RESEARCH ARTICLE SUMMARY

NETWORK SCIENCE

Small vulnerable sets determine large network cascades in power grids

Yang Yang, Takashi Nishikawa,* Adilson E. Motter

INTRODUCTION: Cascading failures in power grids are inherently network processes, in which an initially small perturbation leads to a sequence of failures that spread through the connections between system components. An unresolved problem in preventing major blackouts has been to distinguish disturbances that cause large cascades from seemingly identical ones that have only mild effects. Modeling and analyzing such processes are challenging when the system is large and its operating condition varies widely across different years, seasons, and power demand levels.

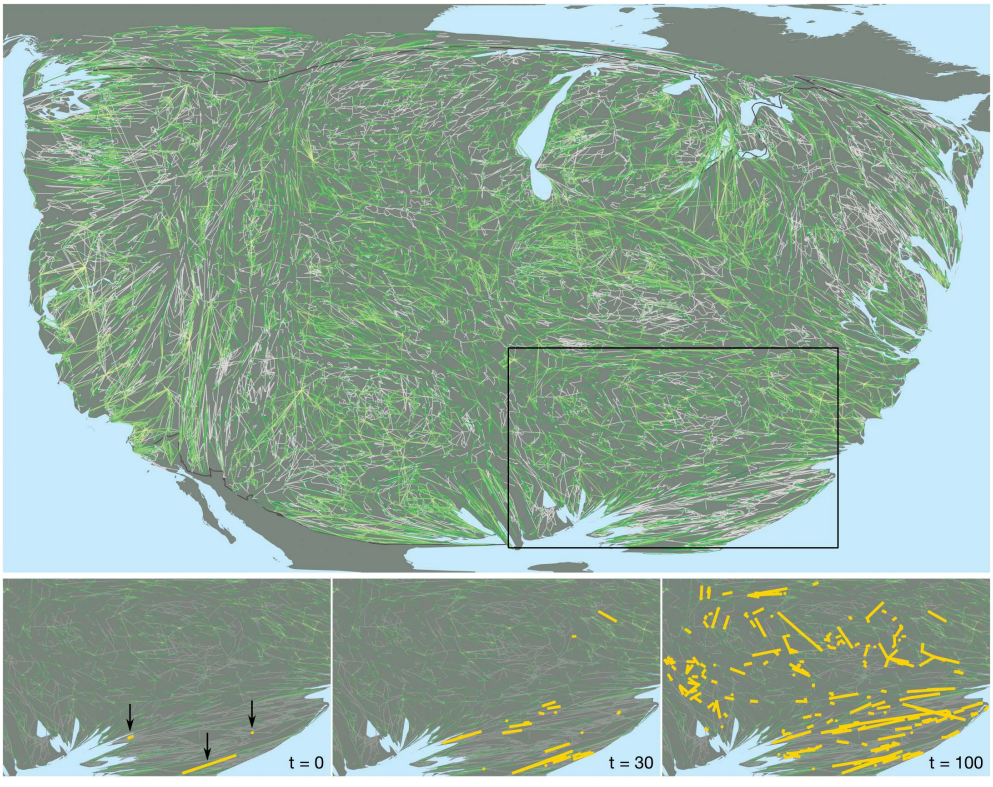
RATIONALE: Multicondition analysis of cascade vulnerability is needed to answer several key questions: Under what conditions would an initial disturbance remain localized rather than cascade through the network? Which network components are most vulnerable to failures across various conditions? What is the role of the network structure in determining component vulnerability and cascade sizes? To address these questions and differentiate cascading-causing disturbances, we formulated an electrical-circuit network representation of the U.S.–South Canada power grid—a large-scale network with more than 100,000 transmission lines—for a wide range of operating conditions. We simulated cascades in this system by means of a dynamical model that accounts for transmission line failures due to overloads and the resulting power flow reconfigurations.

RESULTS: To quantify cascade vulnerability, we estimated the probability that each transmission line fails in a cascade. Aggregating the results from multiple conditions into a single network representation, we created a systemwide vulnerability map, which exhibits relatively homogeneous geographical distribution of power outages but highly heterogeneous distribution of the underlying overload failures. Topological analysis of the network representation revealed that the transmission lines vulnerable to overload failures tend to occupy the network's core, characterized by links between highly connected nodes. We found that only a small fraction of the transmission lines in the network (well below 1% on average) are vulnerable under a given condition. When measured in terms of node-to-node distance and geographical distance, individual cascades often propagate far from the triggering failures, but the set of lines vulnerable to these cascades tend to be limited to the region in which the cascades are triggered. Moreover, large cascades are disproportionately more likely to be triggered by initial failures close to the vulnerable set.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aan3184

CONCLUSION: Our results imply that the same disturbance in a given power grid can lead to disparate outcomes under different conditions—ranging from no damage to a large-scale cascade. The association between large cascades and the triggering failures' proximity to the vulnerable set indicates that the topological and geographical properties of the vulnerable set is a major factor determining whether the failures spread widely. Because the vulnerable set is small, failures would often repeat on the same lines in the absence of interventions. Although the power grid represents a complex system in which changes can have unanticipated effects, our analysis suggests failure-based allocation of resources as a strategy in upgrading the system for improved resilience against large cascades. ■



Cascade-resistant portion of the U.S.–South Canada power grid. The network is visualized on a cartogram that equalizes the density of nodes. (**Top**) Power lines that never underwent outage in our simulations under any grid condition are shown in green, whereas all the other lines—whose vulnerability varies widely—are in gray. (**Bottom**) Spreading of a cascade triggered by three failures at time $t = 0$ (arrows), which resulted in 254 failures at $t = 100$ (the end of the cascade in linearly rescaled time).

The list of author affiliations is available in the full article online.
*Corresponding author. Email: t-nishikawa@northwestern.edu
Cite this article as Y. Yang *et al.*, *Science* **358**, eaan3184 (2017). DOI: 10.1126/science.aan3184

RESEARCH ARTICLE

NETWORK SCIENCE

Small vulnerable sets determine large network cascades in power grids

Yang Yang,[1] Takashi Nishikawa,[1,2]* Adilson E. Motter[1,2]

The understanding of cascading failures in complex systems has been hindered by the lack of realistic large-scale modeling and analysis that can account for variable system conditions. Using the North American power grid, we identified, quantified, and analyzed the set of network components that are vulnerable to cascading failures under any out of multiple conditions. We show that the vulnerable set consists of a small but topologically central portion of the network and that large cascades are disproportionately more likely to be triggered by initial failures close to this set. These results elucidate aspects of the origins and causes of cascading failures relevant for grid design and operation and demonstrate vulnerability analysis methods that are applicable to a wider class of cascade-prone networks.

Cascading failures are inherently large-scale network processes that cannot be satisfactorily understood from a local or small-scale perspective. In blackouts caused by cascading failures in the power grid, a relatively small local disturbance triggers a sequence of grid component failures, causing potentially large portions of the network to become inactive, with costly outcomes. In the North American power grid (*1*), for instance, a single widespread power outage can inflict tens of billions of dollars in losses (*2*), and smaller but more frequent outages can amount to a yearly combined impact comparable with that of the largest blackouts (*3*). Yet, not much is known about what distinguishes disturbances that cause cascades from seemingly identical ones that do not. Despite the substantial advances made through conceptual modeling of general cascades (*4*–*10*) and physics-based modeling of power-grid–specific cascades (*11*–*14*), a major obstacle still remains: the lack of realistic large-scale models and a framework for analyzing cascade vulnerability under variable system conditions. Developing such a framework is challenging for three reasons: (i) Detailed data combining both structural and dynamical parameters are scarce, (ii) the system condition varies on a wide range of time scales, and (iii) computational resources required for modeling grow combinatorially with system size (*15*). These challenges have limited the applicability of most previous studies to vulnerability under a single condition and either to smaller scales than those required to describe large cascades or to models that are not constrained by real data. Similar hurdles exist in studying large-scale failures in the broader context of complex networks (*16*–*18*), including extinction cascades in ecological systems (*19*–*21*) and contagion dynamics in financial systems (*22*, *23*).

Here, we focus on the U.S.–South Canada power grid, which is the largest contiguous power grid amenable to modeling. This system is composed of three interconnections (Texas, Western, and Eastern) (Fig. 1A), which are separate networks of alternating current generators and power consumers connected by transmission lines (network components are illustrated in Fig. 1B). To study this system, we used the data reported in the Federal Energy Regulatory Commission (FERC) Form 715. For each interconnection, the data represent various snapshots of the system, spanning the years 2008 to 2013 and covering multiple seasons as well as both on- and off-peak demand levels, which correspond to different operating conditions. Basic properties of the 46 snapshots we used are listed in table S1. A representation of each snapshot was constructed by processing the parameters of individual power-grid components, including power generation and demand as well as the capacity of transmission lines. Central to the analysis of cascade vulnerability in this system is that our approach distinguishes (i) transmission lines (or simply lines) that have become out of service and do not carry flow because of protective relay actions, equipment malfunctions, operational errors, or physical damages ("primary failures"); and (ii) lines that do not carry flow at the end of the cascades because they are de-electrified owing to the outage of other lines ("secondary failures").

Geographic layout of vulnerabilities

The vulnerability of a given transmission line ℓ can be quantified by the probability p_ℓ that the line fails in a cascade event triggered by a random perturbation to a given snapshot of a given interconnection. To estimate p_ℓ, we used a cascade dynamics model that combines key elements from previous models (*12*, *24*, *25*) to suitably account for the physics of cascading failures. In this model, the initial state of the system for the given snapshot is determined by computing the power flow over all transmission lines and transformers from the power flow equation (supplementary materials, materials and methods). The triggering perturbation was implemented through the removal of a set of n_t lines, representing line outages due to unforeseen events, such as damage to power lines caused by extreme weather and unplanned line shutdowns caused by operational errors. After this initial removal, a cascade event was modeled

[1]Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA. [2]Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60208, USA.
*Corresponding author. Email: t-nishikawa@northwestern.edu

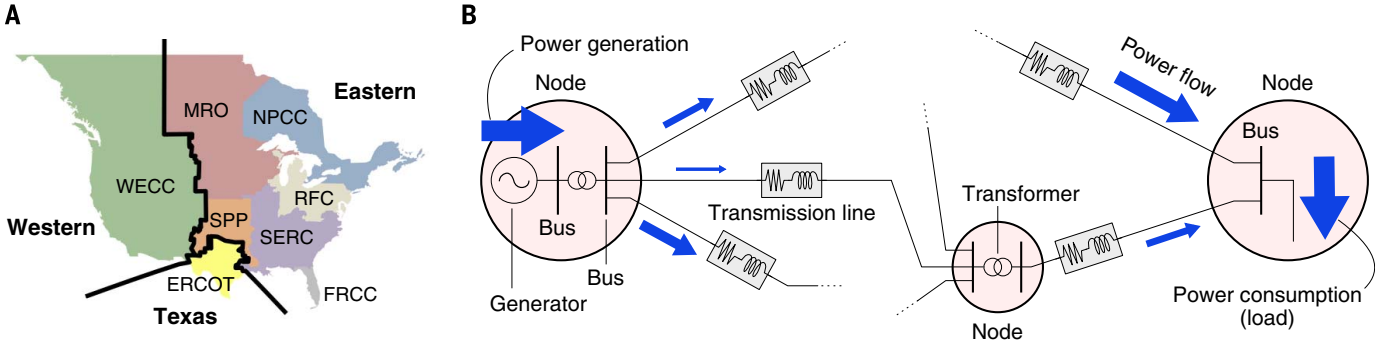


Fig. 1. The U.S.–South Canada power transmission network. (**A**) System map showing the Texas, Western, and Eastern interconnections, as well as the eight NERC regions (acronyms are defined in table S2). (**B**) Schematic diagram of a portion of a transmission network. The vertical lines and pink circles represent buses and nodes, respectively. As indicated by the blue arrows, power injected by the generators flows through this network of transmission lines and is eventually consumed at other points (where the thickness of the arrow represents the amount of power flow).

as an iterative process, with each step consisting of a single power line outage due to overheating (primary failure), followed by the redistribution of power flow in the network to compensate for the lost flow over the failed line. Line overheating was modeled with a temperature evolution equation (*12*), and flow redistribution was determined by solving the power flow equation again; if a primary line failure disconnects the network into multiple parts with unbalanced supply and demand, the power generation and consumption in each part are adjusted (similarly to how generation reserves and power shedding are used in grid operation) to allow for the subsequent power flow calculation. The failure probability p_ℓ was estimated from K such simulated cascade events, including those with no subsequent failures. Further details on the triggering perturbations and cascade dynamics model can be found in the supplementary materials, materials and methods.

We validated the model against historical line outage data available from the Bonneville Power Administration (BPA) with respect to the distribution of cascade sizes measured by the number of (primary) line failures N_f (supplementary materials, materials and methods, and fig. S1A). We also validated the extremal cascade size measured by N_f and power shed P_s (defined as the reduction in the amount of power delivered to the consumers) against the BPA data and grid disturbance data from the North American Electric Reliability Corporation (NERC), respectively (supplementary materials, materials and methods, and fig. S1, B and C). All simulations were performed with $n_t = 3$ because the cascade size distribution for a given snapshot did not differ appreciably for other choices of n_t (fig. S2). However, the distribution exhibited considerable variation across different snapshots, both when cascade size was measured by the power shed P_s (fig. S3) and when measured by the number of line failures N_f (fig. S4).

To aggregate results over different snapshots, we used a node to represent the set of all buses associated with the same geographic location across all snapshots in our data set, where the term "bus" refers to a connection point between components of a power grid, such as transmission lines, transformers, and generators (Fig. 1B). This definition of a node typically corresponds to a substation and can include generators at a nearby power plant and/or an electrical load representing local power consumption. We used a link between a pair of nodes to represent the set of all (parallel) transmission lines directly connecting the same pair of nodes in at least one snapshot, where each of these transmission lines connects two different buses (one from each node in the pair). In this network, the aggregated vulnerability $p_l := \langle p_\ell \rangle$ of a link l, which we refer to

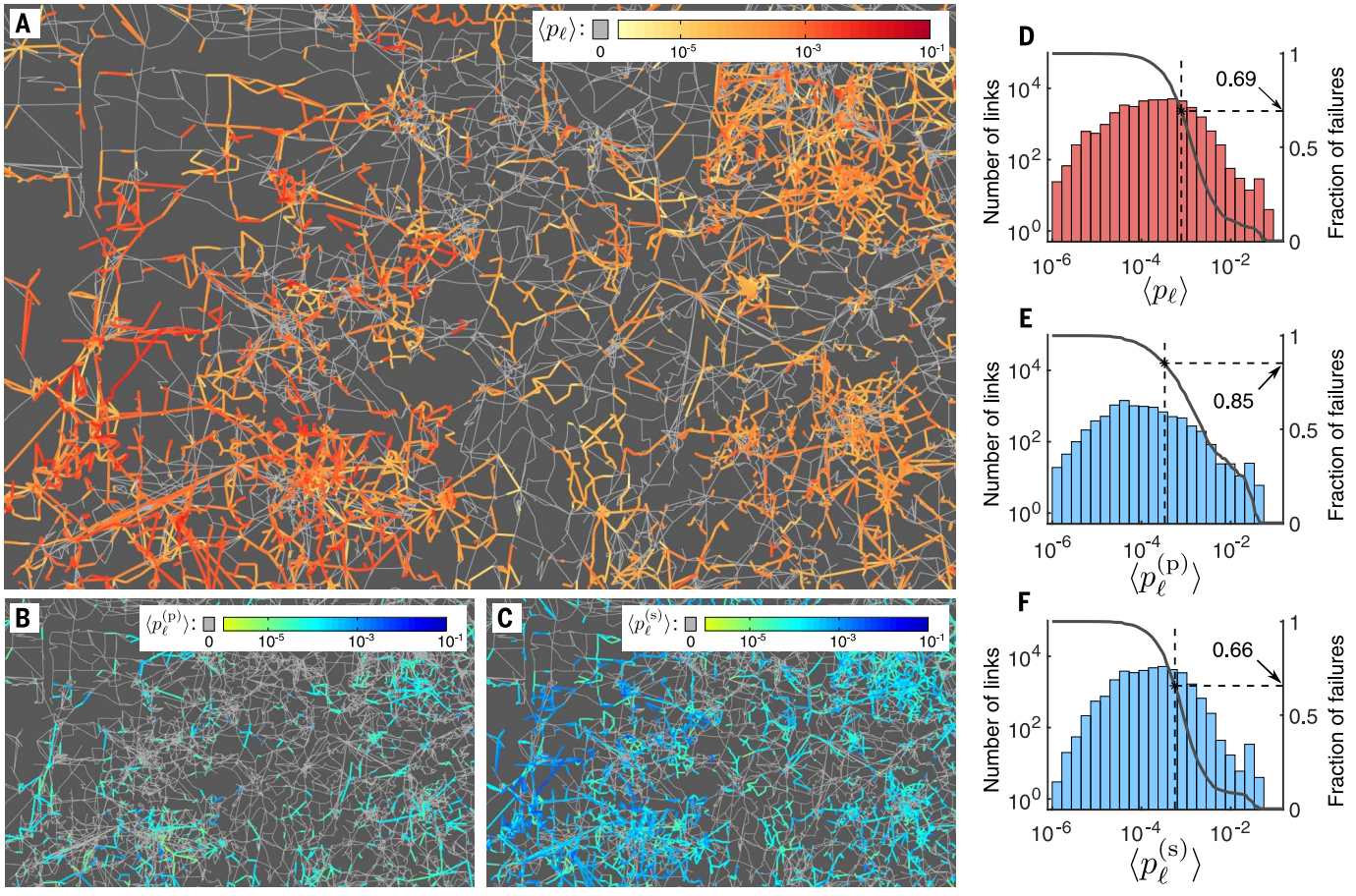


Fig. 2. Vulnerability map of the U.S.–South Canada power grid. (**A**) Averaged failure probability of transmission lines, including primary and secondary failures, expressed as the total A-vulnerability of links $\langle p_\ell \rangle$. Because the structure of the power grid varies slightly from one snapshot to another, we visualize the A-vulnerability using a single network constructed to represent all snapshots of each interconnection by regarding the set of all buses at a given geographical location as a node and all transmission lines connecting two nodes as a single link. Each link is color-coded by the failure probability $\langle p_\ell \rangle$ estimated as a weighted average over all lines in all snapshots, where gray indicates links whose estimated probability is zero. (**B** and **C**) Same as in (A), but color-coded separately for (B) the A-vulnerability to primary failures $\langle p_\ell^{(p)} \rangle$ and (C) the A-vulnerability to secondary failures $\langle p_\ell^{(s)} \rangle$. (A), (B), and (C) correspond to the same unidentified portion of the U.S.–South Canada map. (**D**) Histogram of the A-vulnerability $p_l = \langle p_\ell \rangle$ and the curve for $f(p) := \Sigma' p_l / \Sigma_l p_l$, where Σ' denotes the sum over all links l satisfying $p_l \geq p$. The function $f(p)$ thus represents the fraction of all failures that are associated with links of A-vulnerability p or larger. (**E** and **F**) Same as in (D), but with $\langle p_\ell \rangle$ replaced by $\langle p_\ell^{(p)} \rangle$ and $\langle p_\ell^{(s)} \rangle$, respectively. In (D), (E), and (F), the vertical and horizontal dashed lines indicate, respectively, the minimum A-vulnerability p^* among the most vulnerable 20% of all failing links and the fraction $f(p^*)$ of failures accounted for by these links.

as the A-vulnerability, is a weighted average of the failure probabilities over the lines represented by the link l and over the various snapshots and can be expressed as

$$\langle p_\ell \rangle = \frac{\Sigma_c \Sigma_\ell p_{\ell;c} w_c}{\Sigma_c \Sigma_\ell w_c} \quad (1)$$

where c indexes the different snapshot conditions simulated, and the sum over ℓ is limited to the set of transmission lines defining the link ℓ for the given c. Here, $p_{\ell;c}$ is the probability of line failure in the simulated perturbations of the system (the values of K we used are given in table S1 and justified in fig. S5), and w_c represents the weight assigned to each snapshot (table S1). In our analyses, we present the A-vulnerability separately for primary failures (denoted by $\left\langle p_\ell^{(\mathrm{p})} \right\rangle$), secondary failures (denoted by $\left\langle p_\ell^{(\mathrm{s})} \right\rangle$), and the combination of both primary and secondary failures (denoted by $\langle p_\ell \rangle$ itself).

We constructed the A-vulnerability map of the U.S.–South Canada power grid (shown in Fig. 2, A to C, for a portion of the grid). Over the entire network, we found that only 10.8% of all links ever underwent a primary failure in our simulations and that secondary failures were on average 3.77 times more prevalent than primary ones (table S3). We also found that A-vulnerability was very unevenly distributed among the links, with 20% of the failing links (which in the case of primary failures correspond to only 2.16% of all links) accounting for about 85, 66, and 69% of the primary failures, secondary failures, and combined set of all failures, respectively (Fig. 2, D to F). Also uneven was the geographical distribution of links with nonzero A-vulnerability (Fig. 2, A to C), whose density was correlated positively with population density. This correlation was mainly due to the bias toward higher density of links in more densely populated areas because it disappeared when A-vulnerability was averaged over the links in each geographical area to control for this bias. However, substantial geographical heterogeneity still remained for the averaged A-vulnerability, ranging over several orders of magnitude when calculated for individual U.S. counties. These observations were validated with the U.S. county population data from the 2010 census and the geographic coordinates of county boundaries (fig. S6). Among the 48 states and the District of Columbia represented in the U.S. portion of the network, the three least vulnerable ones were West Virginia (average $\langle p_\ell \rangle$ of 3.2×10^{-5}), Tennessee (average $\langle p_\ell \rangle$ of 3.5×10^{-5}), and Mississippi (average $\langle p_\ell \rangle$ of 3.8×10^{-5}), all in the middle third of the population density ranking. However, some states among the least vulnerable did have relatively high or low population density, such as Illinois and Nebraska, which ranked 13th and 43rd in population density while having the 5th and 6th lowest A-vulnerability, respectively. The heterogeneity of A-vulnerability is visualized in Fig. 3A, with a map representation that equalizes the density of nodes. The breakdown of this representation into primary and secondary failures, presented in Fig. 3, B and C, shows that A-vulnerability to primary failures was more heterogeneously distributed than A-vulnerability to secondary failures. Over all pixels with nonzero A-vulnerability, the standard deviation of $\log \overline{\langle p_\ell \rangle}$ was 0.48 (89.2%), of $\log \overline{\left\langle p_\ell^{(\mathrm{p})} \right\rangle}$ was 0.58 (57.5%), and of $\log \overline{\left\langle p_\ell^{(\mathrm{s})} \right\rangle}$ was 0.41 (87.0%), where the number in parentheses represents the fraction of such pixels. The homogeneity in the distribution of secondary failures, which were several times more numerous than primary failures, underlies the relatively homogeneous aggregated distribution of the resulting power outages observed in Fig. 3A.

Network characterization of vulnerabilites

Our characterization of A-vulnerability allows us to study how the observed cascade dynamics depend on the network structure and to identify the topological centrality of individual links as a determinant. Topological centrality can be quantified through the concept of k-core (*26*–*29*), which is defined as the largest subnetwork in which every node has at least k links (that is, it has degree k). The k-core of a given network can be obtained by recursively removing all nodes with degree $<k$ until all nodes in the remaining network have degree $\geq k$. Repeating this for k = 1, 2, ... determines the k-core decomposition of the network. The coreness of a node is then defined as the (unique) integer c for which this node belongs to the c-core but not to the $(c+1)$–core (*30*). We further extend this concept to links by defining a link's coreness to be the smaller coreness of the two nodes it connects. A network visualization based on this decomposition is illustrated in Fig. 4A.

When this network decomposition was applied to the entire topology of the U.S.–South Canada power system, we found that links of coreness 2 were dominant in all three interconnections (with 81, 67, and 82% of all links in the Texas, Western, and Eastern networks, respectively). This dominance of coreness 2 links was also observed for the cascade-prone portion of the network and was further verified separately for the set of links vulnerable to primary failures ($\langle p_\ell^{(\mathrm{p})} \rangle > 0$) as well as the set of links vulnerable to secondary failures ($\langle p_\ell^{(\mathrm{s})} \rangle > 0$). These results are visualized in Fig. 4B for the case of the Eastern interconnection.

Upon closer inspection, however, the vulnerability revealed a strong correlation with link coreness beyond what can be inferred from the availability of links of a given coreness in the network. For primary failures, almost all links of coreness 1 showed zero A-vulnerability in our simulations, whereas 7 to 19% of higher coreness links were vulnerable (Fig. 4C). The links of coreness 1 are rarely vulnerable because each belongs to a tree subnetwork connected to the rest of the network through a single node, and this protects the link from flow rerouting, which is responsible for most primary failures (for example, flow rerouting accounts for more than 98%

Table 1. Subdivisions of the U.S.–South Canada power grid and its vulnerable sets. The rows represent the regions defined by NERC (Fig. 1A and table S2), within which the simulated cascades are triggered. The columns represent the number of buses, number of transmission lines, and four measures of the vulnerable sets: the number of vulnerable lines $|\mathcal{V}|$, the relative number of lines that are vulnerable in multiple snapshots $|\mathcal{V}_\cap|$, and the mean pairwise normalized topological and geographical distances between vulnerable lines, $\overline{d}_{\mathrm{v\text{-}v}}$ and $\overline{g}_{\mathrm{v\text{-}v}}$, respectively. These quantities are averaged over all snapshots (which is indicated by the notation $\langle\cdot\rangle$). The normalized distances are defined in the supplementary materials, materials and methods.

	Interconnections		Vulnerable sets			
	⟨Buses⟩	⟨Lines⟩	⟨$\lvert\mathcal{V}\rvert$⟩	$\lvert\mathcal{V}_\cap\rvert$	⟨$d_{\mathrm{v\text{-}v}}$⟩	⟨$g_{\mathrm{v\text{-}v}}$⟩
Texas	6161	7637	48	2.9	0.82	0.70
Western	15,891	20,397	81	5.9	0.84	0.95
Eastern	56,740	72,903				
FRCC			37	1.1	0.69	0.70
MRO			32	3.4	0.79	0.97
NPCC			130	2.1	0.85	0.72
RFC			76	4.5	0.94	0.91
SERC			11	11.6	0.92	0.94
SPP			14	3.3	0.66	0.63

of primary failures in the 2010 spring peak snapshot of the Texas network, as shown in the supplementary materials, materials and methods). Among the links that were vulnerable, the level of A-vulnerability increased monotonically with their coreness (Fig. 4D). This is probably because there are more flow paths (from power generators to consumers) that are parallel to a link of higher coreness in general, making the link more likely to be affected by flow rerouted from a failure in these paths.

For secondary failures, the fraction of links that were vulnerable and the A-vulnerability levels of these links followed opposite trends. The decrease in the fraction of vulnerable links shown in Fig. 4E can be understood by noting that a link can experience a secondary failure only if all available flow paths passing through that link are disabled by primary failures. Because links of higher coreness generally have more such paths, they were less likely to fail through this mechanism. Among the vulnerable links, the increase of the average A-vulnerability with coreness shown in Fig. 4F likely arose from the organization of the nodes in each k-core into graph components (maximal subsets of nodes in which every node pair is connected by a network path). Whereas the 2-core formed a single graph component in all three interconnections, the nodes in the 3-core were organized into multiple graph components (3, 11, and 52 components for the Texas, the Western, and the Eastern network, respectively), which were connected sparsely with each other by coreness 2 links. Because of this structure, most secondary failures on links of coreness ≥3 were likely caused by primary failures on the surrounding links of coreness 2 that disconnected a 3-core graph component with no internal power generation from the other 3-core components. This would make the links in these components prone to repetitively undergo secondary failures together. This tendency of co-occurring failures (*31*) among vulnerable links would lead to higher A-vulnerability for those links than for links with lower coreness.

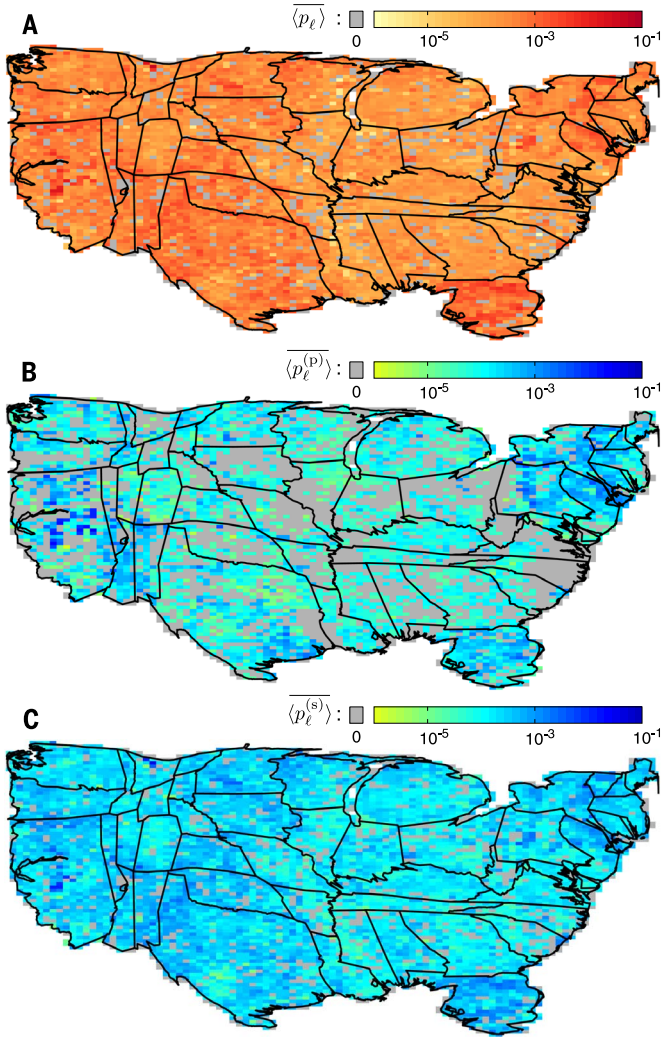


Fig. 3. Vulnerability of the power grid on a density-equalizing map. (**A**) Each pixel is color-coded by the average A-vulnerability $\overline{\langle p_\ell \rangle}$, including both primary and secondary failures, over all links connected to nodes in the area of the pixel. The cartogram was generated by using the diffusion-based method in (*41*) to equalize the density of nodes (supplementary materials, materials and methods) and is limited to the U.S. portion of the network. Color gray marks the pixels with zero average A-vulnerability. (**B** and **C**) Same as in (A) but color-coded separately for (B) the average A-vulnerability to primary failures $\overline{\langle p_\ell^{(p)} \rangle}$ and (C) the average A-vulnerability to secondary failures $\overline{\langle p_\ell^{(s)} \rangle}$.

Relating triggers and network states to vulnerable lines

To characterize the lines at risk of primary failures, we now shift our attention back to individual transmission lines connecting buses in each snapshot, rather than their collective representation as links. For this purpose, we define a "vulnerable" transmission line for a given snapshot to be a line ℓ for which $p_\ell^{(p)} > 0.0005$, with at least 95% Wilson's confidence level (*32*) (which excludes any line with a single failure in 1000 simulated events). This approach for vulnerability analysis is in contrast to previous studies on identifying the line failure combinations that initiate large cascading failures (*25*, *33*). We then define the vulnerable set $\mathcal{V}$ to be the set of all vulnerable lines for the given snapshot. We found that these vulnerable sets not only represented small portions of the grid in each snapshot but also exhibited considerable overlap across different snapshots (although it was rare for the same line to be vulnerable in all snapshots). These findings are presented in Table 1 for each interconnection using, respectively, the weighted average $\langle |\mathcal{V}| \rangle$ of the number of vulnerable transmission lines over all snapshots and the number $|\mathcal{V}_\cap|$ of lines that were vulnerable in two or more snapshots (relative to the number expected if the vulnerable sets were randomly distributed with no correlation). For example, in the Texas interconnection, $\langle |\mathcal{V}| \rangle = 48$ represents only about 0.6% of all the transmission lines, and the relative number of overlapping lines $|\mathcal{V}_\cap|$ is 2.9 (details on the distribution of $p_\ell^{(p)}$ for individual snapshots can be found in fig. S7).

Having a small portion of the grid vulnerable to cascading failures does not imply that these failures stayed localized even for single snapshots. To quantify the degree to which cascades were localized, we used the concepts of topological distance (the number of links along the shortest paths in the network) and geographical distance (the arc length along the Earth's surface), both normalized by the size of the triggering region measured by the respective distances and thus are unitless (supplementary materials, materials and methods). Specifically, the extent of the vulnerable set was measured by $d_{\text{v-v}}$ and $g_{\text{v-v}}$, defined as the normalized topological and geographical distance, respectively, between two transmission lines, averaged over all pairs of lines in the vulnerable set. We further defined $\langle d_{\text{v-v}} \rangle$ and $\langle g_{\text{v-v}} \rangle$ to be the weighted average of $d_{\text{v-v}}$ and $g_{\text{v-v}}$, respectively, over all snapshots. As shown in Table 1, for the Texas and Western networks, both $\langle d_{\text{v-v}} \rangle$ and $\langle g_{\text{v-v}} \rangle$ are comparable with the size of the interconnection, revealing that the spreading of cascades is nonlocal [which is consistent with observations from historical data (*34*), power flow calculations (*35*), and abstract models (*36*, *37*)]. In all cases, $\langle d_{\text{v-v}} \rangle < 1$ and $\langle g_{\text{v-v}} \rangle < 1$ hold true in the Eastern interconnection, where cascades were actually triggered in a local region and could have, in principle, spread widely to the other regions within the interconnection, leading to $\langle d_{\text{v-v}} \rangle > 1$ or $\langle g_{\text{v-v}} \rangle > 1$. This suggests that there is also an aspect of the cascading failures that is local: The propagation of failures in general does not extend too far from the region being perturbed.

The analysis of vulnerable sets provides relevant insights not only into the origins of cascading failures but also into the size of the damage inflicted on the network by individual cascades. In particular, what is the difference between the perturbations that cause large cascades and those that do not? To answer this question quantitatively, we categorized cascades according to their sizes measured by the power shed P_s defined above: small cascades (0.01 MW ≤ P_s < 300 MW) and large cascades (P_s ≥ 300 MW). This choice of

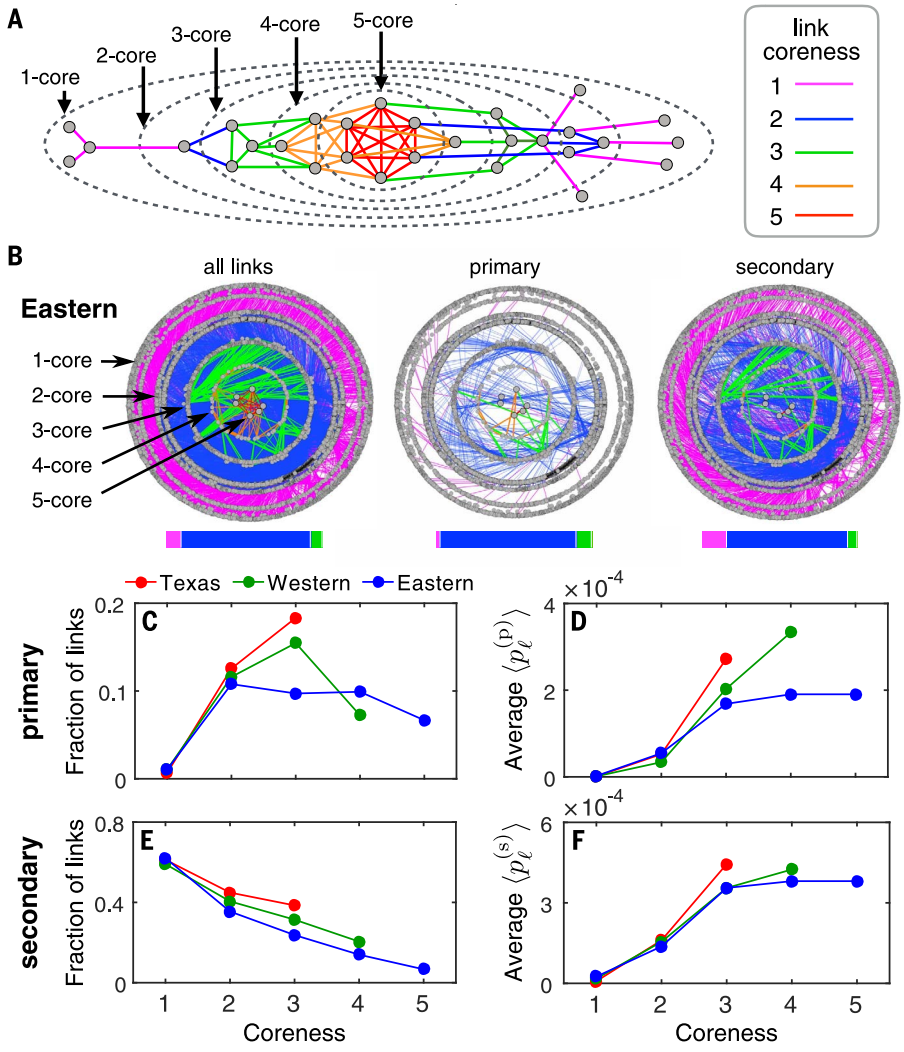


Fig. 4. Characterization of vulnerability through *k*-core decomposition. (**A**) Coreness-based network visualization, where nodes with higher coreness are placed closer to the center. (**B**) Visualization of the *k*-core decomposition of the Eastern interconnection, showing (left) all the links in the network, (middle) only the links with nonzero A-vulnerability to primary failures ($\langle p_\ell^{(p)} \rangle > 0$), and (right) only the links with nonzero A-vulnerability to secondary failures ($\langle p_\ell^{(s)} \rangle > 0$). The bars underneath show the distribution of link coreness, color-coded as in (A). (**C**) Fraction of links with $\langle p_\ell^{(p)} \rangle > 0$ among all links of a given coreness. (**D**) Average of $\langle p_\ell^{(p)} \rangle$ over all links of a given coreness with $\langle p_\ell^{(p)} \rangle > 0$. (**E** and **F**) Counterparts of (C) and (D), respectively, for secondary failures.

measure and threshold is based on the NERC requirement that all blackouts causing more than 300 MW of lost power be reported. We characterized perturbations by three different measures based on (normalized) distances: $d_{t\text{-}t}$, defined as the average pairwise distance among the n_t triggering line failures, as well as $d_{t\text{-}v}$ and $g_{t\text{-}v}$, defined as the minimum topological and geographical distances, respectively, from one triggering line failure to the vulnerable set $\mathcal{V}$. The average of these distances over cascades in each size category ($\overline{d}_{t\text{-}t}$, $\overline{d}_{t\text{-}v}$, and $\overline{g}_{t\text{-}v}$) is shown in Fig. 5 for each region. Cascades resulting in power shed $P_s \geq 300$ MW were associated with a set of triggering line failures that were topologically closer to each other (Fig. 5A), as well as with triggering failures that occurred topologically and geographically closer to a vulnerable line (Fig. 5, B and C).

Conclusions

Our vulnerability analysis of a continent-wide power system distinguishes itself from most previous studies by its scale but also by accounting for (i) the physics of cascading failures (dc-approximated power flow redistribution and heating of line conductors); (ii) grid operation practices (generation reserves and power shedding); and (iii) a wide range of conditions across years, seasons, and power demand levels (over which the average cascade size varies by one to two orders of magnitude). A strength of our approach is that it consists of tools—the definition of vulnerable sets, the method for aggregating multiple network conditions, and the analysis of coreness-vulnerability correlations—that are applicable to any cascade-prone network.

Our analysis separates the set of all failures occurring in cascade events into primary failures, which define the vulnerable set and account for only 1/5 of all failures, and secondary failures, which are more uniformly distributed and, albeit more numerous, are a mere consequence of the primary ones. The vulnerable set is not only surprisingly small but also highly skewed—with few lines far more likely to undergo a primary failure than the others—and patchy even when we control for the heterogeneity in the geographic organization of the grid. Although the vulnerable set is widespread through the network, the portion of it recruited in each cascade is not and is in fact strongly spatially correlated with the location of the triggering line failures; this is counter to the perception that cascades [for being nonlocal with respect to both topological and geographical distances (*31*, *38*)] can spread essentially without spatial constraints.

Our analysis also shows that larger cascades are associated with co-occurring perturbations that are closer both to each other and to the vulnerable set. This validates the existing hypothesis that localized triggering failures amount to bigger cascades (*39*) and reveals a striking relation to the classic threshold model (*4*) used to describe behavioral cascades in social systems, in which large cascades tend to be triggered by perturbations adjacent to the set of "early adopters." This set corresponds to the nodes most susceptible to change and thus plays a role similar to the one the vulnerable set plays in our analysis. The network topology emerged as a significant factor in determining the risk of cascading failures in our analysis based on the *k*-core decomposition, which has also been used to characterize nodes that serve as efficient spreaders in contact-based processes (*40*).

There are never two identical cascades in a network. It may thus come as a surprise that (primary) failures in large cascades are constrained to only a small subset of the network, which will likely experience new failures in the absence of remediating actions. This offers a scientific foundation for failure-based allocation of resources, which in the case of a power grid would be based on prioritizing upgrades of the system on the basis of previous observed failures (*14*)—but only if those are the primary (as opposed to all) failures (although upgrading transmission line capacities in the vulnerable set could create new vulnerable lines outside the set). Future work will be needed to determine the extent to which this applies to other flow networks that are subject to repeated failures, such as supply chains, food webs, and traffic networks.

Methods summary

For each interconnection, the system was modeled as a network of buses connected by transmission lines, given the parameters of individual network components in a given snapshot. The triggering

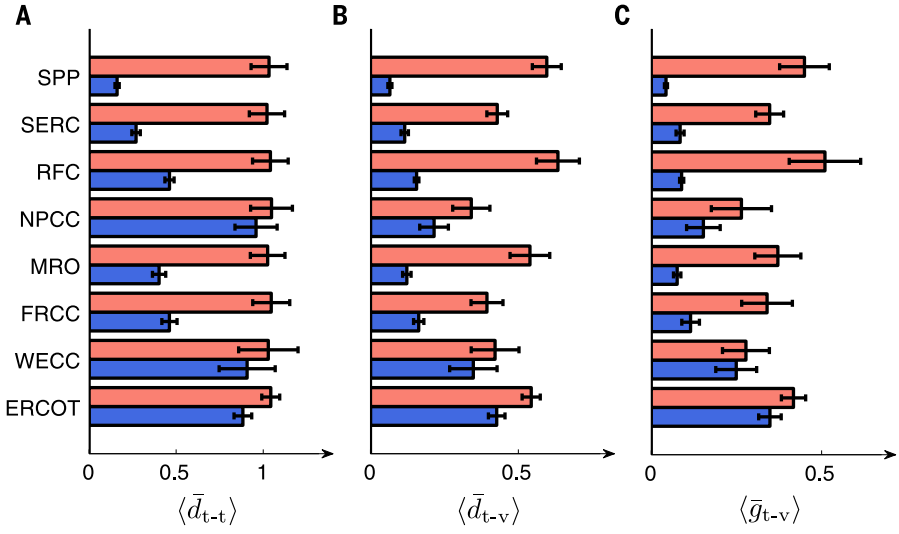


Fig. 5. Cascade size and distances involving triggering line failures. (**A** to **C**) Three types of (normalized) distances are shown for each NERC region (Fig. 1A and table S2): (A) the mean pairwise topological distance between the triggering failures, (B) the topological distance between the set of triggering failures and the vulnerable set, and (C) the geographical distance between the set of triggering failures and the vulnerable set. The distances are averaged separately over large cascades (blue, $P_s \geq 300$ MW) and over small cascades (red, 0.01 MW $\leq P_s <$ 300 MW). In each case, the distances are further averaged over all snapshots. Error bars mark the estimated standard deviation.

perturbations were chosen uniformly from all lines for the Texas and Western networks, whereas for the Eastern network, they were chosen uniformly within one of the six regions defined by NERC (Fig. 1A and table S2). The initial state of the network and the redistribution of power flow after a line removal were both calculated by solving an equation that expresses a balance between incoming and outgoing power flows at each bus. Through a temperature-evolution equation, the heating of a transmission line was modeled as an exponential convergence to the equilibrium temperature determined by the power flow over that line. Mechanisms responsible for the primary failures occurring in a given simulated cascade were identified by using an algorithm we developed to determine the degree to which the change in each generator's output contributes to changes in individual line power flows.

The density-equalizing transformation used to generate Fig. 3 was determined by estimating the density function for the geographical distribution of nodes and evolving it to a uniform-density equilibrium through a linear diffusion process (*41*). The topological and geographical distances between two transmission lines are defined based on the corresponding distances between the buses they connect. Both distances are thus zero between two lines that connect to a common bus. Further details on the formulation of the power flow equation, triggering perturbations, temperature evolution equation, validation of the cascade dynamics model against historical data, calculation of the density-equalizing transformation, algorithm for assigning power flow changes to generators, and the definitions of bus-to-bus distances are all given in the supplementary materials.

REFERENCES AND NOTES

1. The U.S. Department of Energy; www.energy.gov.
2. K. H. LaCommare, J. H. Eto, Cost of power interruptions to electricity consumers in the United States (US). *Energy* **31**, 1845–1855 (2006). doi: 10.1016/j.energy.2006.02.008
3. P. Hines, J. Apt, S. Talukdar, Large blackouts in North America: Historical trends and policy implications. *Energy Policy* **37**, 5249–5259 (2009). doi: 10.1016/j.enpol.2009.07.049
4. D. J. Watts, A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5766–5771 (2002). doi: 10.1073/pnas.082090499; pmid: 16578874
5. K. I. Goh, D.-S. Lee, B. Kahng, D. Kim, Sandpile on scale-free networks. *Phys. Rev. Lett.* **91**, 148701 (2003). doi: 10.1103/PhysRevLett.91.148701; pmid: 14611564
6. P. Crucitti, V. Latora, M. Marchiori, Model for cascading failures in complex networks. *Phys. Rev. E* **69**, 045104 (2004). doi: 10.1103/PhysRevE.69.045104; pmid: 15169056
7. A. E. Motter, Cascade control and defense in complex networks. *Phys. Rev. Lett.* **93**, 098701 (2004). doi: 10.1103/PhysRevLett.93.098701; pmid: 15447153
8. R. Kinney, P. Crucitti, R. Albert, V. Latora, Modeling cascading failures in the North American power grid. *Eur. Phys. J. B* **46**, 101–107 (2005). doi: 10.1140/epjb/e2005-00237-9
9. S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, S. Havlin, Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025–1028 (2010). doi: 10.1038/nature08932; pmid: 20393559
10. C. D. Brummitt, R. M. D'Souza, E. A. Leicht, Suppressing cascades of load in interdependent networks. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E680–E689 (2012). doi: 10.1073/pnas.1110586109; pmid: 22355144
11. D. P. Nedic, I. Dobson, D. S. Kirschen, B. A. Carreras, V. E. Lynch, Criticality in a cascading failure blackout model. *Int. J. Elec. Power* **28**, 627–633 (2006). doi: 10.1016/j.ijepes.2006.03.006
12. M. Anghel, K. A. Werley, A. E. Motter, Stochastic model for power grid dynamics. *Proceedings of the 40th Annual Hawaii International Conference on System Sciences* HICSS'07, Waikoloa, Big Island, HI, USA, vol. 1, p. 113 (2007).
13. I. Dobson, B. A. Carreras, V. E. Lynch, D. E. Newman, Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos* **17**, 026103 (2007). doi: 10.1063/1.2737822; pmid: 17614690
14. A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, G. Zussman, Sensitivity analysis of the power grid vulnerability to large-scale cascading failures. *Perf. Eval. Rev. Si.* **40**, 33–37 (2012). doi: 10.1145/2425248.2425256
15. C. Moore, S. Mertens, *The Nature of Computation* (Oxford Univ. Press, 2011).
16. S. H. Strogatz, Exploring complex networks. *Nature* **410**, 268–276 (2001). doi: 10.1038/35065725; pmid: 11258382
17. A. Vespignani, Predicting the behavior of techno-social systems. *Science* **325**, 425–428 (2009). doi: 10.1126/science.1171990; pmid: 19628859
18. D. Helbing, Globally networked risks and how to respond. *Nature* **497**, 51–59 (2013). doi: 10.1038/nature12047; pmid: 23636396
19. S. Sahasrabudhe, A. E. Motter, Rescuing ecosystems from extinction cascades through compensatory perturbations. *Nat. Commun.* **2**, 170 (2011). doi: 10.1038/ncomms1163; pmid: 21266969
20. J. A. Estes *et al.*, Trophic downgrading of planet Earth. *Science* **333**, 301–306 (2011). doi: 10.1126/science.1205106; pmid: 21764740
21. A. D. Barnosky *et al.*, Approaching a state shift in Earth's biosphere. *Nature* **486**, 52–58 (2012). doi: 10.1038/nature11018; pmid: 22678279
22. P. Gai, S. Kapadia, Contagion in financial networks. *Proc. R. Soc. London Ser. A* **466**, 2401–2423 (2010). doi: 10.1098/rspa.2009.0410
23. A. G. Haldane, R. M. May, Systemic risk in banking ecosystems. *Nature* **469**, 351–355 (2011). doi: 10.1038/nature09659; pmid: 21248842
24. I. Dobson, B. A. Carreras, V. E. Lynch, D. E. Newman, An initial model for complex dynamics in electric power system blackouts. *Proceedings of the 34th Annual Hawaii International Conference on System Sciences* HICSS'01, Maui, HI, USA, vol. 2, p. 2017 (2001).
25. M. J. Eppstein, P. Hines, A "random chemistry" algorithm for identifying collections of multiple contingencies that initiate cascading failure. *IEEE Trans. Power Syst.* **27**, 1698–1705 (2012). doi: 10.1109/TPWRS.2012.2183624
26. S. B. Seidman, Network structure and minimum degree. *Soc. Networks* **5**, 269–287 (1983). doi: 10.1016/0378-8733(83)90028-X
27. B. Bollobás, The evolution of sparse graphs, in *Graph Theory and Combinatorics, Proceedings of the Cambridge Combinatorial Conference in honor of Paul Erdös* (Academic Press, 1984), pp. 35–57.
28. S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, *k*-Core organization of complex networks. *Phys. Rev. Lett.* **96**, 040601 (2006). doi: 10.1103/PhysRevLett.96.040601; pmid: 16486798
29. J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, A. Vespignani, Large scale networks fingerprinting and visualization using the *k*-core decomposition. *Adv. Neur. Inform. Process. Syst.* **18**, 41–50 (2006).
30. J. A. Bondy, U. S. R. Murty, *Graph Theory with Applications* (Macmillan, 1976).
31. Y. Yang, T. Nishikawa, A. E. Motter, Vulnerability and cosusceptibility determine the size of network cascades. *Phys. Rev. Lett.* **118**, 048301 (2017). doi: 10.1103/PhysRevLett.118.048301; pmid: 28186802
32. L. D. Brown, T. T. Cai, A. DasGupta, Interval estimation for a binomial proportion. *Stat. Sci.* **16**, 101–133 (2001). doi: 10.1214/ss/1009213286
33. C. Long, D. You, J. Hu, G. Wang, M. Dong, Quick and effective multiple contingency screening algorithm based on long-tailed distribution. *IET Gener. Transm. Distrib.* **10**, 257–262 (2016). doi: 10.1049/iet-gtd.2015.0885
34. I. Dobson, B. A. Carreras, D. E. Newman, J. M. Reynolds-Barredo, Obtaining statistics of cascading line outages spreading in an electric transmission network from standard utility data. *IEEE Trans. Power Syst.* **31**, 4831–4841 (2016). doi: 10.1109/TPWRS.2016.2523884
35. D. Jung, S. Kettemann, Long-range response in ac electricity grids. *Phys. Rev. E* **94**, 012307 (2016). doi: 10.1103/PhysRevE.94.012307; pmid: 27575148
36. L. Daqing, J. Yinan, K. Rui, S. Havlin, Spatial correlation analysis of cascading failures: Congestions and blackouts. *Sci. Rep.* **4**, 5381 (2014). doi: 10.1038/srep05381; pmid: 24946927
37. D. Witthaut, M. Timme, Nonlocal effects and countermeasures in cascading failures. *Phys. Rev. E* **92**, 032809 (2015). doi: 10.1103/PhysRevE.92.032809; pmid: 26465530

38. P. D. Hines, I. Dobson, P. Rezaei, Cascading power outages propagate locally in an influence graph that is not the actual grid topology. *IEEE Trans. Power Syst.* **32**, 958–967 (2017). doi: 10.1109/TPWRS.2016.2578259
39. Y. Berezin, A. Bashan, M. M. Danziger, D. Li, S. Havlin, Localized attacks on spatially embedded networks with dependencies. *Sci. Rep.* **5**, 8934 (2015). doi: 10.1038/srep08934; pmid: 25757572
40. M. Kitsak *et al*., Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893 (2010). doi: 10.1038/nphys1746
41. M. T. Gastner, M. E. J. Newman, Diffusion-based method for producing density-equalizing maps. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7499–7504 (2004). doi: 10.1073/pnas.0400280101; pmid: 15136719

ACKNOWLEDGMENTS

The authors thank H. Valizadehhaghi for insightful discussions. This work was supported by the Institute for Sustainability and Energy at Northwestern (ISEN) under a Booster Award, the U.S. National Science Foundation under grant DMS-1057128, and the Advanced Research Projects Agency–Energy (U.S. Department of Energy), under award DE-AR0000702. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof. The power-grid data were obtained from FERC under a nondisclosure agreement by following the procedure described at www.ferc.gov/legal/ceii-foia/ceii.asp. The BPA line outage data and the NERC grid disturbance data are both publicly available at https://transmission.bpa.gov/Business/Operations/Outages (Miscellaneous Outage Data and Analysis) and www.oe.netl.doe.gov/OE417_annual_summary.aspx (Electric Disturbance Events, OE-417), respectively. The 2010 U.S. census data and the boundary data for the U.S. counties can be downloaded from https://factfinder.census.gov and www.census.gov/geo/maps-data/data/cbf/cbf_counties.html, respectively.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/358/6365/eaan3184/suppl/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S3
References (*42*–*49*)

27 March 2017; accepted 21 September 2017
10.1126/science.aan3184

Small vulnerable sets determine large network cascades in power grids

Yang Yang, Takashi Nishikawa and Adilson E. Motter

Science **358** (6365), eaan3184.
DOI: 10.1126/science.aan3184

The domino effect in power failure

Sometimes a power failure can be fairly local, but other times, a seemingly identical initial failure can cascade to cause a massive and costly breakdown in the system. Yang *et al.* built a model for the North American power grid network based on samples of data covering the years 2008 to 2013 (see the Perspective by D'Souza). Although the observed cascades were widespread, a small fraction of all network components, particularly the ones that were most cohesive within the network, were vulnerable to cascading failures. Larger cascades were associated with concurrent triggering events that were geographically closer to each other and closer to the set of vulnerable components.

Science, this issue p. eaan3184; see also p. 860