Social Network Analysis: Unit 5

Katia Papakonstantinopoulou

AUEB, Master in Data Science, November 8, 2023

Properties	Models	Algorithms
Small diameter Edge clustering	Small-world model Erdös-Rényi model	Decentralized search
Scale-free	Preferential attachment Copying model	PageRank Hubs and authorities
Weak ties' strength Core-periphery	Kronecker graphs	Community detection
Densification power-law Shrinking diameters	Microscopic model of evolving networks	Link prediction Supervised random walks
Information virality	Independent cascade model Game-theoretic model	Influence maximization Outbreak detection

Outline:

- Network resilience
- Link prediction and network inference
- Social learning on networks

Network Resilience

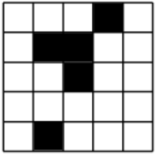
In real world, networks can be vulnerable!
How resilient are they to node/edge losses?

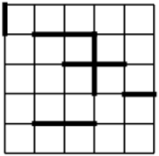
- How node or edge percolation affects the network functionality?
- Resilience of randomly vs. preferentially grown networks
- Resilience to random failures vs. targeted attacks

Assume that a fraction of nodes or edges are removed.
In the remaining graph,

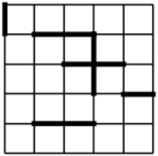
- how large are the connected components?
- what is the average distance between nodes in the components?

This problem is related to percolation.





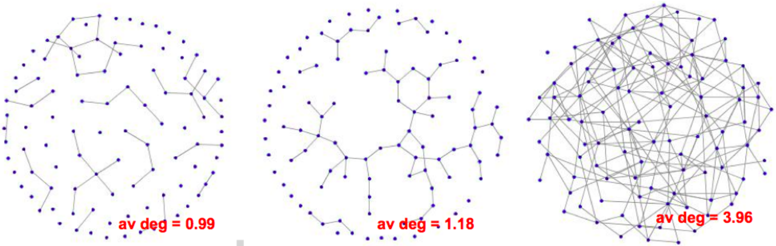
Edge removal and edge percolation



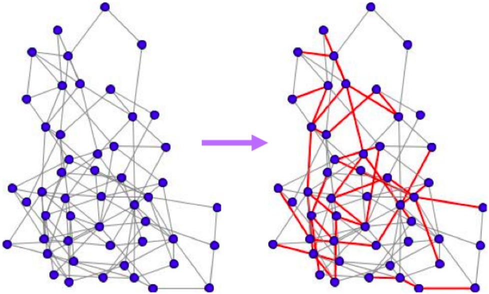
- bond percolation: each edge is removed with probability $(1 - p)$
  It corresponds to random failure of links.
- targeted attack: causing the most damage to the network with the removal of the fewest edges
  - strategies: remove edges that are most likely to break apart the network or lengthen the average shortest path
    e.g. edges with high betweenness

Attacks: reverse process of adding edges

As the average degree increases to 1, a giant component suddenly appears.
Edge removal is the opposite process – at some point the average degree drops below 1 and the network becomes disconnected.



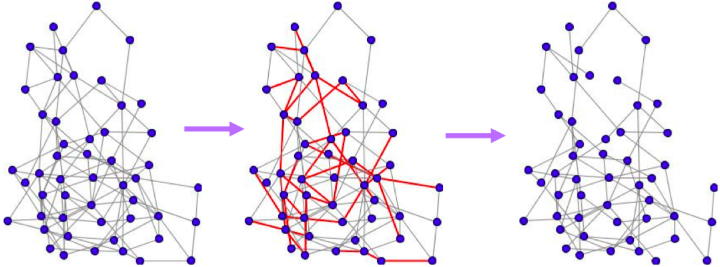
Quiz



Quiz

In this network each node has average degree 4.64.
If you removed 25% of the edges, by how much would you reduce the giant component?

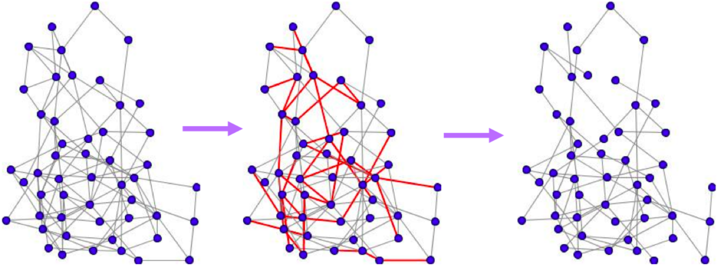
Quiz (cont.)



Answer

- 50 nodes, 116 edges, average degree 4.64
- After the 25% edge removal:
  76 edges, average degree 3.04 – still well above percolation threshold

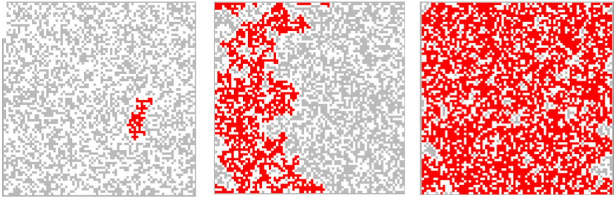
Quiz (cont.)



Answer

- 50 nodes, 116 edges, average degree 4.64
- After the 25% edge removal:
  76 edges, average degree 3.04 – still well above percolation threshold

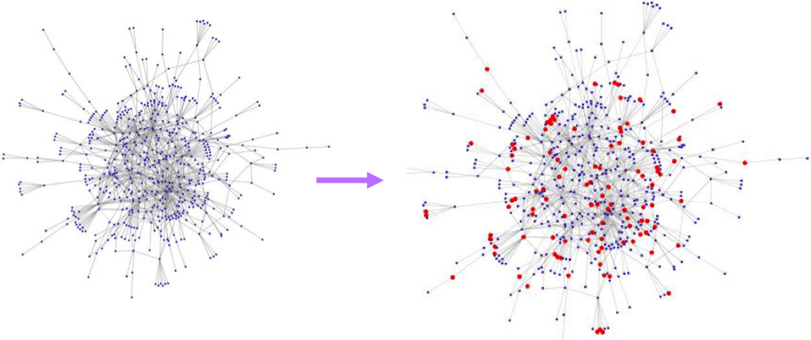
Ordinary Site Percolation on Lattices:
Fill in each site (site percolation) with probability p



- low p: small islands
- p critical: giant component forms, occupying finite fraction of infinite lattice
- p above critical value: giant component occupies an increasingly larger portion of the graph

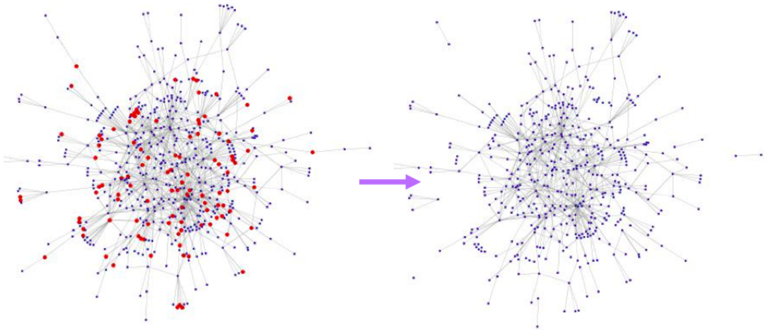
http://www.ladamic.com/netlearn/NetLogo501/LatticePercolation.html

Percolation on networks



- Percolation can be extended to networks of arbitrary topology.
- We say the network percolates when a giant component forms.

Example: gnutella filesharing network, 20% of nodes removed at random



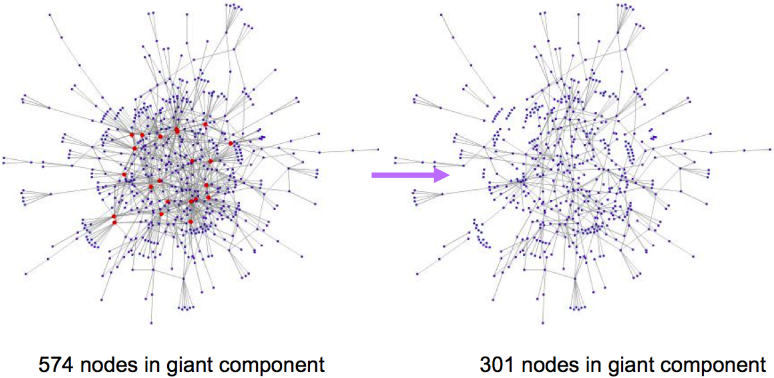
574 nodes in giant component

427 nodes in giant component

Targeted attacks on power-law networks

Power-law networks are vulnerable to targeted attack

Example: same gnutella network, 22 most connected nodes removed (2.8% of the nodes)



Quiz

Why is removing high-degree nodes more effective?

- it removes more nodes
- it removes more edges
- it targets the periphery of the network

Answer

Removing high-degree nodes causes the removal of more edges.

Quiz

Quiz

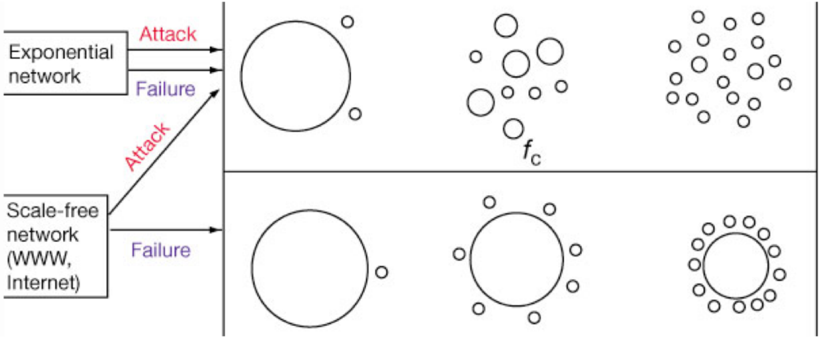
Why is removing high-degree nodes more effective?

- it removes more nodes
- it removes more edges
- it targets the periphery of the network

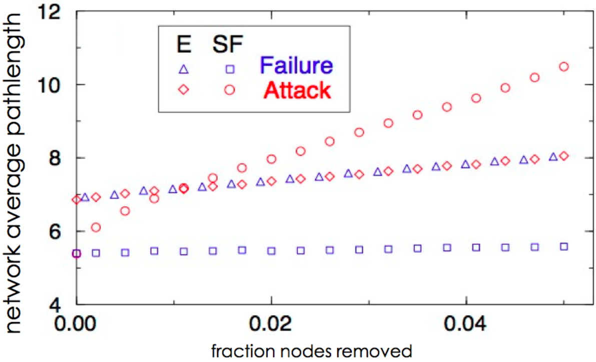
Answer

Removing high-degree nodes causes the removal of more edges.

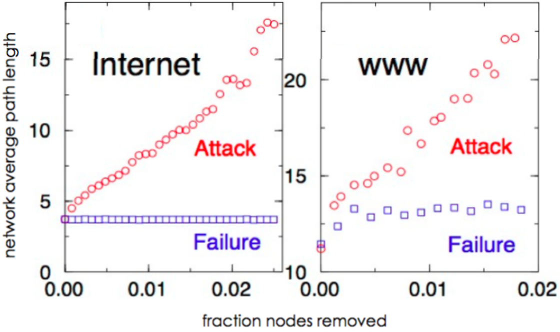
Random failures vs. Attacks



Source: Error and attack tolerance of complex networks. Reka Albert, Hawoong Jeong and Albert-Laszlo Barabasi. Nature 406, 378-382(27 July 2000); `http://www.nature.com/nature/journal/v406/n6794/abs/406378A0.html`



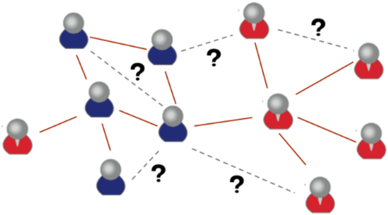
Source: Error and attack tolerance of complex networks. Reka Albert, Hawoong Jeong and Albert-Laszlo Barabasi. Nature 406, 378-382(27 July 2000); `http://www.nature.com/nature/journal/v406/n6794/abs/406378A0.html`



Source: Error and attack tolerance of complex networks. Reka Albert, Hawoong Jeong and Albert-Laszlo Barabasi. Nature 406, 378-382(27 July 2000); `http://www.nature.com/nature/journal/v406/n6794/abs/406378A0.html`

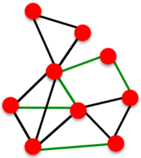
Link prediction

Given a network, which links are possible to be established next?



The link prediction problem

Given $G[t_0, t_0^{'}]$ a graph on edges up to time $t_0^{'}$, output a ranked list L of links (not in $G[t_0, t_0^{'}]$) that are predicted to appear in $G[t_1, t_1^{'}]$.



$G[t_0, t_0']$
$G[t_1, t_1']$

Evaluation of a solution: If n is the number of new edges that appear during the test period $[t_1, t_1^{'}]$, take the top n elements of L and count the correct edges.

Since network data are very sparse, we usually consider a core, e.g., we take into account only nodes with degree of at least 3 (we don't know enough about nodes with less than 3 edges to make good inferences).

Link prediction via Proximity (cont.)

Algorithm:

- For each pair of nodes (x, y) compute score $c(x, y)$
  For example, $c(x, y)$ could be the number of common neighbors of x and y
- Sort pairs (x, y) by the decreasing score $c(x, y)$
  Note: Only consider / predict edges where both endpoints are in the core (degree ≥ 3)
- Predict top n pairs as new links
- See which of these links actually appear in $G[t_1, t_1']$

Social learning

The process according to which which agents in a network learn about an underlying state by observing their neighbors' choices.

Models:

- Bayesian social learning
- Myopic (Non-Bayesian) social learning
- Bayesian observational social learning
  Users observe past actions – Most relevant for markets
- Bayesian communication social learning
  Communication of beliefs or estimates – Most relevant for friendship networks

- It was first introduced by DeGroot (1974) and more recently analyzed by Golub and Jackson (2007).
- Beliefs are updated by taking weighted averages of the neighbors' beliefs
- A finite set $\{1, \ldots, n\}$ of agents
- Interactions captured by an $n \times n$ nonnegative interaction matrix W

  $w_{ij} > 0$ indicates the trust or weight that i puts on j

  W is a stochastic matrix (row sum=1)
- There is an underlying state of the world $\theta \in R$
- Each agent has initial belief $z_i(0)$; we assume $\theta = \frac{1}{n} \sum_{i=1}^{n} z_i(0)$
- Each agent at time k updates his belief $z_i(k)$

Important applications of SNA

Some practical applications of SNA:

- Identifying important people (those linked by strong social ties) within an individual's network neighborhood, e.g., romantic partners
- Identifying influential people in a network
- Detecting outbreaks of contagious diseases

Research trends

- Market trends identification
- Identification of opinion trends
- Biological networks
- etc

Unit 5: Recap

We discussed the following aspects:

- Network resilience
- Link prediction and network inference
- Social learning on networks

Concluding Remarks

Thank you for attending the course!

What I hope you got out of it:

- What are the key properties of empirically observed networks.
- What are the processes that shape them and help us understand the empirical observations.
- What implications of the processes on the functions the network is supposed to carry out.
- Research trends (also biological networks!).

Textbooks:

- David Easley and Jon Kleinberg, Networks, crowds, and markets, Cambridge University Press, 2010.
- Jure Leskovec, Anand Rajaraman and Jeffrey David Ullman, Mining of massive datasets, Cambridge University Press, 2014.
- M. E. J. Newman, Networks – An Introduction, Oxford University Press, 2010 (deriving knowledge using math and analytics, community detection, etc).
- Matthew O. Jackson, Social and Economic Networks, Princeton University Press, 2008 (how is a network created by self-interested users?).
- Albert-Làszlò Barabàsi, Network Science, Cambridge University Press, 2015 (for beginners).