Topological perturbation of complex networks

Perturbations in complex systems can deactivate some of the edges or nodes.
Edge loss: the edge is deleted
Node loss: the node and all its edges are deleted

Effects on the global topology:
• increase of path lengths,
• separation into isolated clusters.

More connected network - less effect of an edge removal
But bridges are definite points of vulnerability!
The effect of a node removal depends on the number and characteristics of its edges.
Resilience to perturbations

Topological resilience: the remaining nodes are still connected.
the average distance does not increase.

We will remove edges/nodes one by one, and look at
• the size of the giant connected component
• the average distance between nodes in the giant connected component

Influencing factors: the type of removal
the original topology
Evolution of a random graph

Assume that the connection probability is a power-law of $N$, $p = cN^z$

Assume that $z$ increases from $-\infty$ to 0

Look for trees, cycles (circuits) and cliques in the graph.

Appearance thresholds: $p \sim N^z$

The graph contains cycles of any length if $z \geq -1$
Clusters in a random graph

- If \( \lim_{N \to \infty} pN = 0 \) the graph contains only isolated trees.
- If \( p = cN^{-1} \) with \( c < 1 \) the graph has isolated trees and cycles.
- At \( p = cN^{-1} \) with \( c = 1 \) a giant cluster appears.
- The size of the giant cluster approaches \( N \) rapidly as \( c \) increases.

\[
S = (f(1) - f(c))N
\]

The graph becomes connected if

\[
\lim_{N \to \infty} \frac{p}{\ln N / N} = \infty
\]
Existence of a giant connected cluster in a general random graph

average size of clusters

\[ \langle s \rangle = 1 + \frac{G'_0(1)}{1-G'_1(1)} \]

\( G_0(x) \) – node degree generating function
\( G_1(x) \) – generating function for the degree of an edge endpoint

\[ G_0(x) = \sum_{k=0}^{\infty} P(k)x^k \]
\[ G_1(x) = \frac{\sum_{k} kP(k)x^{k-1}}{\sum_{k} kP(k)} \]

\[ \langle s \rangle \rightarrow \infty \quad \text{when} \]

\[ G'_1(1) \equiv \frac{\sum_{k} kP(k)}{\sum_{k} k(k-1)P(k)} = 1 \]

equivalent to

\[ \frac{\langle k^2 \rangle}{\langle k \rangle} = 2 \]

A giant connected component exists if the graph is sufficiently heterogeneous.
Edge removal in random graphs

Start with a connected ER random graph with conn. prob. $p$.

$$ \lim_{N \to \infty} \frac{p}{\ln N / N} = \infty $$

Remove a random fraction $f$ of the edges.

Expected result: an ER graph with conn. prob. $p(1-f)$

Connected if

$$ \lim_{N \to \infty} \frac{p(1-f)}{\ln N / N} = \infty $$

For a broad class of starting graphs, there exists a threshold edge removal probability such that if a smaller fraction of edges is removed the graph is still connected.

B. Bollobas, Random Graphs, 1985
Node removal

Removing a node deactivates all its edges.
We can expect that the effect of the node removal will depend on the number of edges it had.
The size of the connected component will decrease at least by one.

Assume we have two networks with the same number of nodes and edges, and remove a fraction $f$ of the nodes.
Can the two networks’ resilience be different?
Numerical simulations of network resilience

Two networks with equal number of nodes and edges

- ER random graph
- scale-free network (BA model)

Study the properties of the network as an increasing fraction $f$ of the nodes are removed.
Node selection: random (errors) the node with the largest number of edges (attack)

Measures: the fraction of nodes in the largest connected cluster, $S$
the average distance between nodes in the largest cluster, $l$

Random networks respond similarly to errors and attacks

The size $S$ and average path length $l$ of the largest cluster

Similar to an inverse graph evolution.
Breakdown transition in general random graphs

Consider a random graph with arbitrary $P(k_0)$

A giant cluster exists if each node is connected to at least two other nodes.

$$\frac{\langle k^2 \rangle}{\langle k \rangle} = 2$$


After the random removal of a fraction $f$ of the nodes,

$$\langle k \rangle = \langle k_0 \rangle (1 - f), \quad \langle k^2 \rangle = \langle k_0^2 \rangle (1 - f)^2 + \langle k_0 \rangle f (1 - f)$$

Breakdown threshold:

$$f_c = 1 - \frac{1}{\frac{\langle k_0^2 \rangle}{\langle k_0 \rangle} - 1}$$
Application: random graphs

Consider a random graph with connection probability $p$ such that at least a giant cluster is present in the graph.

$$\langle k_0 \rangle = pN, \quad \langle k_0^2 \rangle = (pN)^2 + pN$$

Find the critical fraction of removed nodes such that the giant cluster is destroyed.

$$f_c = 1 - \frac{1}{\langle k_0 \rangle} = 1 - \frac{1}{pN}$$

The higher the original average degree, the larger damage the network can survive.
Scale-free networks are more error tolerant, but also more vulnerable to attacks.

• blue symbols: random failure
• red symbols: targeted attack

**Attacks:** same breakdown scenario as for random graphs.
**Failures:** little effect on the integrity of the network.

Is the low peak in average distance a finite size effect?

How does the breakdown threshold depend on the size of the network and the degree exponent?
Breakdown threshold of scale-free random graphs

Scale-free random graph with

\[ P(k) = A k^{-\gamma}, \ \text{with} \ k = m, ... K \]

\[ f_c = 1 - \frac{1}{\frac{\gamma - 2}{\gamma - 3} m - 1} \]

Infinite systems with \( \gamma < 3 \) do not break down under random failure.

Real scale-free networks show the same dual behavior:

- Blue symbols: random failure
- Red symbols: targeted attack

- Break down if 5% of the nodes are eliminated selectively
- Resilient to the random failure of 50% of the nodes.

Similar results have been obtained for metabolic networks and food webs.
Robustness of scale-free networks

Failures
Topological error tolerance

\[ \gamma \leq 3 : f_c = 1 \]
(R. Cohen et al PRL, 2000)
1. Rank order the nodes by your expectation for the effect of their removal. What were your criteria in doing so?

2. For each node, determine what is the effect of its removal on the size of the connected component, and the distances on the connected component.

3. Do the results match your expectations?
Case study: NA powergrid

• Nodes: generators, transmission substations, distribution substations
• Edges: high-voltage transmission lines

• 14099 nodes: 1633 generators, 2179 distribution substations, the rest transmission substations
• 19,657 edges

• The role of the power grid is to route power from generators to distribution substations (and then to customers)
• Connected network: power from any generator is in principle accessible to any substation
Degree distribution

\[ P(k > K) \approx \exp(-0.5K) \]
Betweenness distribution and edge range

Edge range: what would the distance between the endpoints of an edge be if the edge is removed.

\[ P(l > L) \approx (2500 + L)^{-0.7} \]
Resilience of the NA powergrid

- The relevant question is whether distribution substations receive enough power
- Studied measure: how many generators can feed a given distribution substation
- Average connectivity – the fraction of generators able to feed a given substation, averaged over substations

\[ Co = \left\langle \frac{N_g^i}{N_g} \right\rangle_i \]

- Connectivity loss \( CL = 1 - \left\langle \frac{N_g^i}{N_g} \right\rangle_i \) expressed as a percentage
- Generator removal will definitely lead to connectivity loss, transmission substation removal not necessarily.
Connectivity loss for generator removal

Connectivity loss for transmission substation removal

Highest damage if the next substation to be removed is the current highest-load substation
Limitations of topological resilience

• The most relevant measure of connectivity may not be the size of the giant connected cluster

• The effects of removing a node or edge propagate through the network
  – E.g. cascading failure on the power grid, gene mutation
  – Depends on the dynamical properties of the network

• The network topology still determines the boundaries of propagating failure
Case study: Modeling cascading failures in the North American power grid

- Three types of substations within the power grid: generators, transmitters, distributors

- Assume that power is routed through the shortest paths starting from generators and ending with distributors. Thus the *betweenness (load)* of a transmission substation is assumed to be a proxy for the power flowing through it.

- Assume that each transmitter has a tolerance (ability to handle increased load) $\alpha$; so the maximum bearable load is $C=\alpha L_0$.

- Node loss will cause the (reversible) overload of frequently used transmission nodes and the rerouting of power.

Network measures

Efficiency:
Each edge has an initial efficiency $e_{ij}=1$. The efficiency degrades at overload of either $i$ or $j$: $e_{ij}^* = e_{ij} \frac{C_i}{L_i}$ and comes back to 1 if $C_i > L_i$

Path efficiency: $\varepsilon_{ij}$ harmonic sum of edge efficiencies over the path

Network efficiency
$$E = \frac{1}{N_G N_D} \sum_{i \in G_G} \sum_{j \in G_D} \varepsilon_{ij}$$
determined by the shortest paths from generators to distributing stations


• Damage: $D = \frac{E(G_0) - E(G_f)}{E(G_0)}$, normalized

efficiency loss when nodes fail and power is rerouted

One node is removed, then the node loads are recalculated, then the edge/path/network efficiencies are updated, then the node loads are recalculated … until efficiency stabilizes.
Above a critical tolerance value, the removal of a single node has little effect on network; however, below this critical tolerance value, 20% global efficiency loss possible.

Upper line: no efficiency loss after removing a node in this category
Lower curve: tolerance - dependent efficiency loss
Low Damage Node Characteristics

- Correlation between low degree, low load and little damage (see filled circles)

- 90% of no damage nodes have betweenness below 1000 and degree < 3 for generators and load below 2000 and degree = 2 for transmitters

- 72% of no-damage generators have degree 1 thus are expected to cause insignificant damage to power routing
Three separable classes of nodes

- Nodes whose removal causes little or no damage (nearly 60% of nodes)

- Nodes that follow a tolerance-dependent curve

\[ D = D_0 \left( 1 - \frac{(\alpha - 1)^\beta}{K^\beta + (\alpha - 1)^\beta} \right) \]

Maximum damage \( D_0 = 25\% \)

- Nodes that follow the curve, then jump to no damage behavior