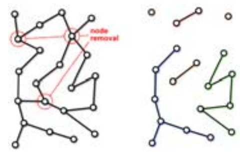
Topological perturbation of complex networks

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 studied in the literature:
the remaining nodes are still connected.
the average distance does not increase.

Ex. Propose other measures of resilience.

Testing resilience to incremental damage:
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

Ex. What factors affect the topological resilience of a network?

Review: components in a random graph

Erdős-Rényi (uniform) 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 connected component appears.
- The size of the giant connected component approaches N rapidly as c increases.
- The graph is connected if $\lim_{N\to\infty} \frac{p}{\ln N / N} = \infty$

Random graph with degree distribution P(k):

- A giant connected component exists if $\langle k^2 \rangle / \langle k \rangle \geq 2$

Ex. How is this related to topological resilience?

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 randomly chosen fraction f of the nodes.

Can the two networks' resilience be different?

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. $\frac{\langle k^2 \rangle}{\langle k \rangle} \geq 2$

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

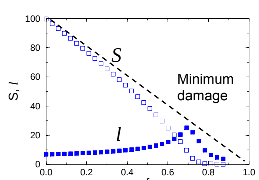
Cohen et al., Phys. Rev. Lett. 85, 4626 (2000).

Application: random graphs

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

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

$$f_c = 1 - \frac{1}{\frac{\langle k_0^2 \rangle}{\langle k_0 \rangle} - 1} = 1 - \frac{1}{pN} = 1 - \frac{1}{\langle k_0 \rangle}$$



The higher the original average degree, the larger damage the network can survive.

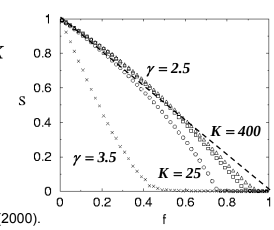
Q: How do you explain the peak in the average distance?

Breakdown threshold of scale-free random graphs

Scale-free random graph with

$P(k) = Ak^{-\gamma}$, with $k = m, \ldots K$

$$\lim_{K \to \infty} f_c = 1 - \frac{1}{\frac{\gamma - 2}{\gamma - 3} m - 1}$$



Cohen et al., Phys. Rev. Lett. 85, 4626 (2000).

Infinite scale-free networks with $\gamma < 3$ do not break down under random node failure.

Q. Do you think there is a flip side of this resilience to random node removal?

Numerical simulations of network resilience

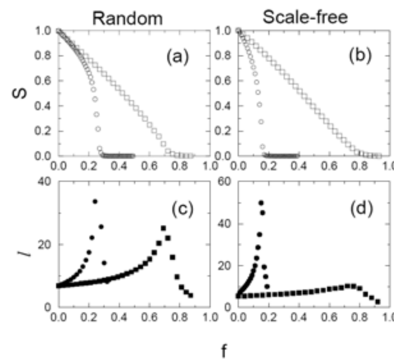
Two networks with equal number of nodes and edges
- ER random graph
- scale-free network (BA model)

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*

R. Albert, H. Jeong, A.-L. Barabási, Nature 406, 378 (2000)

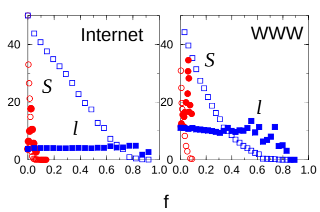
Scale-free networks are more error tolerant, but also more vulnerable to attacks



- squares: random failure
- circles: targeted attack

Failures: little effect on the integrity of the network.
Attacks: fast breakdown

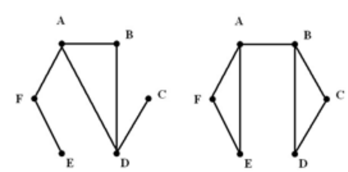
Real scale-free networks show the same dual behavior



- blue squares: random failure
- red circles: targeted attack
- open symbols: *S*
- filled symbols: *l*

- break down if 5% of the nodes are eliminated selectively (always the highest degree node)
- resilient to the random failure of 50% of the nodes.

Similar results have been obtained for metabolic networks and food webs.



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.
3. Do the results match your expectations?

Case study: NA powergrid

- 14,000 nodes: 1600 generators, 10,200 transmission substations, 2200 distribution substations
- 19,700 edges: high-voltage transmission lines
- Exponential degree distribution, long-tailed betweenness distribution

- 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
- 15% of edges are bridges.

- Q: Are there more appropriate measures of network resilience for the power grid?

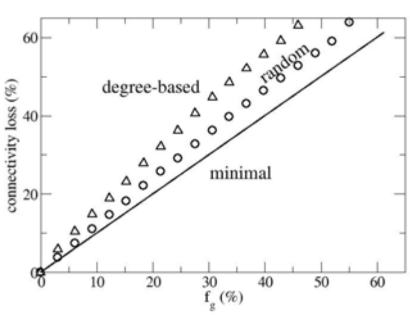
Resilience of the NA powergrid

- The relevant question is whether the 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

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

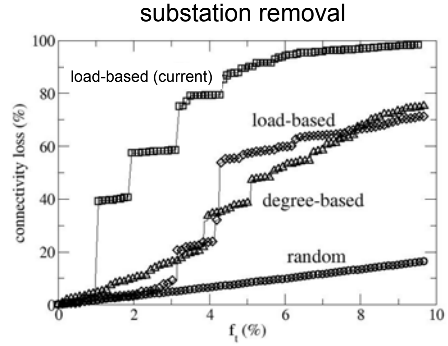
- Connectivity loss $CL = 1 - \left\langle \frac{N_g^i}{N_g} \right\rangle_i$ expressed as a percentage
- Generator removal will definitely lead to connectivity loss, transmission substation removal not necessarily.

Connectivity loss for generator removal



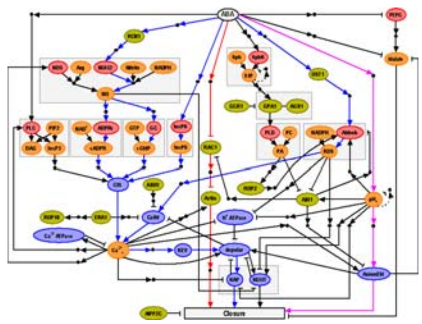
R. Albert, I. Albert, G.N. Nakarado, Phys. Rev. E (2004)

Connectivity loss for transmission substation removal



Highest damage if the next substation to be removed is the current highest-load substation

Resilience of the ABA signal transduction network



Relevant connectivity: the connection of source (ABA) to sink (closure)

At least four separate ABA-closure paths, through Ca^{2+}, through actin, through pH_c and through malate.

4 nodes (e.g depolar, actin, pH_c, malate) need to be simultaneously disrupted to block all ABA- closure paths.

Q: what other connectivity measures could be considered?

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) α; 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.

P. Crucitti, V. Latora, and M. Marchiori, Phys. Rev. E 69, 045104R (2004)

Network measures

- *Efficiency*:

Initial edge efficiency e_{ij}=1. Degrades at overload of either i or j :
$e_{ij}^{'} = e_{ij} C_i/L_i$, returns to 1 if $C_i > L_i$
Path efficiency ε_{ij} : harmonic sum of edge efficiencies over the path

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

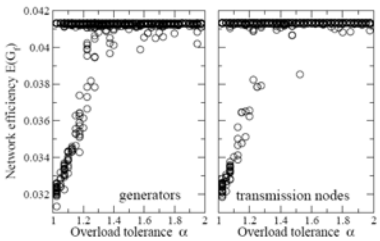
V. Latora, M. Marchiori, Phys. Rev. Lett. 87, 198701 (2001)

- *Damage*: $D = \frac{E(G_0) - E(G_f)}{E(G_0)}$, normalized efficiency loss after rerouting

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.

Single Node Random Removals

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

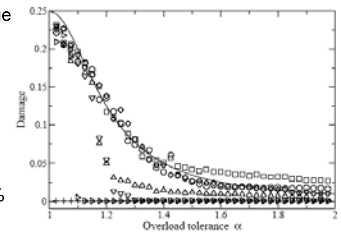
Three separable classes of nodes

- Nodes whose removal causes little or no damage (nearly 60% of nodes)
- Nodes that follow a tolerance-dependent curve

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

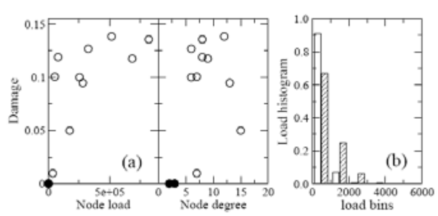
Maximum damage D_0= 25%

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



R. Kinney , P. Crucitti , R. Albert , V. Latora, Eur. Phys. J. B 46 , 101-107 (2005)

Low Damage Node Characteristics



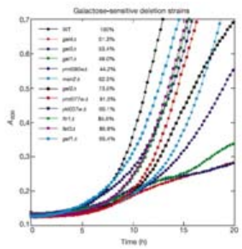
- Correlation between low degree, low load and little damage (see filled circles)
- 90% of no damage nodes have betweenness below 1000 and degree< 3 for generators and load below 2000 and degree =2 for transmitters
- 72% of no-damage generators have degree 1 thus are expected to cause insignificant damage to power routing

Resilience of cellular interaction networks

- Perturbation: knockout mutation of a gene. This means that all products of this gene (mRNA, protein) will be absent.
- Measured outcome: phenotype (e.g. growth behavior) of the mutant strain.
- The literature aims to correlate topological measures of the gene product (usually a protein in a protein interaction network) with the phenotype of the gene mutation.
- Caveats
  - The gene knockout may be incompletely represented by the loss of a protein node in a protein-protein interaction network
  - The effects of knockouts propagate through the network

Systematic deletion of *S.cerevisiae* genes

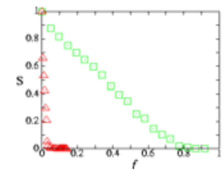
- 5196 gene knockout yeast strains
- Studied growth in rich media and altered environmental conditions
- 19% of genes essential – without them the yeast does not survive even in rich medium
- 15% of knockouts show slow growth in a rich medium
- 15% of strains show morphological alteration – different cell size/shape



Giaever *et al.* 2002 Nature 418: 387

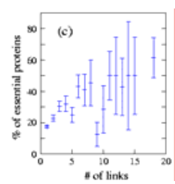
Correlating yeast gene essentiality and protein degree

Start with yeast protein interaction network and knowledge of essential genes.
The network topology displays the error tolerant/attack susceptible behavior seen in other networks.



Green – random node removal
Red – removal of highest degree node at each step.

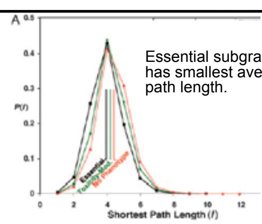
Group proteins by degree, determine the percentage of essential genes (that encode these proteins) in each group.



Highly connected proteins are more **essential**

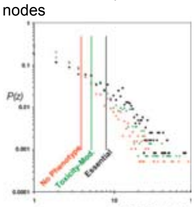
H. Jeong et al., Nature 411, 41 (2001)

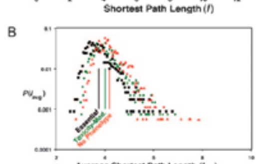
Essential – lethal after deletion
Toxicity modulating – growth inhibition of mutant after exposure to DNA damaging agent
No phenotype
Construct subgraphs of these three node types.
All have giant connected components with >60% of nodes



Essential subgraph has smallest average path length.

Essential subgraph has highest average degree





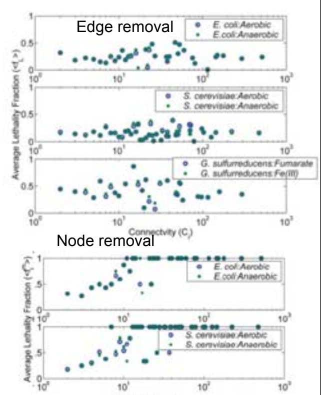
High sensitivity corresponds to higher connectivity, degree and shorter characteristic path length

Distribution of each node's average distance from every other reachable node.

Said, *et al.* 2004 PNAS 101: 18006

Resilience of metabolic networks

- Nodes – metabolites, edges – reactions
- Gene knockouts – removal of the reaction catalyzed by enzyme
- Consider edge removal (=gene knockout) and node (metabolite) removal
- Determine the lethality fraction of edge or node removal from nodes of given degree
- Relatively narrow range of lethality fraction in case of edge removal
- Very highly connected metabolites are 100% lethal, but…
- The lethality fraction of some less connected nodes is higher than the lethality fraction of more connected nodes.



Mahadevan *et al*. 2005 Biophys. Jour. 88: L07-09