UC Davis
UC Davis Previously Published Works

Title

Competitive percolation strategies for network recovery

Permalink

https://escholarship.org/uc/item/6p95r634

Journal

Scientific Reports, 9(1)

ISSN

2045-2322

Authors

Smith, Andrew M
Pósfai, Márton
Rohden, Martin
et al.

Publication Date

2019

DOI

10.1038/s41598-019-48036-0

Peer reviewed

SCIENTIFIC REPORTS

natureresearch

OPEN

Competitive percolation strategies for network recovery

Andrew M. Smith[1], Márton Pósfai[1], Martin Rohden[1], Andrés D. González[2], Leonardo Dueñas-Osorio[3] & Raissa M. D'Souza[1,4,5]

Restoring operation of critical infrastructure systems after catastrophic events is an important issue, inspiring work in multiple fields, including network science, civil engineering, and operations research. We consider the problem of finding the optimal order of repairing elements in power grids and similar infrastructure. Most existing methods either only consider system network structure, potentially ignoring important features, or incorporate component level details leading to complex optimization problems with limited scalability. We aim to narrow the gap between the two approaches. Analyzing realistic recovery strategies, we identify over- and undersupply penalties of commodities as primary contributions to reconstruction cost, and we demonstrate traditional network science methods, which maximize the largest connected component, are cost inefficient. We propose a novel competitive percolation recovery model accounting for node demand and supply, and network structure. Our model well approximates realistic recovery strategies, suppressing growth of the largest connected component through a process analogous to explosive percolation. Using synthetic power grids, we investigate the effect of network characteristics on recovery process efficiency. We learn that high structural redundancy enables reduced total cost and faster recovery, however, requires more information at each recovery step. We also confirm that decentralized supply in networks generally benefits recovery efforts.

Resilience of complex networks is one of the most studied topics of network science, with an expanding literature on spreading of failures, mitigation of damage, and recovery processes[1–6]. The level of functionality of a network is typically quantified by its connectedness, e.g., size of the largest component[1], average path length[7,8], or various centrality metrics[9]. Such simple topology-based metrics ensure mathematical tractability and allow us to analyze and compare networks that can be very different in nature, providing general insights into the organization of complex systems. However, such a perspective necessarily ignores important system-specific details. For example, abstracted topological models of infrastructure networks recovering from damage or catastrophic failure aim to rapidly restore the largest component[10–13]. But, extensive connectivity is not a necessary condition to guarantee that all supply and demand can be met. For instance, consumers of a power grid can be served if they are connected to at least one power source and that source satisfies operational constraints[14]. The concept of "islanding", a technique of intentionally partitioning the network to avoid cascading failures, is actually a practical strategy used to improve security and resilience during restoration efforts in power grids[15–17]. Indeed, after the 2010 earthquake in Chile the recovery process first created five islands, which were only connected to each other in the final steps of reconstruction[18].

The restoration of critical infrastructure operation after a catastrophic event, such as a hurricane or earthquake, is a problem of great practical importance and is the focus of a significant body of work in civil and industrial engineering disciplines. The goal of engineering based models of recovery is to provide system-specific predictions and actionable recommendations. This is achieved by incorporating component level details and realistic transmission dynamics into the models, often in the form of generalized formulations of network design problems (NDPs), which satisfy network flows. In this context, objective functions of such NDP-based models aim to minimize the construction and/or operational costs of recovering edges and nodes in a utility network. Basic forms of the NDP have been well-studied[19,20], and have recently been combined with scheduling and

[1]Department of Computer Science and Complexity Sciences Center, University of California, Davis, CA, 95616, USA. [2]School of Industrial and Systems Engineering, University of Oklahoma, Norman, OK, 73019, USA. [3]Department of Civil & Environmental Engineering, Rice University, Houston, TX, 77005, USA. [4]Department of Mechanical and Aerospace Engineering, University of California, Davis, California, 95616, USA. [5]Santa Fe Institute, Santa Fe, New Mexico, 87501, USA. Correspondence and requests for materials should be addressed to A.M.S. (email: amsmi@ucdavis.edu)

1

resource allocation problems to model the entire restoration process[21–23]. While these models provide a principled manner to obtain optimal, centralized recovery strategies, their complexity (at least NP-complete)[19] renders computation not scalable, and interpretation restricted in scope to small instances. More efficient approximate solutions for NDPs have been found using optimization meta-heuristics such as hybrid ant system[24] and gradient descent[25] methods. Such algorithms are generally applicable to global search problems and were designed to reduce computational complexity by not guaranteeing optimality; therefore, they provide limited insight into the mechanism of network formation during recovery. We will analyze the output of an NDP algorithm, and leveraging on these observations, we will develop a percolation-based model for network recovery with the goal of uncovering important principles of network formation and recoverability.

Percolation processes, often used for studying properties of stochastic network formation, have recently been applied to network recovery[26]. In the kinetic formulation of random percolation[27], we start with N unconnected nodes and consider a discrete time process. At each timestep, an edge is selected from the set of all possible edges at random, and added to the network. Initially the largest connected component (LCC) is sublinear in N; above a critical edge density it spans a finite fraction of the network and the LCC is referred to as the giant component. Controlling location of the critical point is of great interest in many systems–for instance, suppressing the formation of the giant component may reduce the likelihood of virus spreading in social contact networks. This can be achieved by selecting $M > 1$ candidate edges at each timestep, and adding the edge that optimizes some criteria. The general class of models that results from this choice is referred to as competitive percolation or an *Achlioptas process*[28,29]. While simple, Achlioptas processes often have the benefit of being scalable, numerically analyzable, and provide a parameter, M, for tuning how close the formation process is to matching the desired criteria. Note that when M is equal to the number of possible edges, we always add the edge that is optimal with respect to the selection criteria. Previous percolation-based recovery models typically measure solution quality by how quickly the LCC grows[11,12], or assume nodes which are not connected to the LCC to be nonfunctional[10,13]. However, empirical studies of recovery scenarios suggest that these assumptions do not apply to infrastructure networks after catastrophic scenarios[18].

In this work, we aim to narrow the gap between topology-based recovery approaches and computationally difficult optimization approaches by incorporating features which mirror infrastructure restoration processes. We start by applying a generalized version of a well-studied NDP recovery algorithm[22,23,30,31] to a small case study, and we identify that the satisfaction of demand is a key driving force in the initial periods of recovery, outranking operational efficiency and direct repair costs of network elements in importance. Motivated by this finding, we define a simple, competitive percolation-based model of recovery that aims to maximize the satisfaction of consumer demand in a greedy manner. We show that component size anti-correlates with the likelihood of further growth– leading to islanding and the suppression of the emergence of large-scale connectivity, similar to explosive percolation transitions[29] and in contrast with traditional topological recovery models. We apply our recovery algorithm to synthetic power grids to systematically investigate how realistic structural features of the network affect the efficiency of the recovery process. We learn that high structural redundancy (related to the existence of multiple paths between nodes) allows for reduced total cost and faster recovery time; however, to benefit from that redundancy, an increasing amount of information needs to be considered at each step of reconstruction. We also study the role of the ratio of suppliers and consumers and find that decentralized supply generally benefits recovery efforts, unless the fraction of suppliers becomes unrealistically high. Our model deepens our understanding of network formation during recovery and of the relationship between network structure and recoverability. We anticipate that our work can lead to efficient approximations of the NDP algorithm by leveraging the important mechanisms uncovered by our competitive percolation model.

Model

Problem statement and the optimal recovery model.

We are interested in the problem of restoring the operation of a critical infrastructure system after sustaining large-scale damage. The infrastructure network is represented by a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the set of N nodes corresponding to substations and $\mathcal{E}$ is a set of E edges corresponding to transmission lines, e.g., power lines, water or gas pipes. We introduce the parameter d_i representing the commodity demand of node i: if $d_i < 0$, node i is a net consumer; if $d_i > 0$, node i is a net supplier. We normalize d_i such that the total consumption (or production) sums to unity, i.e., $\frac{1}{2}\sum_{i=1}^{N} |d_i| = 1$. Following a catastrophic event, a subset of the network $\mathcal{G}' = (\mathcal{N}', \mathcal{E}')$ becomes damaged. We study a discrete time reconstruction process: in each timestep we fix one damaged component, and the process ends once the entire network is functional. Our goal is to identify a sequence in which to repair the elements such that the total cost of recovery is minimized. In this manuscript, we focus on the fundamental case where all links are damaged but nodes remain functional, i.e., $\mathcal{E}' = \mathcal{E}$ and $\mathcal{N}' = \varnothing$.

Optimization frameworks are often used in order to explore the space of possible repair sequences and identify best solutions. We implement an algorithm for the time-dependent network design problem (td-NDP)[22,23,30], which is a well known example of an optimization algorithm for network recovery developed by the civil engineering community. Out of the recovery processes we examine in this paper, td-NDP is the most realistic, and therefore the most computationally complex. It is formulated as a mixed integer problem which optimizes a cost that includes reconstruction costs of network components, operational costs, and penalties incurred for unsatisfied demand, while taking constraints on flows of commodities into account. In general, mixed integer programs are known to be NP-hard except in special cases. For the td-NDP this means that problems become exponentially harder as the size of the network to be reconstructed increases; therefore, it is common practice to break up the recovery into time windows of length T, and find the locally optimal solution in each window. Previous studies showed that often even $T = 1$ yields adequate approximation of the globally optimal solution[32]. Here, depending on the size of the network, we use up to $T = 5$, striking a compromise between computational efficiency and
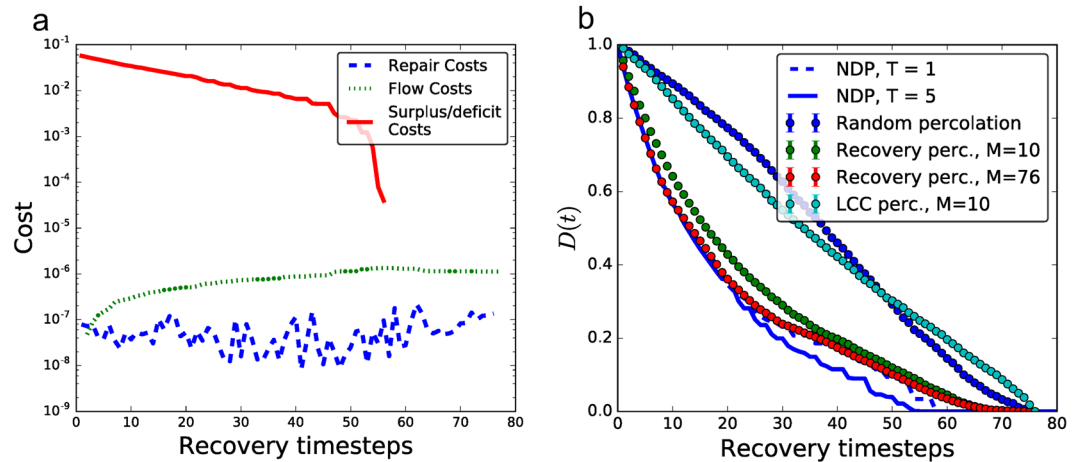
Figure 1. Recovery of the Shelby County power network. (**a**) Breakdown of normalized recovery costs for the locally optimal td-NDP solution. Surplus/deficit costs are the overwhelmingly dominant factor in early stages of recovery. Note that the surplus/deficit cost (red line) becomes zero once all nodes belong to a single connected component and all demand is met. The total recovery time is equal to the number of links $E = 76$ in the network. (**b**) Total commodity deficit during the recovery of the Shelby County power network, averaged over 100 independent realizations for percolation results; the recovery percolation model closely approximates the td-NDP baseline surplus/deficit cost curve (especially as M increases), while LCC percolation deviates.

establishing a hard upper bound to the true but unknown minimum cost configuration. A formal definition of td-NDP is provided in the Methods section.

To uncover the key driving factors and properties of the recovery process, we apply the td-NDP to a representative example, the transmission power grid of Shelby County, Tennessee, which consists of 9 suppliers and 37 consumers (with 14 junction nodes where $d_i = 0$), connected by $E = 76$ transmission lines. Network topology and necessary parameters were obtained from refs[22,23]. Figure 1a shows the total repair cost as a function of time as we perform the td-NDP with $T = 5$ on a network that was initially completely destroyed, with cost broken down by the type of expense. We see that deficit cost (i.e., the penalty accrued for unsatisfied demand) is dominant and exponentially decreasing throughout the initial stages of recovery. Investigating how this impacts the growth of the components, Fig. 2a shows the commodity deficit or supply of each connected component throughout the recovery process, with circle sizes representing the component size, and colors representing if a component is over- (blue) or undersupplied (red). We see that the td-NDP process results in many small components initially, with relatively small surplus/deficit, and only towards the end of the process all components are joined. This is consistent with islanding techniques discussed in engineering practice. Our goal is to develop a simple and computationally efficient model of the recovery process that captures these key features.

Competitive percolation optimizing for LCC growth.   Previous topology-based recovery processes prioritize the rapid growth of the largest connected component (LCC)[10–13]. Models vary in the details, such as the type of failure (random, localized, catastrophic) and additional secondary objectives (such as prioritizing nodes based on population), but the metric for the quality of the solution is directly related to how quickly the LCC grows.

As a representative example of topology driven recovery strategies, we implement an Achiloptas process using a selection rule that maximizes the sum of the resulting component, which we refer to as *LCC percolation*. In this process, we randomly select $M > 0$ candidate edges out of the set of damaged edges at each discrete timestep. We then examine the impact that repairing each individual edge would have and select the edge that, when added, creates the largest connected component. More specifically, let s_i denote the size of the component to which node i belongs. If nodes i and j belong to separate components, repairing edge (i, j) creates a component with size $S_{ij} = s_i + s_j$; if they belong to the same component, the size of the component does not change and we set $S_{ij} = 0$. Out of the M candidate edges, we repair the one that maximizes S_{ij}; if multiple candidate links have the same maximal S_{ij}, we select one of them uniformly at random. If $M = 1$, the process is equivalent to traditional percolation. If $M = E$, the process is largely deterministic, we always repair an edge that is optimal with respect to the selection criteria.

We now apply LCC percolation as a model for the recovery of the Shelby County power grid and compare the results to the benchmark td-NDP process. Figure 2b shows that if we use growth of the LCC as our objective, the LCC, represented by the largest circle, grows rapidly throughout the process as expected. However, the deficit/surplus of this component fluctuates greatly. As indicated in Fig. 1b by the magnitude of total commodity deficit D, i.e., the total unsatisfied demand in the network, such a recovery algorithm is costly and leaves large portions of the grid without power until the final steps of the recovery algorithm. To conclude, we find that a recovery process based on quick growth of the LCC is neither cost efficient nor effective for satisfying consumer's demand as quickly as possible. As a result, we do not consider this algorithm further.
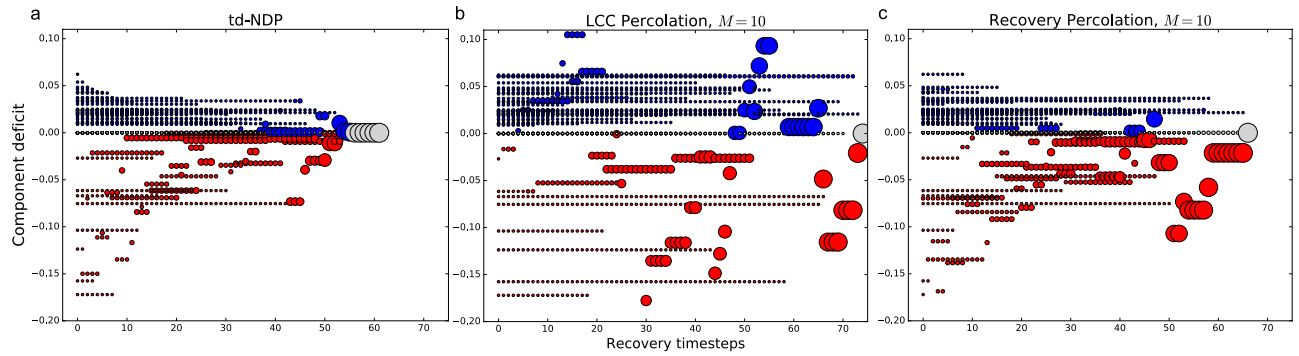
Figure 2. Commodity deficit of each connected component over time throughout various recovery processes on the Shelby County dataset: (**a**) td-NDP, (**b**) LCC percolation, and (**c**) recovery percolation. Each point represents a component, where the size of the point indicates the relative size of the component. Blue circles (above 0 on the vertical axis) indicate excess supply, while red circles (below 0 on the vertical axis) indicate unmet demand. Grey circles indicate that supply and demand are perfectly matched (at 0 on the vertical axis). The final circle corresponds to the first timestep where all nodes belong to the LCC; the recovery process, however, ends when all $E = 76$ links are repaired, which may occur after all nodes are connected. The optimization process shown in (**a**) keeps size and commodity deficit low throughout the recovery process, consistent with islanding practices. LCC percolation in (**b**) is vastly different, showing varying sizes and commodity deficits throughout the process, resulting in higher recovery costs. Our competitive recovery percolation method in (**c**) shows a signature closer to (**a**) and is much more computationally efficient, choosing the next edge to repair from a sample size of only $M = 10$ at each timestep.

Competitive percolation optimizing for demand satisfaction.

We have shown using an example power grid that the key driving factor in the recovery process is the reduction of the total commodity deficit, i.e., the total unsatisfied demand, and that optimizing LCC fails to capture this. We anticipate that this also holds for the recovery of other critical infrastructure networks, such as gas and water supply networks. To capture the essence of real recovery strategies, we propose a competitive percolation process which we refer to as *recovery percolation* that, instead of optimizing for LCC growth, aims to directly reduce the unsatisfied demand. In addition to network topology, this recovery process also takes into account the net demand or production of the individual nodes.

We define D_i as the commodity deficit of the connected component to which node i belongs. We assume that capacity constraints of the transmission lines are sufficient and thus do not limit the flow of commodities during the recovery process, a common practice in infrastructure recovery literature[14,22,23]. Therefore the commodity deficit of a component is the sum of demand or supply of individual nodes belonging to the component, i.e., $D_i = \sum_{j \in \mathcal{C}_i} d_j$, where $\mathcal{C}_i$ is the set of nodes belonging to the component containing node i.

We use commodity deficit of the components as a selection criteria for the competitive percolation model to account for the goal of balancing supply and demand. Similar to LCC percolation, we randomly select $M > 0$ damaged candidate edges, from which one is chosen to be repaired and added to the network at each timestep as follows. We first consider how much demand would be met by adding each of the M edges individually to the network. More specifically, if nodes i and j belong to components such that $D_i D_j < 0$, then repairing edge (i, j) reduces the total commodity deficit by $\Delta D = \min(|D_i|, |D_j|)$; if $D_i D_j \geq 0$, then there is no commodity deficit reduction, i.e., $\Delta D = 0$. Out of the M candidate edges, we repair the one that maximizes ΔD; if multiple candidates have the same maximal ΔD, we repair one of them chosen uniformly at random. If $M = E$ the process always selects an edge that is optimal with respect to the selection criteria; if M is reduced, the process becomes more stochastic as it approaches 1. Overall, the computational complexity of this algorithm is $O(EM)$, since E links are repaired and to repair each link M candidates are considered. This polynomial runtime is in contrast with the exponential td-NDP algorithm.

Figure 1b shows that for the Shelby County power grid the total commodity deficit during the recovery percolation for $M = E = 76$ well approximates the td-NDP, especially at the beginning of the recovery process when costs are much higher. We also see that even when $M = 10$, corresponding to only ~13% of the total edges considered at each timestep, the approximation remains very effective. Figure 2c shows similar dynamical behavior in recovery percolation as in the td-NDP solution (cf. Fig. 2a): larger components delay formation, and tend to have smaller commodity deficit.

Results

In the following, we apply the recovery percolation model to various synthetic network topologies with realistic features to identify important mechanisms driving network formation and to understand how network structure affects the efficiency of recovery efforts. For each synthetically generated network, the demand distribution is chosen to approximate the demand observed in real power grids (details are provided in the Methods section).

Recovery percolation on complete networks.

We have shown that recovery percolation follows our benchmark td-NDP solution closely on a real-world topology. We also observed that the growth of connected components is suppressed via recovery percolation as compared to LCC percolation. To understand this behavior,
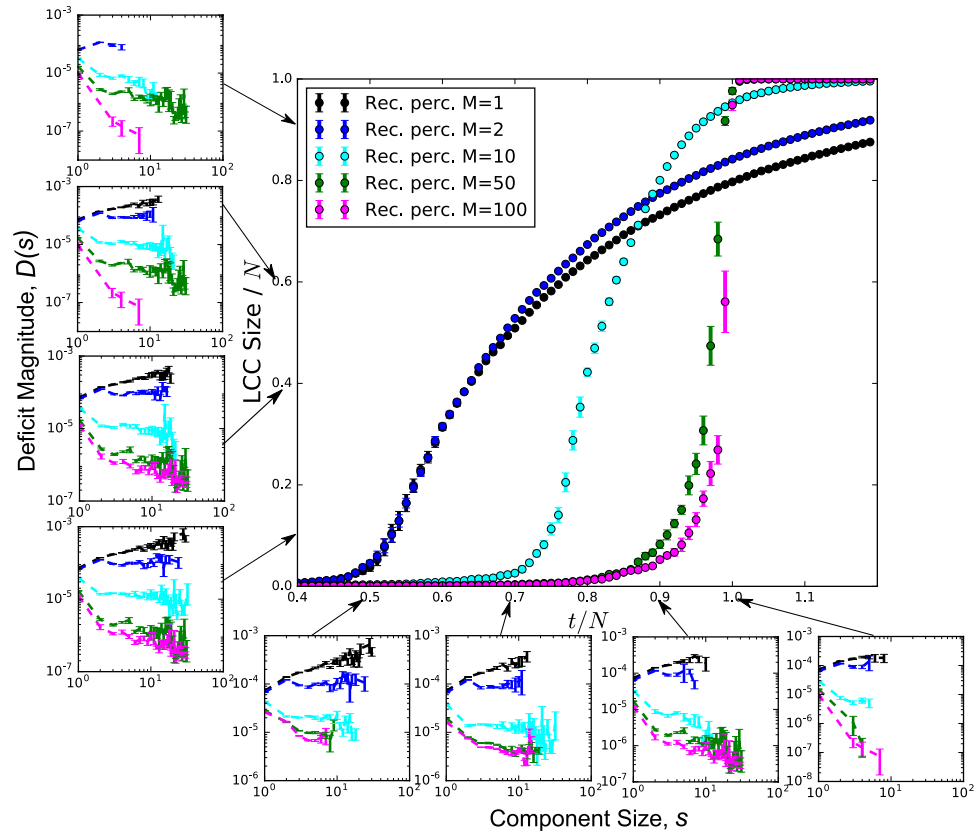
Figure 3. Recovery percolation on the complete graph, showing growth of the LCC with increasing number of edges repaired (center figure). We see that as M is increased, the transition point is delayed. Small plots show the average undersupply of connected components as a function of component size. Left column plots distributions for various M values when the LCC is of a certain size; bottom row plots shows the same for fixed times. Rare events (component sizes that occur <1% of the time) are not plotted to eliminate noise. We note that as M is increased, the slopes trend towards being more negative, indicating that larger components have less deficit magnitude throughout the recovery process when more recovery choices are presented to the recovery percolation process. All data points are the average of 10 independent realizations with $N = 10^4$ nodes;, error bars indicate the standard error of the mean.

we study large systems with $N = 10^4$ nodes and we allow potential edges to exist between any node pair, removing underlying topology constraints. Note that the td-NDP process is intractable for networks of this size.

Figure 3 (main figure) shows the growth of the LCC for a range of M values. For $M = 1$, the model reduces to random percolation which has a second-order phase transition at $t/N = 0.5$, and above this critical point the LCC becomes proportional to N. As we increase M, the apparent transition point shifts to higher values and approaches 1, indicating that the appearance of large-scale connectivity is suppressed; however, once the transition point is reached, the growth of the LCC becomes increasingly abrupt. This observation is analogous to explosive percolation, where links are chosen to be constructed explicitly to delay component growth[29]. In contrast, in recovery percolation it is an indirect consequence of a practical restoration strategy.

To understand the underlying mechanism of component growth, we plot the average component sizes and their corresponding average undersupply at various points during the reconstruction process in the bottom row of plots in Fig. 3. Note that average oversupply behaves in a similar manner, but is omitted for clarity. The left column of plots in Fig. 3 shows the same quantities such that the size of the LCC is fixed. The main trend we observe at any given point in time is that for large enough M there is negative correlation between component size and undersupply, and this correlation becomes stronger as M increases. This means that as components grow their commodity deficit is reduced and therefore the likelihood of further growth is also reduced, ultimately suppressing the appearance of large scale connectivity.

The observed two features also describe islanding, an intentional strategy in resilience planning and recovery in real-world power grids. This islanding behavior is already observed in early stages of the restoration process, becoming more apparent as t approaches the transition point.

Recovery percolation on synthetic power grids. So far we investigated the recovery process on an underlying graph without topological constraints. We also wish to analyze more realistic networks and turn our attention to synthetic power grids[33,34]. This allows us to systematically investigate how typical structural features of power grids affect the efficiency of the recovery process.
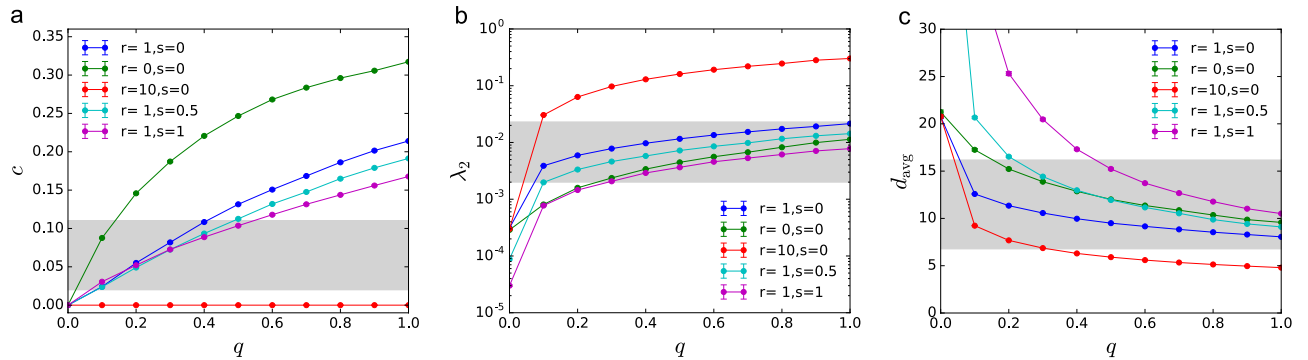
Figure 4. Network structure of the synthetic power grids. (**a**–**c**) Increasing q increases both local and global redundancy, i.e., clustering c and algebraic connectivity λ_2; while decreases the average shortest path length d_{avg} of the network. High r (solid red) favors long distance shortcuts, increasing λ_2 and decreasing c; while small r (green) has the opposite effect. Increased line splitting s (light blue and purple) increases d_{avg}, especially for low q; while reducing both c and λ_2. The shaded area indicates a realistic regime, corresponding to the range spanned by a set of real networks with more than 500 nodes provided in ref.[34]. Each data point is an average of 100 independent network realizations with $N = 1000$ nodes; the error bars indicate the standard error of the mean and are smaller than the markers.

Power grids are spatially embedded networks, and physical constraints, particularly at the high-voltage transmission level, limit the maximum number of connections a node can have. Their degree distributions, therefore, have an exponential tail, in contrast to many complex networks that display high levels of degree heterogeneity. Transmission power grids typically have average degree $\langle k \rangle$ between 2.5 and 3[34,35]. An important requirement of power grids is structural redundancy, meaning that the failure of a single link cannot cause the network to fall into disconnected components. A network without redundancy has tree structure, has average degree 2 and all node pairs are connected through a unique path. Any additional link creates loops and improves redundancy. Structural redundancy can be characterized locally by counting short range loops. For example, power grids have a high clustering coefficient c, typically ranging between 0.05 and 0.1[34]. The algebraic connectivity, denoted by λ_2, is the second smallest eigenvalue of the network's Laplacian matrix and captures a measure of global redundancy: it is related to the number of links that have to be removed in order to break the network into two similarly sized components, with high values corresponding to high redundancy. The exact value of λ_2 depends on system size, where for a given number of nodes, λ_2 is minimal for tree structure, and monotonically increases as further links are added[36].

To generate networks that exhibit the features of typical transmission power grids, we use a simplified version of a practical model developed by Schultz *et al.*[34]. The model generates spatially embedded networks mimicking the growth of real-world power grids. The process is initiated by randomly placing N_0 nodes on the unit square and connecting them with their minimum spanning tree. To increase redundancy, qN_0 ($0 \leq q \leq 1$) number of links are added one-by-one connecting a random node i to node j, such that it minimizes the redundancy-cost trade-off function

$$f(i, j) = \frac{(d_{net}(i, j) + 1)^r}{d_{euc}(i, j)},$$
(1)

where i and j are two nodes not connected directly, $d_{net}(i, j)$ is their shortest path distance in the network, and $d_{euc}(i, j)$ is their Euclidean distance (In the original model, with some probability an additional redundancy link is connected to the newly added node in each timestep. We found that the roles of these two types of redundancy links are almost identical in the context of recovery percolation; therefore we omitted the latter to simplify discussion. For further details see ref.[34]). The $r \geq 0$ parameter controls the trade-off between creating long loops to improve redundancy and the cost of building power lines. After the initialization, we add $N - N_0$ nodes through a growth process. In each time step a new node is added: with probability $1 - s$ the node is placed in a random position and connects to the nearest node; with probability s a randomly selected link (i, j) is split and a new node is placed halfway between nodes i and j and is connected to both of them. To increase redundancy, in each time step an additional link is added with probability q connecting a randomly selected node i to node j, such that $f(i, j)$ is minimized. Finally, a fraction of nodes p_s are randomly selected to be suppliers, the rest are assigned to be consumers.

Changing parameters q, r, and s allows us to systematically explore how these parameters impact the structure of these model power grids (Fig. 4): q controls the average degree $\langle k \rangle = 2(1 + q)$ and adds redundancy to the network; r controls how loops are formed, where small r favors short distance connections leading to high c and low λ_2, while large r favors long loops leading to low c and high λ_2; and s increases typical distances in the network, lowering both c and λ_2. Extreme choices of these parameters, however, may produce networks with unrealistic properties. On Fig. 4, we highlighted a realistic regime corresponding to the range spanned by a set of real networks with more than 500 nodes provided in ref.[34].

Although later we focus on larger networks, for completeness we note that using this network generation procedure we can create synthetic networks with approximately the same properties as the Shelby County dataset.
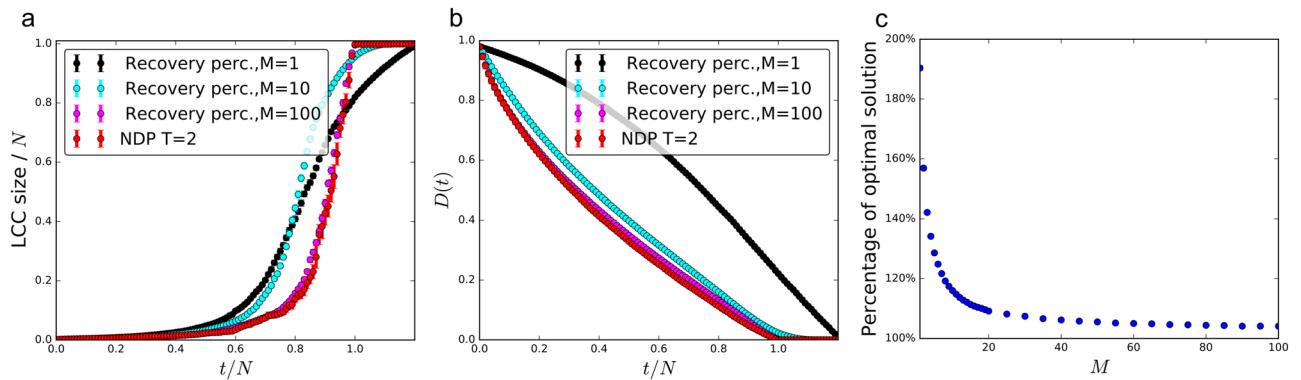
Figure 5. Recovery percolation and td-NDP on synthetic power grids. (**a**) Size of the LCC as a function of recovery time and (**b**) total commodity deficit during the restoration. High M values well capture the behavior of the td-NDP process. (**c**) How well the total cost during recovery percolation approximates the cost of the td-NDP solution as a function of M. All data points are the average of 10 independent realizations with $N = 1000$ nodes; error bars indicate the standard error of the mean and are typically smaller than the markers.

The Shelby County transmission grid has number of nodes $N = 60$, average degree $\langle k \rangle = 2.53$, clustering coefficient $c = 0.078$, algebraic connectivity $\lambda_2 = 0.073$, and average shortest path length $d_{avg} = 5.35$. Setting the parameters to, for example, $N = 60$, $N_0 = 8$, $q = 0.27$, $r = 1$, and $s = 0.4$, generates networks with $c = 0.078 \pm 0.038$, $\lambda_2 = 0.059 \pm 0.019$, and $d_{avg} = 5.13 \pm 0.46$, where the reported values are an average of 10^3 realizations and the error is the standard deviation.

Comparing recovery percolation and td-NDP. We first consider a set of parameters that yield typical topologies and compare the performance of recovery percolation to that of the locally optimal td-NDP recovery. We choose the parameters to create networks similar to the Western US grid following the specifications of ref.[34] ($N = 10^3$, $q = 0.33$, $r = 1$, and $s = 0$). For the td-NDP analysis, we reduce the time window from $T = 5$, as used in the Shelby County model, down to $T = 2$ for tractability reasons since our synthetic networks are much larger (increasing T causes an exponential increase in complexity). Note that lower values of T lead to more localized search, resulting in a suboptimal solution. However, $T = 2$ is still a useful touchstone since (i) it represents the best estimate with the available tools and (ii) even $T = 1$ provides a reasonable approximation of the optimal solution (see Fig. 1b and ref.[32]). Figure 5a shows the growth of the LCC for the td-NDP process and recovery percolation varying M from 1 to 100. For recovery percolation we find similar behavior to what we observed for complete networks: as M is increased the growth of the LCC is suppressed and the formation of large-scale connectivity is delayed, but when it forms it grows more rapidly. For large M, the recovery percolation closely resembles the td-NDP recovery in terms of LCC formation.

As the dominant cost factor in recovery of infrastructure networks is the total commodity deficit $D(t)$, this is the most important metric in network recoverability, beyond the size of the LCC. Figure 5b shows $D(t)$ reduction throughout the recovery process. As M increases, we see a closer fit with td-NDP, especially in the more expensive early stages of recovery. We observe that $M = 10$ approximates total commodity deficit quite well, which is significantly less computationally intensive than td-NDP or even the deterministic version of recovery percolation ($M = E$). This is consistent with what we found for the power grid of Shelby County (Fig. 1b), although the difference between $M = 10$ and E is less drastic in that case.

To better understand how the choice of M affects the quality of recovery percolation, we calculate the total cost C_M, defined here as the area under the curve $D(t)$ over time (i.e., $C_M = \sum_t D(t)$) as a function of M. Figure 5c shows that C_M rapidly approaches C_∞, its value at $M = E$. For this particular case, we only need to consider $M = 20$, that is less than 2% of edges, at each timestep to get within 10% of the optimal cost.

It is worth highlighting that, for sufficiently large M (but still small compared to the total number of edges E), recovery percolation captures the essential properties of the td-NDP process for $T = 2$, despite the fact that recovery percolation only considers commodity deficit, while td-NDP takes into account such details as heterogeneous repair costs of individual power lines, operational costs, performs network flows, and selects optimal recovery actions considering two timesteps.

Effect of network structure on recovery percolation. Recovery percolation together with the synthetic transmission power grids provide a stylized model to extract key network features that impact the efficiency of the restoration process. For this we systematically investigate how typical structural features affect the following quantities:

1. Total optimal cost of recovery C_∞, which is the minimum cost obtainable with recovery percolation ($M = E$).
2. Time to recovery t_{90}, the number of timesteps needed to reduce total commodity deficit by 90% percent.
3. Characteristic M^*, which captures the approximability of the process. It is defined as the smallest M value for which $C_M \leq 1.2 C_\infty$.

7
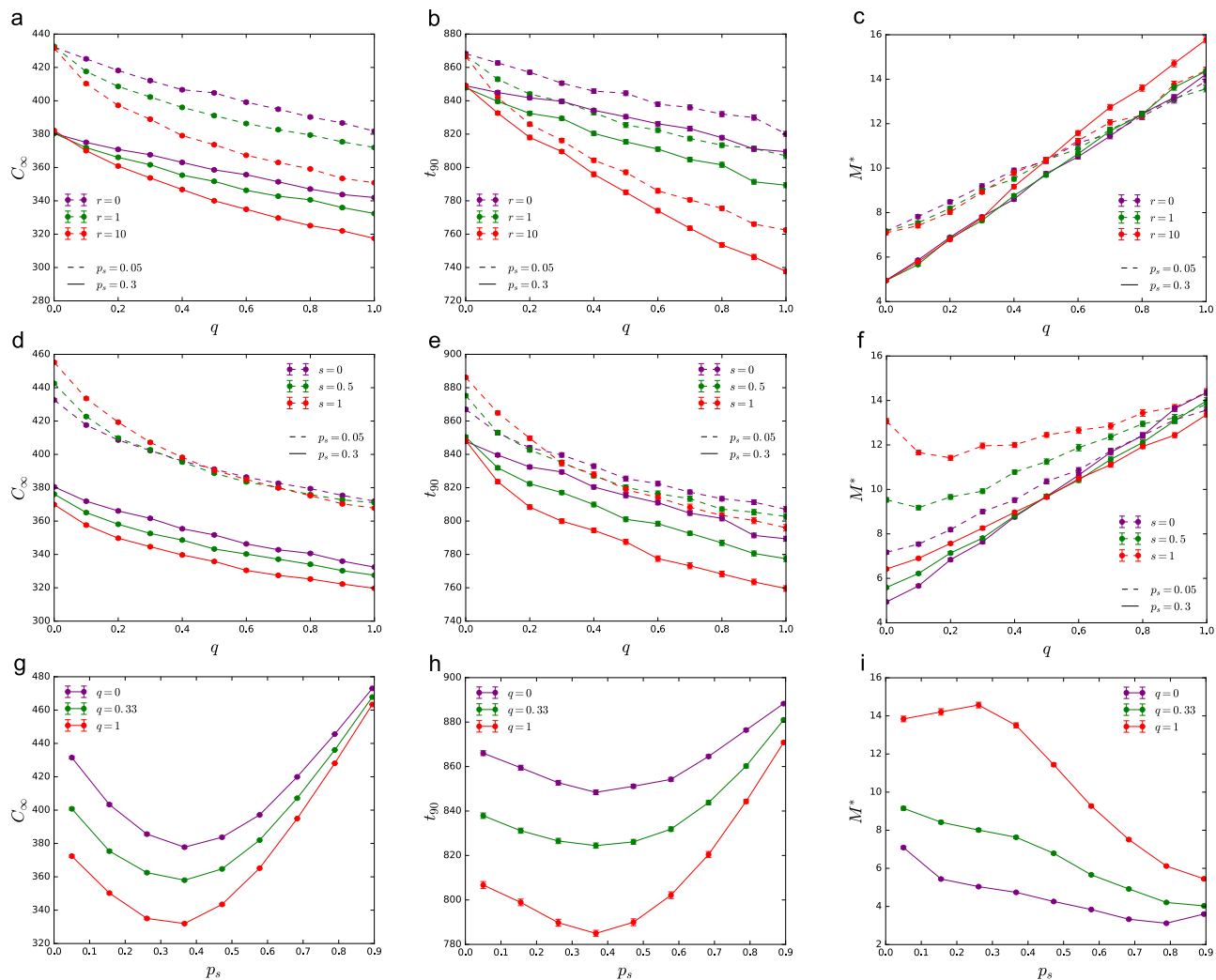
Figure 6. Effect of network structure on recovery. (**a–f**) Redundancy q reduces optimal cost C_∞ and recovery time t_{90} and increases M^*, the number of edges needed to be sampled to approximate the optimal solution. This effect is independent of the value of r, s, and p_s (dashed and solid lines). (**g–h**) Centralized and unrealistically distributed supply both increase C_∞ and recovery time t_{90}, with a minimum at intermediately distributed supply. (**i**) Increasing p_s, improves approximability. Total cost C_∞ is measured in units of (total demand × recovery timestep). Networks were generated using parameters $N = 1000$, $q = 0.33$, $r = 1$, $s = 0$, and $p_s = 0.3$, unless otherwise indicated in the figure. Each data point is an average of 100 independent network realizations; the error bars indicate the standard error of the mean.

Simulations show that redundancy q, which controls the average degree, has a strong effect (Fig. 6a–c). Increasing redundancy lowers both optimal total cost C_∞ and recovery time t_{90}; however, it increases M^*, meaning that to approximate the optimal solution more edges need to be sampled. This observation is robust to the choice of other parameters. Redundancy increases possible ways to reconnect the network, allowing less costly reconstruction strategies, but this also means that more paths must be explored to pick out the optimal one. The effect of r is more subtle, we find that long range shortcuts ($r = 10$) further decrease C_∞ and t_{90}; while short cycles ($r = 0$) have the opposite effect. The value of r has little effect on M^*.

The effect of line splitting depends on both the fraction of suppliers p_s and the redundancy q (Fig. 6d–f). For centralized supply ($p_s = 0.05$), we find that in case of low redundancy, s increases cost C_∞ and recovery time t_{90}; while in case of high redundancy, s has the opposite effect, reducing C_∞ and t_{90}. Line splitting s increases the characteristic M^*, and this increase is particularly significant for low values of q. For distributed supply ($p_s = 0.3$), we find that both C_∞ and t_{90} are decreased by s independently of the value of q. While the value of M^* is increased by s for low q, and decreased for high q.

Finally, the fraction of suppliers p_s also strongly influences the recovery process (Fig. 6g–i). Total optimal cost C_∞ and recovery time t_{90} are high for very centralized (low p_s) and very distributed (high p_s) supply, with a minimum in between. If the demand and supply follow the same distribution, the minimum is at $p_s^* = 0.5$. For our choice, the demand is more heterogeneously distributed than the supply, resulting $p_s^* < 0.5$. Increasing p_s, also

allows easier approximation of the optimal solution, i.e., M^* decreases with increasing p_s (with the exception of a limited regime with high q and low p_s and low q and high p_s).

Overall, we find that high structural redundancy reduces the optimal cost and time of recovery; however, higher edge sampling M is needed to benefit from this reduction. Long range shortcuts in the network further reduce the cost, without significantly increasing M. We also benefit from distributed supply, reducing both cost and recovery time, and depending on the level redundancy, may also improve approximability.

Discussion

We investigated the problem of optimal cost reconstruction of critical infrastructure systems after catastrophic events. We started by analyzing realistic recovery strategies for a small-scale case study, the power grid of Shelby County, TN. We identified the penalty incurred for over- and undersupply of commodities as the main contribution to the cost, outranking operational and repair costs by orders of magnitude in the initial periods of recovery. Motivated by this observation, we introduced the recovery percolation model, a competitive percolation model that in addition to network structure also takes the demand and supply associated with each node into account. The advantage of our stylized model is that it is computationally tractable and easy to interpret compared to the complex optimization formulations used in the engineering literature, while adequately reproducing important features of realistic recovery processes. This allows us to identify underlying mechanisms of the recovery process. For example, we showed that component size anti-correlates with the unsatisfied demand, which suppresses the emergence of large-scale connectivity through a process analogous to explosive percolation. Such a suppression of large-scale connectivity can be in fact observed in real recovery events[18]. The model also allowed us to systematically investigate the effect of typical network characteristics on the efficiency of the recovery process using synthetic power grids, finding that high structural redundancy, long range connections, and decentralized supply benefit recovery efforts.

Although our analysis focused on transmission power grids, recovery percolation is readily applicable to other transportation networks where the commodity transported is interchangeable (i.e., demand could be satisfied from different sources) such as gas and water infrastructure or supply-chain networks. Possible future work may extend recovery percolation to networks where the items transferred each have a specific destination, for example, passengers in human transportation networks[37].

The computational complexity of identifying actionable reconstruction strategies is an open issue, especially in the case of interdependent and decentralized recovery scenarios, where systems are larger, and the optimization problem must be solved numerous times[22,30,38]. Our stylized model is efficient, but as such, it ignores certain details. For example, our approach assumes that link capacities are sufficiently large to service the network flows during the recovery process. Future work may explore how to extend recovery percolation to take into account such constraints. Similar to td-NPD, these strategies provide scenarios that may be useful for developing recovery operator based approaches to mathematically model the dynamics of recovery and enable development of data-driven control approaches[39]. Further work is needed to extend our model to simultaneous recovery of multiple critical infrastructure systems explicitly taking into account interdependencies between the systems. Competitive percolation strategies in general, can provide opportunities for modeling real-world processes. For instance, in addition to this application to recovery, there is recent work of applying competitive percolation strategies to suppress the outbreak of epidemics via targeted immunization[40].

Methods

Time-dependent NDP.   Here, we define our benchmark model for network recovery: the time-dependent network design problem (td-NDP). Our version follows the more general formulation developed by Gonzalez *et al.*[22,23,30]. The td-NDP takes a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is a set of nodes, and $\mathcal{E}$ is the set of edges connecting nodes. At the beginning of the recovery process the td-NDP uses the destroyed graph, $\mathcal{G}' = (\mathcal{N}', \mathcal{E}')$, where $\mathcal{N}'$ and $\mathcal{E}'$ represents the nodes and edges that are not functioning, respectively. The objective function (cf. Eq. (2a)) minimizes the total reconstruction cost over a given time domain $\mathcal{T}$ with $t \in \mathcal{T}$, which includes the cost to repair nodes, q_{it}, cost to repair edges, f_{ijt}, cost of flow on each edge, c_{ijt}, and oversupply and undersupply penalties for each node, M_{it}^+ and M_{it}^-. These costs usually depend on multiple factors, such as the level of damage, the type and size of the components to be restored, their geographical accessibility, the amount of labor and resources required, and the social vulnerability of the affected areas, among others[22,41,42]. To keep track of demand satisfaction, each node also has a supply capacity (demand if negative), b_{it}. In the most general formalization of the problem, node supply b_{it} can depend on time t, but in this paper we only consider constant values. The variables δ_{it}^+ (δ_{it}^-) account for oversupply (or undersupply) of node i. We refer to the sum of the absolute values of oversupply and undersupply ($|\delta_{it}^+| + |\delta_{it}^-|$) as the commodity deficit of node i. The td-NDP includes as decision variables the amount of flow on each edge, x_{ijt}, whether or not a node i [edge (i,j)] is chosen to be recovered at timestep t, $\widetilde{w}_{it}$ ($\widetilde{y}_{ijt}$), and whether or not a node i [edge (i,j)] is functional at timestep t, w_{it} (y_{ijt}). Constraints 2b–2o are imposed to ensure that conservation of flow properties are held and that only recovered and functional nodes can produce or consume flow.

The td-NDP formulation is a mixed integer program, which has been shown to be, in general, NP-hard (and becomes exponentially harder as $|\mathcal{T}|$ and $|\mathcal{G}'|$ grows). The number of variables and constraints also become larger as the input graph becomes larger. For many reasonable size problems, computing a global optimal (i.e., where $\mathcal{T}$ contains the entire time horizon for recovery) is intractable. Therefore, heuristics are used to restrict the size of $\mathcal{T}$ by dividing the total recovery time into smaller windows, and finding the locally optimal solutions within these windows[22]. It has been shown that such heuristic finds solutions very close to the optimal; however, the computational complexity is still relatively high as a result of the underlying mixed-integer program, underscoring the need for efficient approximate methods.

Supply and demand distribution. For our computational experiments, we generate our demand distribution by following the load distribution of the European power grid[42]. This dataset was chosen due to its large system size ($N = 1463$, $E = 2199$) and its high resolution. Our goal is not to identify the true analytic form of the load distribution, but to generate statistically similar samples through bootstrapping. We found that an exponentiated Weibull distribution of the form $f(x, a, c) = ac(1 - \exp(-x^c)^{(a-1)}) \exp(-x^c) x^{(c-1)})$, where $a = 3.59$ and $c = 0.8$ well approximates the features of the demand distribution. Suppliers' capacities are uniformly distributed to balance the total demand. We set the ratio of suppliers (0.3) to consumers (0.7) according this dataset, unless otherwise noted. We assume that the link capacities are sufficiently large to service the network flows in normal operation and during the recovery process.

$$\text{td-NDP} = \text{minimize} \sum_{t \in \mathcal{T}|t>0} \left(\sum_{(i,j) \in \mathcal{E}'} f_{ijt} \widetilde{y}_{ijt} + \sum_{i \in \mathcal{N}'} q_{it} \widehat{w}_{it} \right)$$
$$+ \sum_{t \in \mathcal{T}} \left(\sum_{i \in \mathcal{N}} (M_{it}^+ \delta_{it}^+ + M_{it}^- \delta_{it}^-) + \sum_{(i,j) \in \mathcal{E}} c_{ijt} x_{ijt} \right) \tag{2a}$$

subject to,

$$\sum_{j:(i,j) \in \mathcal{E}} x_{ijt} - \sum_{j:(j,i) \in \mathcal{E}} x_{jit} = b_{it} - \delta_{it}^- + \delta_{it}^+, \qquad \forall i \in \mathcal{N}, \forall t \in \mathcal{T}, \tag{2b}$$

$$x_{ijt} \leq u_{ijt} w_{it}, \qquad \forall (i,j) \in \mathcal{E}, \forall t \in \mathcal{T}, \tag{2c}$$

$$x_{ijt} \leq u_{ijt} w_{jt}, \qquad \forall (i,j) \in \mathcal{E}, \forall t \in \mathcal{T}, \tag{2d}$$

$$x_{ijt} \leq u_{ijt} y_{ijt}, \qquad \forall (i,j) \in \mathcal{E}, \forall t \in \mathcal{T}, \tag{2e}$$

$$\sum_{i \in \mathcal{N}'} \widehat{w}_{it} + \sum_{(i,j) \in \mathcal{E}'} \widetilde{y}_{ijt} \leq 1 \qquad \forall t \in \mathcal{T}|t > 0, \tag{2f}$$

$$w_{it} \leq \sum_{\tilde{t}=1}^{t} \widehat{w}_{i\tilde{t}}, \qquad \forall i \in \mathcal{N}', \forall t \in \mathcal{T}|t > 0, \tag{2g}$$

$$y_{ijt} \leq \sum_{\tilde{t}=1}^{t} \widetilde{y}_{ij\tilde{t}}, \qquad \forall (i,j) \in \mathcal{E}', \forall t \in \mathcal{T}|t > 0, \tag{2h}$$

$$\delta_{it}^+ \geq 0, \qquad \forall i \in \mathcal{N}, \forall t \in \mathcal{T}, \tag{2i}$$

$$\delta_{it}^- \geq 0, \qquad \forall i \in \mathcal{N}, \forall t \in \mathcal{T}, \tag{2j}$$

$$x_{ijt} \geq 0, \qquad \forall (i,j) \in \mathcal{E}, \forall t \in \mathcal{T}, \tag{2k}$$

$$w_{it} \in \{0, 1\}, \qquad \forall i \in \mathcal{N}, \forall t \in \mathcal{T}, \tag{2l}$$

$$y_{ijt} \in \{0, 1\}, \qquad \forall (i,j) \in \mathcal{E}, \forall t \in \mathcal{T}, \tag{2m}$$

$$\widehat{w}_{it} \in \{0, 1\}, \qquad \forall i \in \mathcal{N}', \forall t \in \mathcal{T}, \tag{2n}$$

$$\widetilde{y}_{ijt} \in \{0, 1\}, \qquad (i,j) \in \mathcal{E}', \forall t \in \mathcal{T}. \tag{2o}$$

References
1. Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nat.* **406**, 378–382 (2000).
2. Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.* **85**, 4626 (2000).
3. Motter, A. E. & Lai, Y.-C. Cascade-based attacks on complex networks. *Phys. Rev. E* **66**, 065102 (2002).
4. Li, D., Yinan, J., Rui, K. & Havlin, S. Spatial correlation analysis of cascading failures: Congestions and blackouts. *Sci. Reports* **4** (2014).
5. Zhao, J., Li, D., Sanhedrai, H., Cohen, R. & Havlin, S. Spatio-temporal propagation of cascading overload failures in spatially embedded networks. *Nat. Commun.* **7** (2016).

6. Zhong, J. Restoration of interdependent network against cascading overload failure. *Phys. A: Stat. Mech. its Appl.* **514**, 884–891 (2018).
7. Latora, V. & Marchiori, M. Efficient behavior of small-world networks. *Phys. Rev. Lett.* **87**, 198701 (2001).
8. Ash, J. & Newth, D. Optimizing complex networks for resilience against cascading failure. *Phys. A: Stat. Mech. its Appl.* **380**, 673–683 (2007).
9. Holme, P., Kim, B. J., Yoon, C. N. & Han, S. K. Attack vulnerability of complex networks. *Phys. Rev. E* **65**, 056109 (2002).
10. Majdandzic, A. *et al.* Spontaneous recovery in dynamical networks. *Nat. Phys.* **10**, 34 (2014).
11. Hu, F., Yeung, C. H., Yang, S., Wang, W. & Zeng, A. Recovery of infrastructure networks after localised attacks. *Sci. Reports* **6** (2016).
12. Shang, Y. Localized recovery of complex networks against failure. *Sci. Reports* **6** (2016).
13. Di Muro, M., La Rocca, C., Stanley, H., Havlin, S. & Braunstein, L. Recovery of interdependent networks. *Sci. Reports* **6** (2016).
14. Quattrociocchi, W., Caldarelli, G. & Scala, A. Self-healing networks: Redundancy and structure. *Plos One* **9** (2014).
15. Panteli, M., Trakas, D. N., Mancarella, P. & Hatziargyriou, N. D. Boosting the power grid resilience to extreme weather events using defensive islanding. *IEEE Transactions on Smart Grid* **7**, 2913–2922 (2016).
16. Mureddu, M., Caldarelli, G., Damiano, A., Scala, A. & Meyer-Ortmanns, H. Islanding the power grid on the transmission level: less connections for more security. *Sci. Reports* **6**, 34797 (2016).
17. National Research Council. *Terrorism and the electric power delivery system* (National Academies Press, 2012).
18. Rudnick, H., Mocarquer, S., Andrade, E., Vuchetich, E. & Miquel, P. Disaster management. *IEEE Power Energy Mag.* **9**, 37–45 (2011).
19. Johnson, D., Lenstra, J. & Kan, A. The complexity of the network design problem. *Networks* **8**, 279–285 (1978).
20. Balakrishnan, A., Magnanti, T. L. & Wong, R. T. A dual-ascent procedure for large-scale uncapacitated network design. *Oper. Res.* **37**, 716–740 (1989).
21. Nurre, S. G., Cavdaroglu, B., Mitchell, J. E., Sharkey, T. C. & Wallace, W. A. Restoring infrastructure systems: An integrated network design and scheduling (inds) problem. *Eur. J. Oper. Res.* **223**, 794–806 (2012).
22. González, A. D., Dueñas-Osorio, L., Sánchez-Silva, M. & Medaglia, A. L. The interdependent network design problem for optimal infrastructure system restoration. *Comput. Civ. Infrastructure Eng.* **31**, 334–350 (2016).
23. González, A. D., Dueñas-Osorio, L., Sánchez-Silva, M. & Medaglia, A. L. The time-dependent interdependent network design problem (TD-INDP) and the evaluation of multi-system recovery strategies in polynomial time. *The 6th Asian-Pacific Symp. on Struct. Reliab. its Appl.* 544–550 (2016).
24. Poorzahedy, H. & Rouhani, O. M. Hybrid meta-heuristic algorithms for solving network design problem. *Eur. J. Oper. Res.* **182**, 578–596 (2007).
25. Gallo, M., D'Acierno, L. & Montella, B. A meta-heuristic approach for solving the urban network design problem. *Eur. J. Oper. Res.* **201**, 144–157 (2010).
26. Li, D., Zhang, Q., Zio, E., Havlin, S. & Kang, R. Network reliability analysis based on percolation theory. *Reliab. Eng. Syst. Saf.* **142**, 556–562 (2015).
27. Krapivsky, P. L., Redner, S. & Ben-Naim, E. *A kinetic view of statistical physics* (Cambridge University Press, 2010).
28. Achlioptas, D., D'Souza, R. M. & Spencer, J. Explosive percolation in random networks. *Sci.* **323**, 1453–1455 (2009).
29. D'Souza, R. M. & Nagler, J. Anomalous critical and supercritical phenomena in explosive percolation. *Nat. Phys.* **11**, 531–538 (2015).
30. González, A. D., Chapman, A., Dueñas-Osorio, L., Mesbahi, M. & D'Souza, R. M. Efficient infrastructure restoration strategies using the recovery operator. *Comput. Civ. Infrastructure Eng.* **32**, 991–1006 (2017).
31. Gomez, C., González, A. D., Baroud, H. & Bedoya-Motta, C. D. Integrating Operational and Organizational Aspects in Interdependent Infrastructure Network Recovery. *Risk Analysis* (2019).
32. González, A. D. *Resilience Optimization of Systems of Interdependent Networks*, Ph.D. dissertation, Rice University, Houston, Texas http://hdl.handle.net/1911/105511 (2017).
33. Wang, Z., Scaglione, A. & Thomas, R. J. Generating statistically correct random topologies for testing smart grid communication and control networks. *IEEE transactions on Smart Grid* **1**, 28–39 (2010).
34. Schultz, P., Heitzig, J. & Kurths, J. A random growth model for power grids and other spatially embedded infrastructure networks. *The Eur. Phys. J. Special Top.* **223**, 2593–2610 (2014).
35. Li, J., Dueñas-Osorio, L., Chen, C., Berryhill, B. & Yazdani, A. Characterizing the Topological and Controllability Features of U.S. Power Transmission Networks Physica A: Statistical Mechanics and Its Applications **453**, 84–98 (2016).
36. Van Mieghem, P. *Graph spectra for complex networks* (Cambridge University Press, 2010).
37. D'Souza, R. M. Curtailing cascading failures. *Science* **358**(6365), 860–861 (2017).
38. Smith, A. M., González, A. D., Dueñas-Osorio, L. & D'Souza, R. M. Interdependent network recovery games. *Risk Analysis* (2017).
39. Chapman, A., González, A. D., Mesbahi, M., Dueñas-Osorio, L. & D'Souza, R. M. Data-guided control: Clustering, graph products, and decentralized control. In *Decision and Control* (CDC), 2017 IEEE 56th Annual Conference on, 493–498 (IEEE, 2017).
40. Clusella, P., Grassberger, P., Pérez-Reche, F. J. & Politi, A. Immunization and targeted destruction of networks using explosive percolation. *Phys. Rev. Lett.* **117**, 208301 (2016).
41. FEMA. Multi-hazard Loss Estimation Methodology, Earthquake Model - Technical Manual, Hazus - MH 2.1. *Tech. Rep.*, Washington D.C (2013).
42. Hutcheon, N. & Bialek, J. W. Updated and validated power flow model of the main continental european transmission network. In *PowerTech* (POWERTECH), 1–5 (IEEE, 2013).

Acknowledgements

Author Contributions

A.S. and M.P. preformed the numerical simulations and analysis; all authors contributed to the design of the study and participated in writing the manuscript.

Additional Information