LIFE DATA EPIDEMIOLOGY

Lecture 9: Robustness

Leonardo Badia

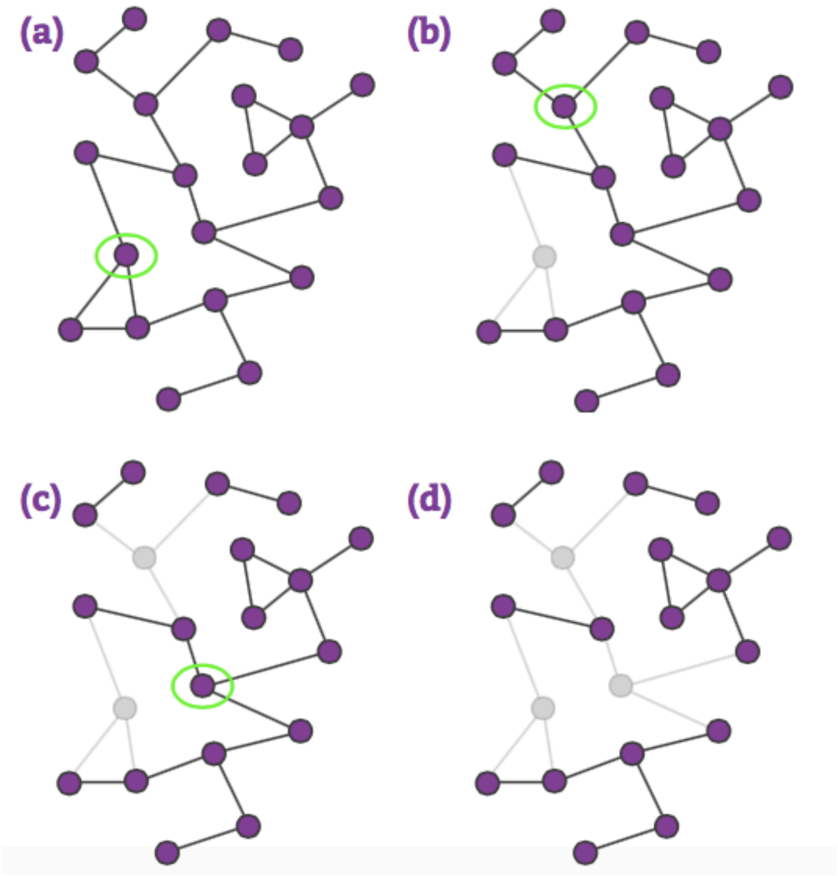
leonardo.badia@unipd.it

Robustness

- Network science is often interested in understanding robustness to failures
- Reason: real-world networks work under imperfect conditions / malfunctioning
  - technological networks are subject to link breakage or node failure
  - metabolic networks have mutations and chemical transcription mistakes
  - for epidemics, this is to contrast them!

Robustness

- What if our models are missing nodes?
- Would the network still "work"?
  - Failures can lead to either just isolating nodes/groups or breaking the whole network apart



Percolation

- A mathematical attempt can be made through percolation theory
  - Consider a lattice (e.g., a square grid)
  - each position in the lattice is occupied by a peeble with probability *p*
  - lattice links are also created automatically between positions occupied by peebles
- What is the resulting network structure?

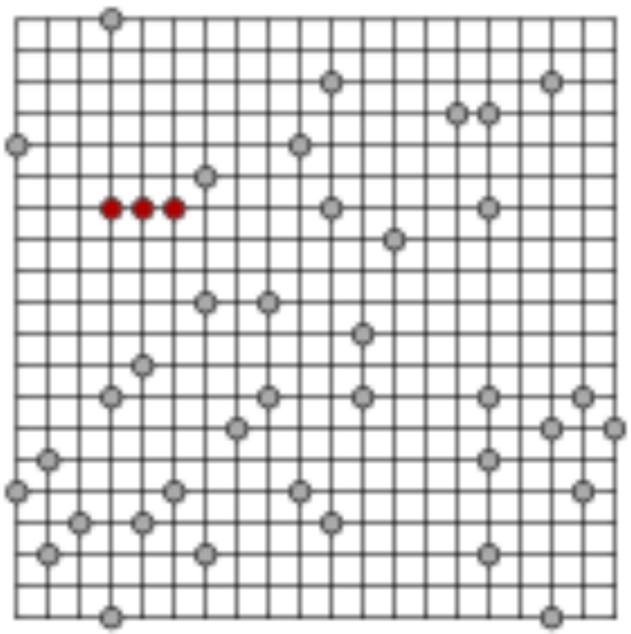
Percolation

- It can be found that the behavior is not smooth, but rather has a phase transition around a critical value p_c
- As p grows, a giant component appears with size that suddenly becomes infinite → it involves the entire lattice when $p \approx p_c$
- Other network metrics experiences a similar transition as well around value p_c

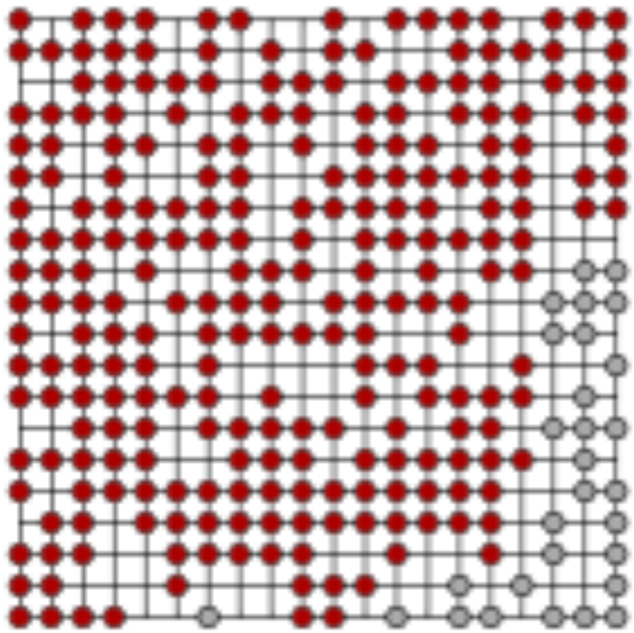
Percolation

- Critical transition for $p_c \approx 0.6$

$p = 0.1$

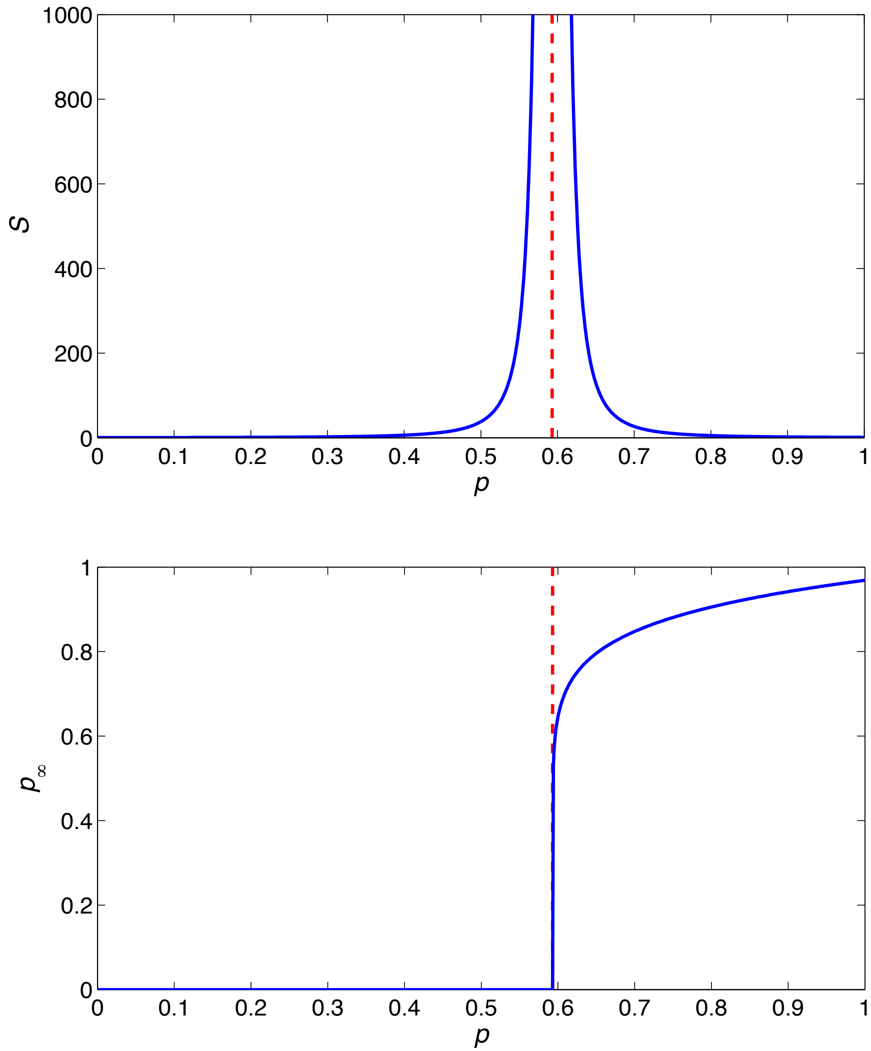


$p = 0.7$



Percolation

- At p_c a phase transition appears
- A giant component appears and many network metrics change behavior



Node vs link creation or break

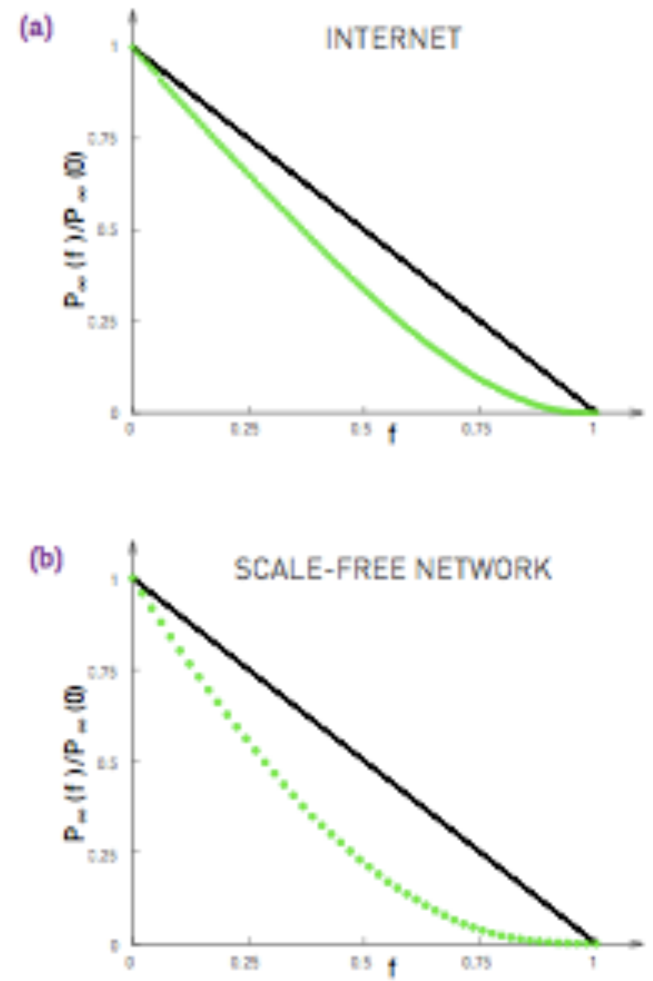
- Actually, percolation theory can be applied to two similar processes
  - node addition/removal
  - link addition/removal
- In the following, we will derive the analysis for node-based percolation, but everything is directly extendable to a link-based case
  - so that networks that are robust node-wise are also so if links are considered

Percolation → scale-free

- What if we apply node removal to scale-free networks (instead of regular lattices)?
- We observe an increased robustness
  - reason: the presence of the **hubs**, which were missing in a regular lattice
  - of course this is because removals are still entirely random, so removing a big hug is very bad, but hubs are few special nodes, so they are hard to pick going randomly

Scale-free network robustness

- Robustness of the Internet due to its scale-free nature
  - often working even during earthquakes/hurricanes
  - routers linked to the GC after random removal with rate *f* → still large if *f* < 1
  - experiments aligned with a scale-free model



Critical transition in scale-free

- Apparently, scale-free networks are critical only if fraction $f = 1-p$ of node removal is a very high value f_c (→ **breakup threshold**)

- Let us verify this analytically based on:
  - to have nodes belonging to a GC, this GC must exist in the first place
  - in a scale-free network, nodes are randomly wired (differently from a lattice): how many of them do we need to keep a GC together?

Molloy-Reed criterion

- To hold a GC together in a randomly wired network, at least 2 links needed per node
- **Molloy-Reed criterion**. Any randomly wired network has a GC if and only if:

$$\kappa = \langle k^2 \rangle / \langle k \rangle > 2$$

  - That is, networks with $\langle k^2 \rangle < 2 \langle k \rangle$ do not have a giant component and are fragmented
  - A criterion valid for any degree distribution!

Molloy-Reed criterion

- Let verify the criterion for a random graph
- Degree distribution is Poisson, so:

$$\langle k \rangle = \sigma^2 = 1/\lambda \quad \text{but} \quad \sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$$

  - Thus, $\langle k^2 \rangle = \langle k \rangle (1+\langle k \rangle)$
  - Molloy-Reed criterion implies $\langle k \rangle > 1$ which we already verified to be the condition for the existence of a GC in a random graph

Molloy-Reed criterion

- **Formal proof**. Consider node i in the GC
  - Actually, that i belongs to the GC can only be derived recursively as i being linked to j where $j \in$ GC. Write this condition as $i \rightarrow j_{\mathrm{GC}}$
- What is the average degree of the GC? It must be $\langle k_i \mid i \rightarrow j_{\mathrm{GC}} \rangle > 2$ or the GC is not held together. Thus, we need to prove

$$\langle k_i \mid i \rightarrow j_{\mathrm{GC}} \rangle = \langle k^2 \rangle / \langle k \rangle$$

Molloy-Reed criterion

- **Formal proof (cont'd)**. We can also write
$$\langle k_i \mid i \rightarrow j_{GC} \rangle = \Sigma_i \; k_i \, P(k_i \mid i \rightarrow j_{GC})$$
- If p_k = degree distribution, by Bayes' rule:
$$P(k_i \mid i \rightarrow j_{GC}) = P(i \rightarrow j_{GC} \mid k_i) \, p_{k_i} \, / \, P(i \rightarrow j_{GC})$$
- Probability of $i \rightarrow$ arbitrary node j does not involve that $j \in GC$, therefore:
$$P(i \rightarrow j_{GC}) = 2L \, / \, N \, / \, (N-1) = \langle k \rangle \, / \, (N-1)$$
$$P(i \rightarrow j_{GC} \mid k_i) = k_i \, / \, (N-1)$$
- Thus: $\langle k_i \mid i \rightarrow j_{GC} \rangle = \Sigma_i \; k_i^2 \, p_{k_i} \, / \, \langle k \rangle = \langle k^2 \rangle \, / \, \langle k \rangle$

Breakup threshold in scale-free

- What critical fraction f_c of a network can be removed without destroying the GC?
- Removing "f" of nodes changes degree $k \rightarrow k'$ (and their distribution) in two ways:
  - it erases some nodes, so there are fewer nodes with some old degree $k \rightarrow$ however this is irrelevant if removals are iid random
  - it also removes the links associated to them, thus changing their neighbors' degree

Breakup threshold in scale-free

- What is the probability that a removal of a fraction f of nodes changes $k \longrightarrow k'$?

$$P(k \rightarrow k') = \binom{k}{k'} f^{k-k'} (1-f)^{k'} \qquad (\text{for } k > k')$$

- Thus: $$p_{k'} = \sum_{k=k'}^{\infty} p_k \binom{k}{k'} f^{k-k'} (1-f)^{k'}$$

Breakup threshold in scale-free

- We use this to derive the new values of first and second moments, denoted as $\langle k' \rangle_f$ and $\langle k'^2 \rangle_f$ (to indicate removal rate f)

$$\langle k' \rangle_f = \sum_{k'=0}^{\infty} k' \sum_{k=k'}^{\infty} p_k \frac{k!}{k'!\,(k-k')!} f^{k-k'} (1-f)^{k'}$$

$$= \sum_{k'=0}^{\infty} \sum_{k=k'}^{\infty} p_k \frac{k!}{(k'-1)!\,(k-k')!} f^{k-k'} (1-f)^{k'}$$

Breakup threshold in scale-free

- Observe that the two summations over: $k' \geq 0,\ k \geq k'$ → rewrite as $k \geq 0,\ 0 \leq k' \leq k$

$$\langle k' \rangle_f = \sum_{k=0}^{\infty} k\, p_k (1-f) \boxed{\sum_{k'=0}^{k} \frac{(k-1)! \quad f^{k-k'} (1-f)^{k'-1}}{(k'-1)!\ (k-k')!}}$$

$$= \sum_{k=0}^{\infty} k\, p_k (1-f)\ \cancel{(1-f+f)^{k-1}} = (1-f)\langle k \rangle$$

Breakup threshold in scale-free

- We obtained $\langle k' \rangle_f = (1-f)\langle k \rangle$ → the new value of the average degree after node removal depends only on f and the old $\langle k \rangle$
- To derive $\langle k'^2 \rangle_f$, write it as $\langle k'(k'-1) + k' \rangle_f$
  - $\langle k'(k'-1) \rangle_f$ is obtained similar to before; the trick is to rewrite the summations in the same way, and to take out both k' and $(k'-1)$
  - this results in $\langle k'(k'-1) \rangle_f = (1-f)^2 \langle k(k-1) \rangle$
  - thus, $\langle k'^2 \rangle_f = (1-f)^2 \langle k^2 \rangle + (1-f)f\langle k \rangle$

Breakup threshold in scale-free

- Use Molloy-Reed to see if, after removing a fraction *f* of nodes, there still is a GC → breakup threshold f_c @critical point κ=2:
  $\langle k'^2 \rangle_{f_c} = (1-f_c)^{\cancel{2}} \langle k^2 \rangle + \cancel{(1-f_c)} f_c \langle k \rangle = 2\cancel{(1-f_c)}\langle k \rangle$
- Resulting in: $f_c\,(\langle k \rangle - \langle k^2 \rangle) = (2\langle k \rangle - \langle k^2 \rangle)$
  - that can be rearranged into

$$f_c = 1 - (\langle k^2 \rangle / \langle k \rangle - 1)^{-1}$$

Breakup threshold in scale-free

- Remarkably, f_c only depends on the ratio between $\langle k^2 \rangle$ and $\langle k \rangle$, so in turn only on p_k
  - E.g., for a random graph (Erdős-Rényi) we have $\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle$, hence the breakup happens for $f_c = 1 - 1 / \langle k \rangle$; i.e. still a GC as long as $1 / \langle k \rangle$ of the nodes are left alive
  - in general, $f_c = 1 - (\langle k^2 \rangle / \langle k \rangle - 1)^{-1}$ means that networks with big hubs (giving a big deviation from $\langle k \rangle$) are hard to die

Example of applications

- Robustness aligned with theory:
  - The Internet survives without 92% nodes
  - The citation network has $f_c = 96\%$
  - The actor network $f_c = 98\%$
- This can serve us to characterize:
  - air transportation under random strikes
  - social contacts even when someone is off
  - destroying of criminal/terror networks
  - **eradication of an epidemics**

Enhanced robustness

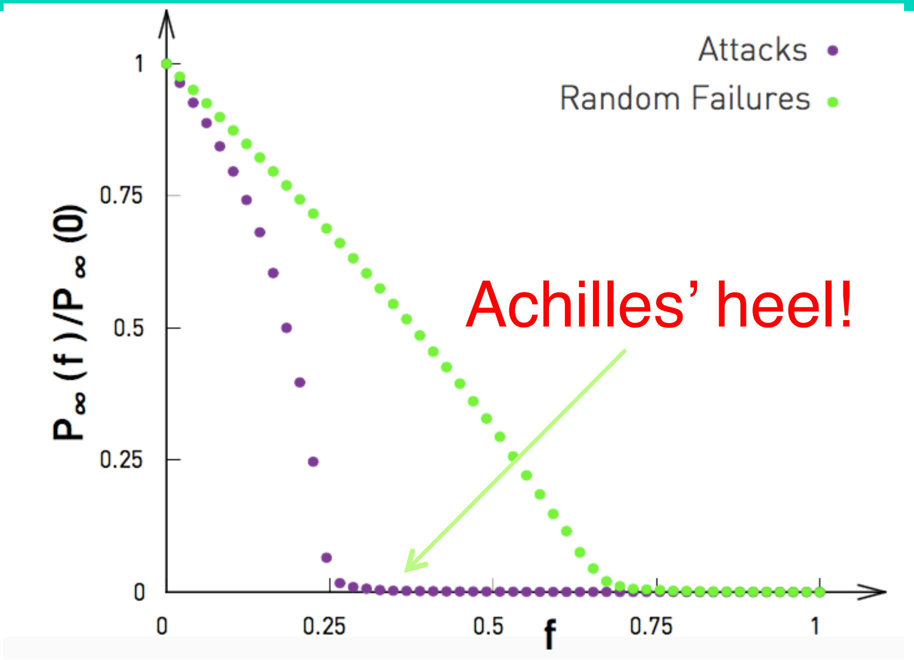
- For a random graph $f_c^{ER} = 1 - 1 / \langle k \rangle$
- A network has **enhanced robustness** if its breakup threshold $f_c > f_c^{ER}$
  - does not need to be scale-free for it (it only needs $\langle k^2 \rangle > \langle k \rangle (1 + \langle k \rangle)$
  - however, scale-free networks surely have it

Robustness against attacks

- What if removals are not by chance, but caused by an adversary with sufficient insight on our network structure?
  - such an adversary may be interested in causing the worst possible damage
  - and it is immediate to see that their worst action would be removing the nodes with highest degrees (the hubs), thereby causing the biggest disruption of service

Robustness against attacks

- An attack meant to cripple a scale-free network should go for the hubs first to create most havoc



- Result: a breakup threshold similar to random failures, but now it is finite and actually has a much smaller value

Robustness against attacks

- Scale-free networks are not very robust to targeted attacks exactly because they have vulnerable hubs
- Recall that: $f_c = 1 - (\langle k^2 \rangle / \langle k \rangle - 1)^{-1}$
  - meaning that robustness depends on $\kappa = \langle k^2 \rangle \,/\, \langle k \rangle$ (the larger the better)
  - removing hubs decreases $\langle k^2 \rangle$, thus making the network more vulnerable (it decreases $\langle k \rangle$ too, but $\langle k^2 \rangle$ decreases faster)

Robustness against attacks

- Take a scale-free network with $p_k = c\, k^{-\gamma}$
  - actually, limited to $k_{min} \leq k \leq k_{max}$
  - and where $c = (\gamma\text{-}1) / (\, k_{max}{}^{1\text{-}\gamma}\, k_{min}{}^{1\text{-}\gamma}\,)$
- A targeted attack removing f% nodes again changes $k \rightarrow k'$ in two ways:
  - it erases some nodes, and now they are all the biggest hubs: $k_{max} \rightarrow k'_{max} \ll k_{max}$
  - and it also removes the links associated to them, thus changing their neighbors' degree

Robustness against attacks

- Focus on the first effect, search the new cutoff k'_{max} through $$f = \int_{k'_{max}}^{k_{max}} p_k \, \mathrm{d}k$$
- We have: $$f = \frac{\cancel{\gamma - 1}}{k_{min}^{-\gamma+1} - k_{max}^{-\gamma+1}} \; \frac{k'^{\,-\gamma+1}_{max} - k_{max}^{-\gamma+1}}{\cancel{\gamma - 1}}$$
  - if $k'_{max} \ll k_{max}$ (true if network large enough) we can also neglect the terms with k_{max}
- So, $f = (k'_{max} / k_{min})^{1-\gamma} \rightarrow \quad k'_{max} = k_{min} \, f^{1/(1-\gamma)}$

Robustness against attacks

- For the second effect, evaluate *g* that is fraction(links) deleted by removing nodes

$$g = \int_{k'_{\max}}^{k_{\max}} k\, p_k \,\mathrm{d}k \Big/ \int_0^{k_{\max}} k\, p_k \,\mathrm{d}k = \frac{c}{\langle k \rangle} \int_{k'_{\max}}^{k_{\max}} k^{1-\gamma} \,\mathrm{d}k$$

$$= \frac{1}{\langle k \rangle} \frac{1-\gamma}{2-\gamma} \frac{k'^{\,-\gamma+2}_{\max} - k^{-\gamma+2}_{\max}}{k^{-\gamma+1}_{\min} - k^{-\gamma+1}_{\max}} \approx \frac{1}{\langle k \rangle} \frac{1-\gamma}{2-\gamma} \frac{k'^{\,-\gamma+2}_{\max}}{k^{-\gamma+1}_{\min}} = \left(\frac{k'_{\max}}{k_{\min}} \right)^{-\gamma+2}$$

neglecting the $k_{\max}$

because $\langle k \rangle = k_{\min} (\gamma-1) / (\gamma-2)$

Robustness against attacks

- We found $g = (k'_{max} / k_{min})^{2-\gamma}$ and we can combine it with $k'_{max} = k_{min} f^{1/(1-\gamma)}$ to obtain

$$g = f^{(2-\gamma)/(1-\gamma)}$$

  - for $\gamma \longrightarrow 2$ all links destroyed even for small f
  - remember the graph → hub-and-spoke
- Now, link removal rate g is random, so:

$$p_{k'} = \sum_{k=k_{min}}^{k'_{max}} p_k \binom{k}{k'} g^{k-k'} (1-g)^{k'} \quad \text{for } k_{min} \le k' \le k'_{max}$$

Robustness against attacks

- Once corrected for new k'_{max} and new $p_{k'}$, finding f_c is the same robustness problem
- Similar to previous analysis, we can derive

$$\kappa = \left|\frac{2-\gamma}{3-\gamma}\right| \begin{cases} k_{min} & \text{for } \gamma > 3 \\ k'^{3-\gamma}_{max} \; k^{\gamma-2}_{min} & \text{for } 3 > \gamma > 2 \end{cases}$$

- →manipulations→ a parametric equality:

$$f_c^{\frac{2-\gamma}{1-\gamma}} = 2 + \frac{2-\gamma}{3-\gamma} k_{min} \left(f_c^{\frac{3-\gamma}{1-\gamma}} - 1 \right)$$

Improving robustness

- How to make the network more robust to both random failures and attacks?
  - both aspects depend on $\kappa = \langle k^2 \rangle / \langle k \rangle$
  - if the number of nodes is fixed but we are allowed to add **redundant** links, we should do so to increase the variance of the degree
- Similar to information theory: we cannot avoid errors but we create alternate ways to hold the network together

Improving robustness

- It can be shown that the best distribution to achieve robustness is bimodal
  - meaning a fraction r of nodes with degree k_{max} and a fraction $(1-r)$ with k_{min}

$$p_k = r\,\delta(k-k_{max}) + (1-r)\,\delta(k-k_{min})$$

- Adding links in intermediate degree nodes is not helpful, better to concentrate them in few nodes to create hubs (deg k_{max})

Improving robustness

- Obviously a bimodal distribution is robust
  - against random removals because of hubs
  - against attacks to < *r* of the hubs
  - even if all hubs are removed, the network can still survive if $k_{\min}$ is large enough, as now $\langle k \rangle = k_{\min} > 1$, so we still have a GC
- Thus, if the goal is max f_c, it can be shown that *r* does not need to be very large

Immunization

- Efforts to stop epidemics ↔ directed towards increasing the infection threshold
  - randomly immunizing a fraction *f* of the nodes is same as decreasing $\langle k \rangle$ to $(1-f)\langle k \rangle$
  - so, lower spreading rate $\alpha=\beta/\mu \rightarrow (1-f)\,\alpha$
- Yet, check vs threshold $\alpha^{(SIR)} > \langle k \rangle / (\langle k^2 \rangle - \langle k \rangle)$
  - e.g., for random networks $\alpha^{(SIR)} = 1/\langle k \rangle$
  - but for scale-free (vanishing threshold) = 0

Immunization

- For example: virus sent as attachment (thus spreading on email network)
  - thus $\langle k \rangle$=3.26; if α=1 and we assume the network is random → we need $f = 0.76$
  - but network is scale free and $\langle k^2 \rangle$=1271; hence, in reality we need $f = 0.997$
- To be fully protected, we would need to install anti-virus on every computer!

Targeted immunization

- For a scale-free network with $\gamma < 3$, vanishing threshold due to big $\kappa = \langle k^2 \rangle / \langle k \rangle$
  - implying network robustness vs attacks
  - yet now network=infection, attacks=vaccine!
- Solution to decrease $\langle k^2 \rangle$: target the hubs
  - better immunize the super-spreaders!
- **Possible strategy**: just vaccine all nodes with degree k higher than k_0

Targeted immunization

- As for network robustness, this implies:
  - the maximum degree goes from k_{max} to k_0
  - we remove f % nodes and g% links:

    $f = (k_0 / k_{min})^{1\text{-}\gamma}$ $\qquad$ $g = (k_0 / k_{min})^{2\text{-}\gamma}$
  - and degree distribution becomes

$$p_{k'} = \sum_{k=k_{min}}^{k_0} p_k \binom{k}{k'} g^{k-k'} (1-g)^{k'}$$

  - resulting in $\langle k' \rangle = (1\text{-}g) \langle k \rangle$,

    $\langle k'^{\,2} \rangle = (1\text{-}g)^2 \langle k^2 \rangle + g(1\text{-}g)\langle k \rangle$

Targeted immunization

- So $\alpha_C^{(SIS)} = \dfrac{(1-g)\langle k\rangle}{(1-g)^2\left\langle k^2\right\rangle + g(1-g)\langle k\rangle} = \dfrac{1}{(1-g)\kappa + g}$

  where we recall $\kappa = \dfrac{\gamma-2}{3-\gamma} k_0^{3-\gamma}\ k_{\min}^{\gamma-2}$

- This implies $1/\alpha_C^{(SIS)}$ (resp. $1/\alpha_C^{(SIR)}$) is

  $$\frac{\gamma-2}{3-\gamma} k_0^{3-\gamma}\ k_{\min}^{\gamma-2} - \frac{\gamma-2}{3-\gamma} k_0^{5-2\gamma}\ k_{\min}^{2\gamma-4} + k_0^{2-\gamma}\ k_{\min}^{\gamma-2} - 1$$

  - computations for SIR are analogous

- If $k_0 \gg k_{\min}$ (sensible) then $\alpha_C^{(SIS)} \approx \alpha_C^{(SIR)} \approx 1/\kappa$

Problem with finding the hubs

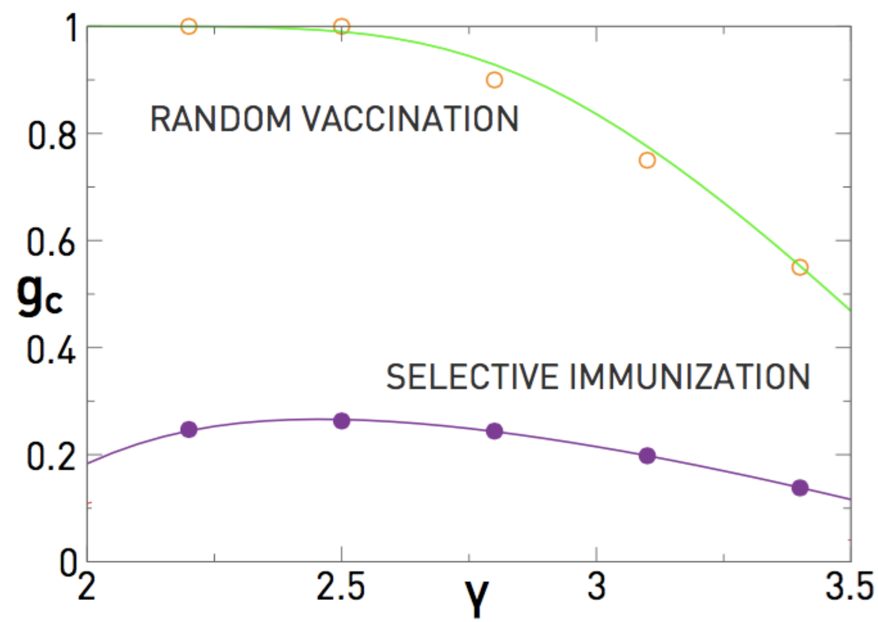
- Generally, hubs are not easy to identify
  - for a sexual network, we need #partners
  - for online social networks, #friends is easy but most contacts are fake (to show off)
  - for influenza, hard to detect in advance who are the super-spreaders
- We avoid the hassle of finding the hubs by relying on the **friendship paradox**

Problem with finding the hubs

- A possible "smart" strategy:
  - start from Group 0 (n_g random nodes)
  - then choose a neighbor for each node in Group 0 → we obtain a n_g–sized Group 1
  - immunize Group 1
- This strategy works because on average Group 1 nodes have degree > Group 0
  - in practice: ask some individuals to name a recent contact/acquaintance/partner

Immunization performance

- Scale-free networks with variable γ
  - g_c=req. %vaccination
- Random vaccination has poor results: we need very high g_c



- Selective immunization instead has g_c always below 30% and ≈ insensitive to γ

Travel restrictions

- Another possible control technique for epidemics is to put travel restrictions
  - serious economic implications!
- Generally, this is hard to incorporate in the analysis as the epidemic itself is already causing self-imposed travel limitations!
  - an epidemic at its peak can cause up to 40% of travel reduction