parshani.roni@gmail.com.

This article contains supporting information online at www.pnas.orglookup/suppl/doi:10.1073/pnas.1008404108/-/DSSupplemental.
transitions in lattice, ER, and SF networks. The fraction of nodes in the giant
$q = 1$ for ER (squares) and SF (circles) networks with the same average degree
(dashed line) are plotted as a function of $q$. The process starts with the initial failure of two nodes (marked in red). The connectivity links connected to them also fail (marked in red). (B) Percolation process—in this stage all the nodes and the connectivity links that are connected to them, that are not connected to the giant cluster (largest cluster) by connectivity links, also fail (marked in red). (C) Dependency process—the nodes that depend (connected by dependency links) on the failing nodes also fail (marked in red). (D) The next step of connectivity failure in which two more nodes fail because they are not connected to the largest cluster (currently containing only two nodes).

reveals a relation between topology and the robustness to random failure: Networks with a broader degree distribution of connectivity links are more vulnerable to random failure in the presence of dependency links. This relation is opposed to the known result for networks containing only connectivity links, where networks with a broader degree distribution are significantly more robust to random failures. Fig. 2 A and B show that, when comparing ER and SF networks with the same average degree, SF networks with a high density of dependent nodes are more vulnerable to random failures than ER networks.

**Formalism**

Next we present an analytical approach for studying the robustness to random failure of networks containing the two types of links. Without loss of generality, we define a model in which only pairs of nodes depend on one another, forming dependency groups of size two. When the dependency group contains more than two nodes, the cascade effect is even more extreme and the transition from the regular second-order percolation transition to a first-order transition occurs for more stable networks (SI Text). Therefore, the properties we present for the case of dependency groups of size two are also valid in the general case of larger dependency groups (see SI Text). The model is defined as follows: A network containing $N$ nodes is randomly connected by connectivity links with a degree distribution $P(k)$ and an average degree $\langle k \rangle$. In addition, pairs of nodes are connected by dependency links as follows: (i) A node can only have one dependency link. (ii) If node $i$ depends on node $j$, then node $j$ depends on node $i$. For this model, we denote by $q$ the fraction of nodes that have dependencies.

We start by presenting the formalism describing the iterative process of cascading failures for the simple case of $q = 1$ (see SI Text). Each iteration (step) includes failures that are the result of the percolation process and failures that are the result of the dependency process. The goal of the formalism is to describe the accumulated process up to step $n$ as an equivalent single random removal, $r_n$, from the original network. The remaining fraction of nodes after such a removal is $\beta_n = 1 - r_n$. The new network, after the removal of a fraction $r_n$ of the nodes, has a giant component consisting of a fraction $g(\beta_n)$ of the remaining nodes which is a fraction $\alpha_n = \beta_n g(\beta_n)$ from the original network. The iterative process is initiated by the removal of a fraction $r_0 = 1 - p$ of the network nodes. The remaining part of the network is $\beta_0 = p$. This initial removal will cause additional nodes to disconnect from the giant cluster due to the percolation process. The fraction of nodes that remain functional after the percolation process is $\alpha_1 = \beta_0 g(\beta_0)$. Each node from the nonfunctional part ($1 - \alpha_1$) will cause the node that depends on it to also fail (dependency process). The probability that a node depending on a nonfunctional node has survived until now is $\alpha_1$. Therefore the fraction of new nodes that will fail due to dependencies is $\delta_1 = (1 - \alpha_1) \alpha_1$. The accumulated failure, including the initial failure of $1 - \beta_0$ and $\delta_1$, is equivalent to a random removal of $r_1 = (1 - \beta_0) + (1 - \alpha_1) \beta_0$ from the original network (see SI Text). The remaining fraction of nodes after the new removal is therefore $\beta_1 = 1 - r_1 = \beta_0 (1 - \delta_1) = \beta_0 g(\beta_0)$. The remaining functional part of the giant component is now $\alpha_2 = \beta_1 g(\beta_1)$. To calculate the fraction $\delta_2$ of nodes that are disconnected due to dependencies at the second stage, recall that at the previous stage a fraction $\delta_1$ failed from $\alpha_1$. The remaining part of $\alpha_1$ was therefore $\alpha_1 - \delta_1 = \alpha_1^2$. Thus $\delta_2 = (\alpha_2 / \alpha_1^2) (\alpha_1^2 - \alpha_1) = (1 - (\alpha_2 / \alpha_1^2)) \alpha_1$, which is equivalent to a random removal of $r_2 = (1 - \beta_2) + (1 - \alpha_2 / \alpha_1) \beta_1$ from the original network. The remaining fraction of nodes is $\beta_2 = 1 - r_2 = \alpha_2 \alpha_1^2 / \beta_1 = \beta_2 g(\beta_1)$. Following this approach, we can construct the sequence, $\beta_n$, of the remaining fraction of nodes in the network after each iteration.

$$
\beta_0 = p.
\beta_1 = p^2 g(\beta_0).
\beta_2 = p^2 g(\beta_1) = \beta_2 g(\beta_2).
\beta_0 = \beta_0 g(\beta_0).
$$

Following a similar approach for the general case of $0 \leq q \leq 1$ (see SI Text) yields the sequence $\beta_n = q p^n g(\beta_{n-1}) + p(1 - q)$. Given, $\beta_n$, the fraction of nodes in the giant cluster is
\(\alpha_{n+1} = \beta_n g(\beta_n) = p(1 - q(1 - p(g(\beta_n))))g(\beta_n).\) Fig. 3A compares theory and simulations of \(\alpha_n\), for the case of an ER network.

To determine the state of the system at the end of the cascade process, we analyze \(\beta_n\) at the limit of \(n \to \infty\). This limit must satisfy the equation \(\beta_\infty = \beta_{n+1}\) because, at the end of the process, the cluster is not further fragmented. Denoting \(\beta_n = \beta_{n-1} = x\) we arrive at the equation

\[
x = p^2 q g(x) + p(1 - q) .
\]  

[1]

This equation can be solved graphically as the intersection of a straight line \(y = x\) and a curve \(y = p^2 q g(x) + p(1 - q)\). When \(p\) is small enough, the curve increases very slowly and does not intersect with the straight line (except at the origin which corresponds to the trivial solution). The critical case for which the nontrivial solution emerges, corresponds to the case when the line touches the curve at a single point \(x\) and in this point we have the condition \(1 = p^2 q x g(\beta)\), which together with Eq. 1 gives the solution for the critical fraction of failing nodes that will fragment the network and the critical size of the giant component.

**Analytical Solution**

An exact analytical solution can be obtained using the apparatus of generating functions. As in refs. 23-25, we introduce the generating function of the degree distribution \(G_\theta(\xi) = \sum \theta^k G(\theta)\). Analogously, we also introduce the generating function of the underlying branching process, \(G_\theta(\xi) = G_\theta(\xi)/G_\theta(1).\) A random removal of a fraction \(1 - p\) of nodes will change the degree distribution of the remaining nodes, so the generating function of the new distribution is equal to the generating function of the original distribution with the argument \(\xi\) replaced by \(1 - p(1 - \xi)\) (23).

The fraction of nodes that belong to the giant component after the removal of \(1 - p\) nodes is \(g(p) = 1 - \sum \theta^k \left(1 - p(1 - \xi)\right)\), where \(f = f(p)\) satisfies a transcendental equation \(f = G_\theta(1 - p(1 - f))\) (25).

In the case of an ER network with a Poisson degree distribution (12-14), the problem can be solved explicitly because \(G_\theta(\xi) = G_\theta(\xi) = \exp(\xi(\xi - 1))\). Accordingly, \(g(x) = 1 - f\) and \(f = \exp(\langle k \rangle x - 1)\) where \(x\) is defined in Eq. 1. The fraction of nodes in the giant component at the end of the cascade process is then given by \(\alpha_\infty = \beta_\infty g(\beta_\infty) = p(1 - q(1 - p(1 - f)))(1 - f)\).

\[\alpha_\infty = \beta_\infty g(\beta_\infty) = p(1 - q(1 - p(1 - f)))(1 - f)\]

The equation \(f = f(q, p, k)\) has a trivial solution at \(f = 1\). The nontrivial solutions of \(f\) can be presented by the crossing points of the two curves in a system of equations that are given with respect to \(x\) and \(f\):

\[
\begin{align*}
x &= p^2 q(1 - f) + p(1 - q) \\
f &= \frac{\ln f}{\ln (1 - \alpha)} .
\end{align*}
\]

For the trivial solution at \(f = 1\), the size of the giant component is zero (\(\alpha_\infty = 0\)). For the solutions that are the crossing points of the two curves, \(f < 1\), i.e., \(\alpha_\infty > 0\). Thus, the case where the curves tangentially intersect corresponds to a first-order phase transition point \((p = p^f)\) where \(\alpha_\infty\) abruptly jumps from a finite size above \(p^f\) to zero below \(p^f\) (26). The condition for the first-order transition is that the derivatives of the equations of system [2] with respect to \(f\) are equal. Together with system [2] this yields

\[
(p^f)^2 < \frac{q}{1 - q/c} + \ln f(1 - 1)^2 .
\]

However, for a solution of system [2] where \(f \to 1\) (\(\alpha_\infty = 0\)) there is no jump in the size of the giant cluster and thus the transition is a second-order transition \((p = p^s)\). Solving system [2] for \(f \to 1\) yields

\[
p^s = \langle k \rangle \left(1 - \frac{1}{1 - q/c}\right) = 1 .
\]

The analysis of Eqs. 3 and 4 shows that the first-order transition at \(p = p^f\) occurs for networks with a high density of dependency links \((q > q_c)\), whereas the second-order transition at \(p = p^s\) occurs for networks with a low density of dependency links \((q < q_c)\). This behavior is confirmed by Fig. 4 which compares theory and simulations for \(p^f(q)\) and \(p^s(q)\). The critical value of \(q_c\) and \(p_c\) for which the phase transition changes from first order to second order is obtained when the conditions for both the first- and second-order transitions are satisfied simultaneously. Applying both conditions, we obtain

\[
\begin{align*}
q_c &= \langle k \rangle + 1 + \frac{1}{2(\langle k \rangle + 1)} \\
p_c &= 1 + \frac{1}{1 + 1}.
\end{align*}
\]

**Simulations**

Next, we support our analytical results by simulations. Finding the transition point via simulations is always a difficult task that
requires high precision. In the case of the first-order transition, we are able to calculate the transition point with good precision by identifying the special behavior characterizing the number of iterations (NOI) in the cascading process. At the first-order transition point, the NOI scales as \( N^{1/4} \) (see SI Text), which is also demonstrated by the long plateau in Fig. 3A. This number sharply drops as the distance from the transition point is increased, because away from the transition point, \( p' \), the NOI scales as \( \log N/(p - p') \) (see SI Text). Thus, plotting the NOI as a function of \( p' \) provides a useful and precise method for identifying the transition point \( p' \) at the first-order region. For the second-order region, a similar behavior exists for the size of the second largest cluster which also reaches its maximum at its transition point (17). Fig. 3B presents simulation results of the NOI. The transition point, \( p' \), can easily be identified by the sharp peak characterizing the transition point. Fig. 3B, Inset, presents a similar behavior for the size of the second largest cluster near the second-order transition point, \( p'' \). Fig. 4A compares simulation results and theory for the transition points \( p'(q) \) at the first-order region (solid line) and \( p''(q) \) at the second-order region (dashed line). The transition points were obtained using the NOI and the second cluster size techniques, respectively. The theoretical results for different values of \( q \) and \( (k) \) were calculated by solving system [2] together with Eqs. 3 or 4, respectively. Fig. 2B illustrates the values of the transition points \( p'(q) \) and \( p''(q) \), respectively, between SF and ER networks with the same average degree. For networks with a small fraction of dependencies (second-order transition region), SF networks are more robust to random failure (lower \( p'' \)). Networks with a high fraction of dependencies (first-order transition region), SF networks become more vulnerable (higher \( p' \)). Fig. 4B compares simulation and theory for \( \alpha_{\text{NOI}} \), the number of nodes in the giant cluster at the transition point. Above \( q_c' \), \( \alpha_{\text{NOI}} \) is finite characterizing a first-order transition, whereas below \( q_c' \), \( \alpha_{\text{NOI}} \) is zero as expected for a second-order transition.

**Discussion**

Here we argue that, in order to properly model real networks, two different types of links are needed: connectivity links and dependency links. We present an analytical formalism for a general network model including both connectivity and dependency links. According to our model, networks with high density of dependency links are extremely vulnerable to random failure and when a critical fraction of nodes fail the network disintegrates in a form of a first-order phase transition. Networks with a low density of dependency links are significantly more robust and disintegrate in a form of a second-order phase transition. In the limit of zero fraction of dependency links, our general solution yields the known results for networks with only connectivity links. Our framework also provides an analytical solution for the critical density of dependency links for which the phase transition changes from a first-order to a second-order percolation transition. We develop a powerful simulation method to accurately estimate the transition point, based on the unique behavior of the NOI (number of iterations in the iterative process of cascading failures) that diverges at the first-order transition point. Using this method, we are able to provide very accurate simulation results supporting our analytical results.

**ACKNOWLEDGMENTS.** We thank the European Epidemic Forecast Infrastructure Framework Project, the Israel Science Foundation, the Office of Naval Research, and the Defence Threat Reduction Agency for financial support. S.V.B. thanks the Office of the Academic Affairs of Yeshiva University for funding the Yeshiva University high-performance computer cluster and acknowledges the partial support of this research through the Dr. Bernard W. Gammson Computational Science Center at Yeshiva College.
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 takes place at a higher density of nodes that cannot cause other nodes to fail, i.e., the fraction of nodes that fail from dependencies is $\beta_q = q(1 - \alpha_1)\alpha_1$ (multiply by $\alpha_1$ because we are interested in the fraction of new nodes that fail, i.e., the fraction of nodes that fail from $\alpha_1$), which is equivalent to a new random removal of $r_1 = (1 - \beta_q) + q(1 - \alpha_1)\beta_0$ from the original network. To understand why, note that when substituting $\alpha_1 = 0$, the equations become valid only for $p > p'$.

Supporting Information

Number of Stages in the Iterative Process of Cascading Failures. We show that at the first-order region, above the transition point $p' (p > p')$ the average number of stages $<n>$, in the iterative process of cascading failures scales as $<n> \sim \ln N / \sqrt{p - p'}$, below $p'$ ($p < p'$), $<n> \sim 1 / \sqrt{p - p'}$ and at $p = p'$, $<n> \sim N^{1/4}$. 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 $p'$ at the first-order transition.

For $p > p'$, 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

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

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

For $p < p'$, 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 $d \sim p - p'$ 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 $d = 0$ gives $<n> = \pi / \sqrt{d} \sim 1 / \sqrt{p - p'}$.

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

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

These iterations converge to $x_c$ as $1/n$ which can be seen by plugging into this equation $x_{n+1} - x_n = C/n + o(1/n)$, where $C$ and $\beta$ 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 $\mu_c$ has a Gaussian spread and thus the distance from criticality scales as $(\mu_c - 1)^{-1 / \sqrt{N}}$. Therefore the distributions of the number of stages in the cascade has an exponential
tail \( \exp[-\alpha n \sqrt{\mu_a - p}] \), in which \((\mu_a - p_c)\) 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 \(\alpha'\) 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, \(\alpha_{\infty}\), 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, \(\alpha_n\) is the remaining fraction after the percolation process \((\alpha_0 = 1)\), and \(\delta_n\) 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 $N$. 