Robustness and Miscellaneous topics

Social and Technological Networks

Rik Sarkar

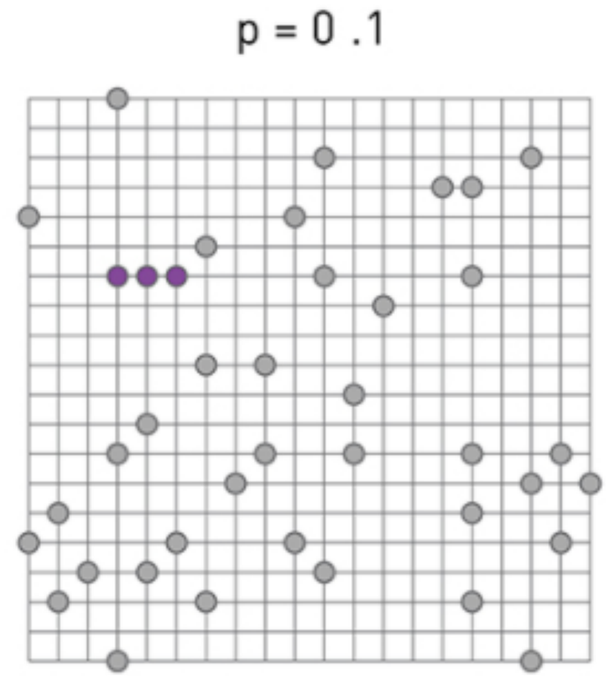
University of Edinburgh, 2017.

Robustness, redundancy

- Ecology and environment
  - Think food webs, dependencies, symbiosis
- Biology
  - Metabolic networks
- Engineering
  - Communication networks, Internet routing
  - Road networks, infrastructure, supply chains
- http://barabasi.com/networksciencebook/

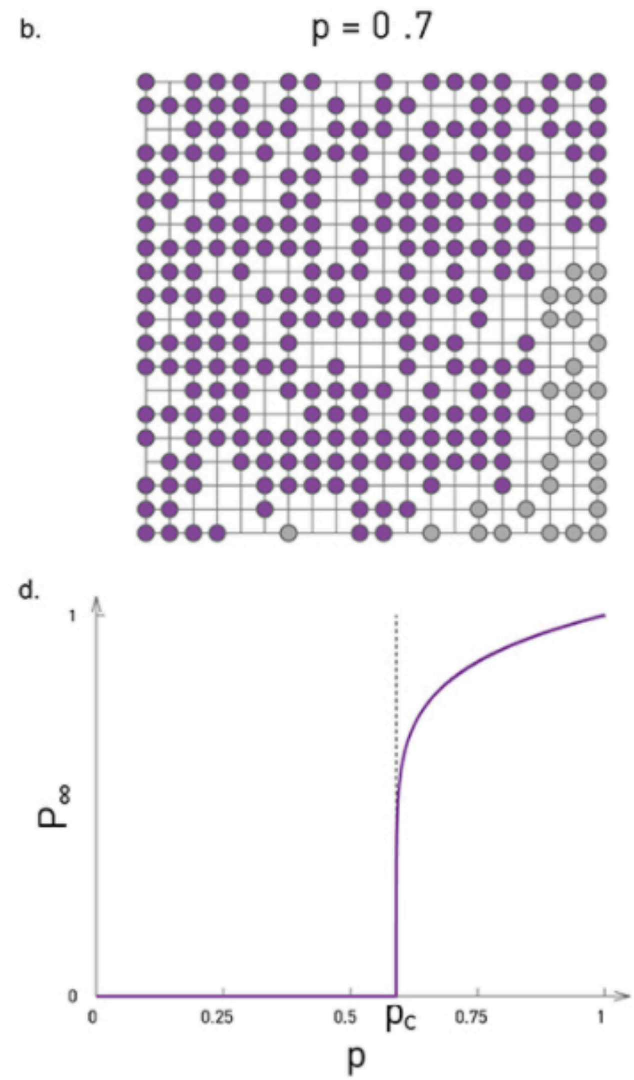
What is the probability that a graph is connected?

- We have seen emergence of giant component in random graphs
  - Phase transition at p=1/n
- Suppose we take a grid graph
- And place a pebble on each node with probability p
  - E.g. there is an attack and each node survives with probability p
- Is there are giant component?

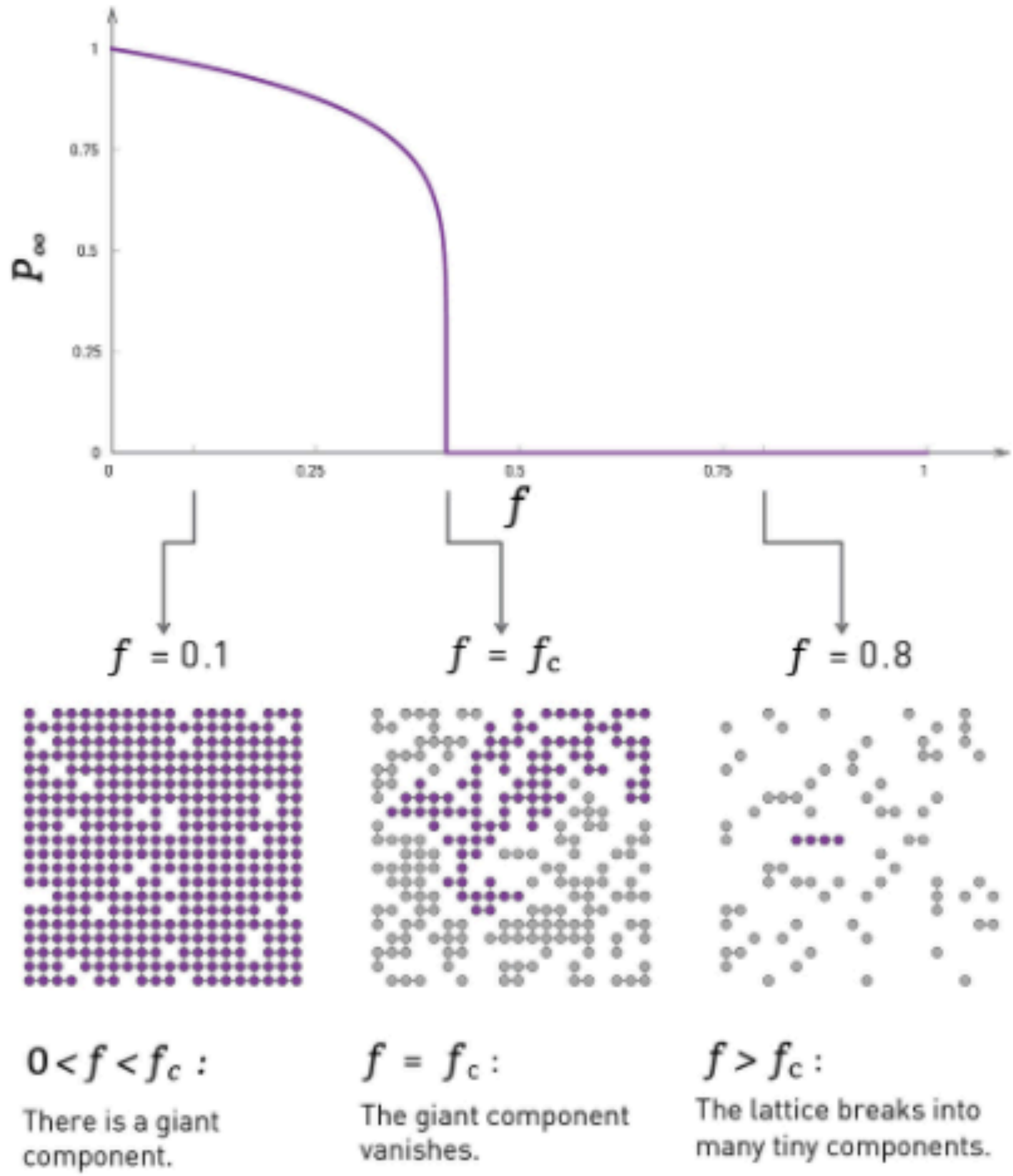


Percolation Threshold

- Yes, for $p > 0.593$
- Varies for other types of grids
  - But exists
- Percolation also shows tipping point and giant component



- S



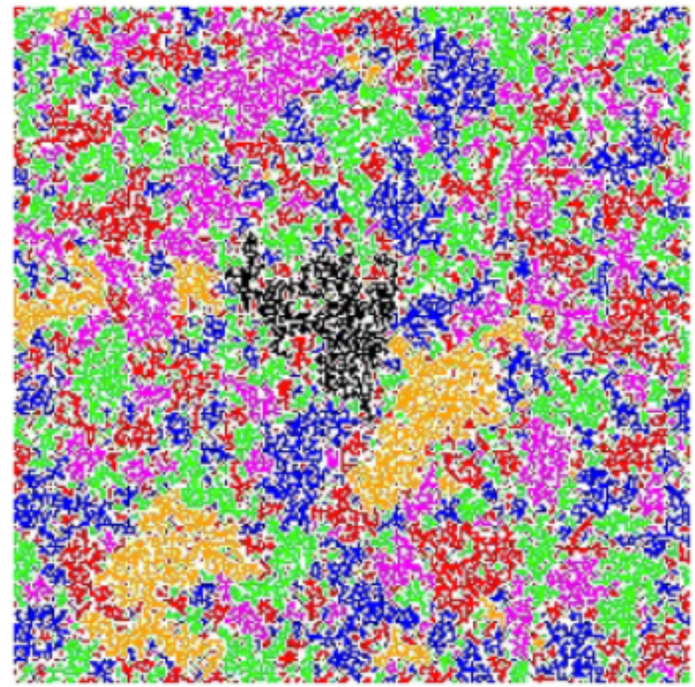
1
0.75
0.5
0.25
0
P_∞
0
0.25
0.5
0.75
1
f
$f = 0.1$
$f = f_c$
$f = 0.8$
$0 < f < f_c$:
There is a giant component.
$f = f_c$:
The giant component vanishes.
$f > f_c$:
The lattice breaks into many tiny components.

Network collapse

- May occur suddenly
- Financial or business networks may suddenly run out of money
- Ecological networks can disappear
  - https://www.youtube.com/watch?v=xZ3OmlbtaMU

What if the collapse is infective?

- Fire spreads in a forest
- A power node failing can cause other nodes to fail
- A traffic blockage at a junction can cause nearby junction to block



Infective/cascading failure

- Suppose every edge uv has a probability p_{uv} that a failure on u will cause a failure of v
- Is there a set of critical targets?
- Is there a small set of nodes that can be targeted to bring down most of the network?

- How do you solve this problem?

Infective/cascading failure

- Size of cascading failure (in power grids) observed to follow power law
  - Most failures are small
  - Some big failures

Robustness of Power law networks

- Sometimes called scale free networks
- If nodes fail randomly
  - Size of giant component decreases gradually
  - Close to zero only for large fractions of (nearly all) nodes failing

Robustness of Power law networks

- The robustness to random failure comes from low probability of hubs failing
- However, removing starting from hubs (highest degree nodes) causes rapid failure
  - Susceptible to planned attack
  - Grids on the other had do not have obvious failure points.

Link prediction

- Given a network
- Can you predict which links are likely to form in future in a reasonable time interval?
- May be because two people become friends
  - Or they are already friends, but the link becomes visible

Link prediction

- Basic idea:
  - Similar people are likely to form links
- Homophily
  - People with similar attributes/interests form links
  - If we have external attributes (locations, interests) then we use them
- Also, friends of friends often become friends
  - Predict links based on common friends and neighborhoods
  - Note that this indirectly incorporates homophily effects

Prediction methods

- Give a score to each pair of nodes based on how likely they are to form link
- Example scoring strategies:
  - Graph distance (shortest path length)
  - Number of common neighbors
  - Jaccard similarity of neighborhoods
  - Preferential attachment
  - Random walk (hitting time based methods)
    - How soon does a random walk from x hit y?
  - Others

Results

- In reality, many unknown external factors affect links
- So raw accuracy itself is low
- However, we can compare them with baselines like random links
- Most methods perform much better than random links
- Nowell, Kleinberg. Link prediction problem. CIKM 03.

Friendship paradox

- Your friends have more friends than you do!
- Are you less social than others?

Friendship paradox

- The paradox:
- If you ask everyone to report their degrees and take average, you get the average degree
- If you ask everyone to report the average degrees of their friends and take the averages of all,
  - you get more than the overall average degree!
- Most of us have some popular friends (hence they are popular)
- If you pick a random friend of a random person, (random edge)
  - This friend is relatively likely to be popular, since popular nodes have more edges

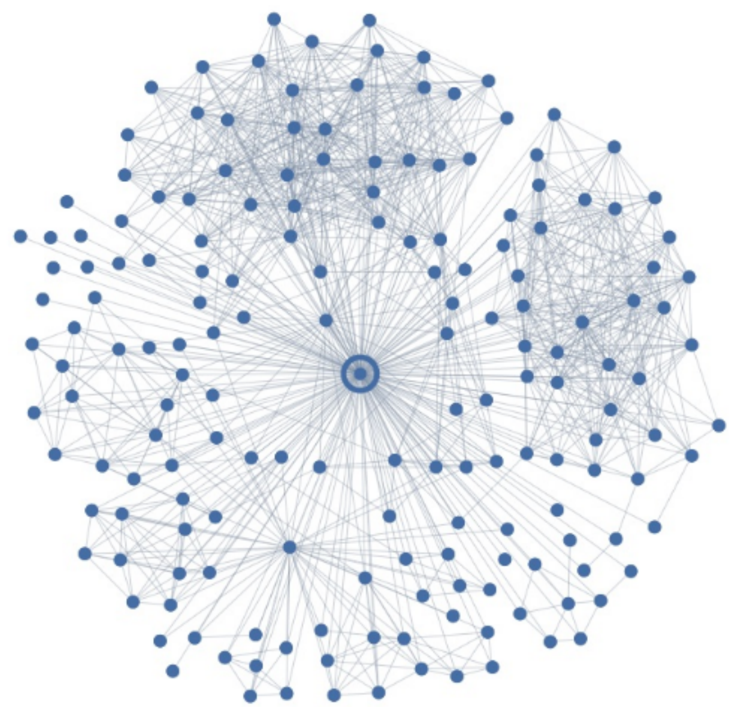
- Average degree of nodes:
- A node with degree d(v) contributes d(v) once
- Average degree of a friend:
- Each person picks a friend and counts degree
- A node with degree d(v) contributes d(v) times, with total contribution $d(v)^2$
- A few nodes with relatively high d(v) can skew the count
- https://en.wikipedia.org/wiki/Friendship_paradox
- S. L. Feld, Why your friends have more friends than you do, American journal of sociology, 1991

Identify spouses or romantic partners

Identify spouses or romantic partners

- Tie strengths are important
- Romantic ties tend to be of high strength, more likely to transmit information
- Do you expect romantic links to have high embeddedness (number/fraction of common friends)?

- People have clusters of friend circles
- Work, school, college, hobbies
- Edges in these have high embeddedness, even if they are not strong friends



- Spouses usually know some friends in each-others different circles
  - The edge does not have high embeddedness
  - Compared to links in groups such as school/ college

Dispersion

- But, it has a dispersed structure:
  - There are several mutual friends, but the mutual friends are not well connected among themselves

Dispersion

- dispersion between u,v
- Notations:
  - C(u,v): Common friends of u, v
  - G_u : Subgraph induced by u and all neighbors of u
  - d_{uv} : distance measured in G_u-{u,v}: Without using u or v

$$disp(u,v) = \sum_{s,t \in C(u,v)} d_{uv}(s,t)$$

Dispersion

$$disp(u,v) = \sum_{s,t \in C(u,v)} d_{uv}(s,t)$$

- Increases with more mutual friends
- Increases when these friends are far in the graph
- It is possible to use other distance measures
- Good results with d = 1 if no direct edge, 0 otherwise

Normalized dispersion

- Use norm(u,v) = disp(u,v)/embed(u,v)
  - 48% accuracy
- Apply recursively, to weigh higher nodes with high dispersion
  - Gives 50.5% accuracy
  - 60% accuracy for married couples
- High accuracy considering hundreds of friends
- Works better than usual machine learning based on posts, visits, photos etc
- Best results with combination of features
- Backstrom and Kleinberg. Romantic partnerships and dispersion of social ties, ACM CSCW 2014