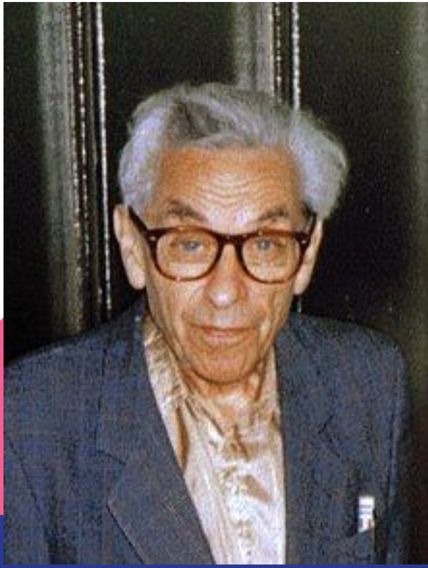
Large Random Graphs and Epidemics (Not fair game on tests, but very useful in the real world)

By Grayson York

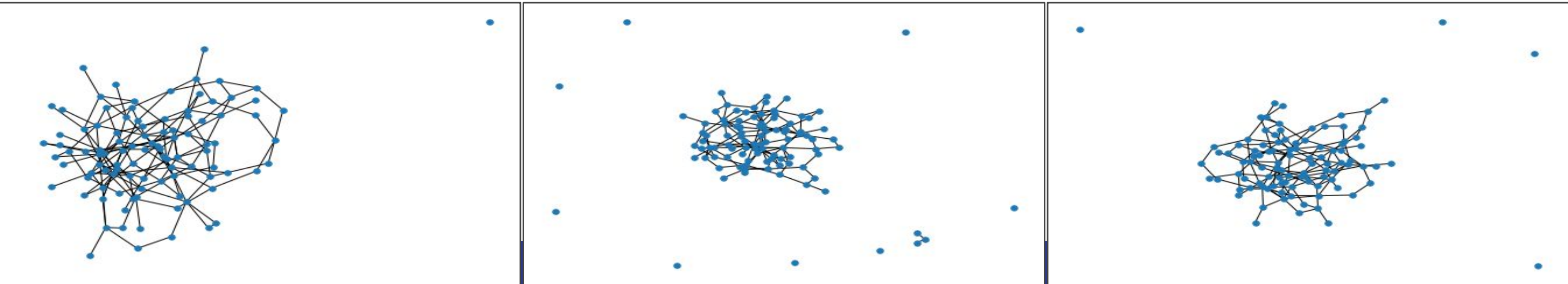
Paul Erdős

- Second most prolific mathematical publisher of all time behind Euler
- One of the stranger characters in modern math
- Famous for proving results using the “probabilistic method”
- Spent most of his life travelling to different universities living out of a suitcase
- Addicted to amphetamines (though he claimed he could quit)
- Reportedly did not know how to open a carton of orange juice
- Ron Graham handled everything except math for him
- Mathematicians now talk about their “Erdos number”
    - Those who publish with Erdos get a number of 1
    - Those who publish with Erdos number n get Erdos number n+1
    - John Reif’s Erdos number is 2 (Through Bob Tarjan)



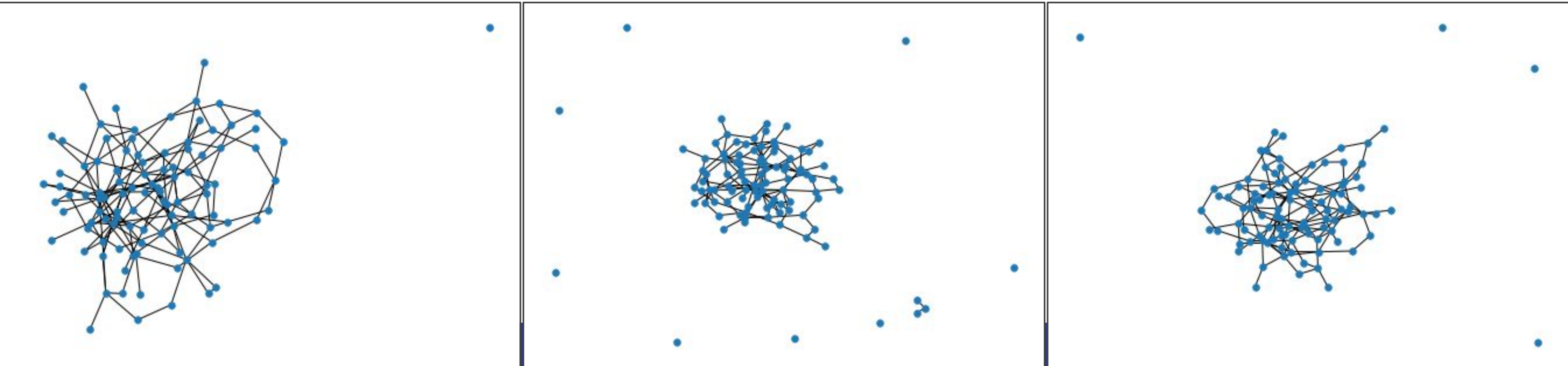
ER Random Graphs

- An Erdos Renyi Random Graph has 2 parameters n, and p
- The graph will have n vertices, and each edge will occur with probability p
- The total number of possible edges in the graph is n(n-1)/2
- The expected number of edges in the graph is pn(n-1)/2
- The average degree of a vertex is pn
- Note that as n grows large with a fixed p, the degree of each vertex goes to infinity



A Useful Reparameterization

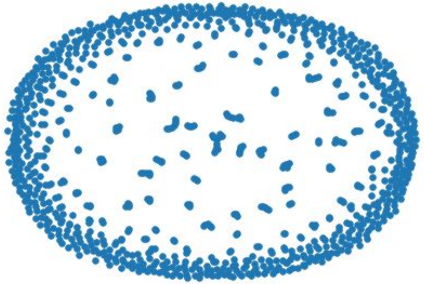
- What if we did not think in terms of p
- Instead, let us define μ = np
- Now, we can hold μ constant, and let n grow large and p will adapt
- These are ER graphs with n = 100, μ = 3
- Note that almost all of the vertices are in the same component



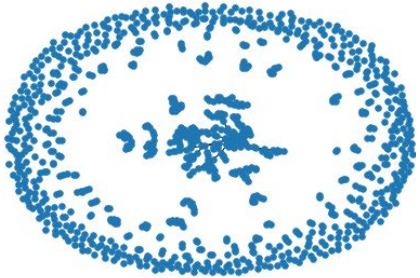
Giant Component (GC) Formation

- Under what circumstances will this giant component form?
- Lets make an informal assumption, lets assume that no cycles exist.
- Now, we can think about this like a tree.
- Starting from a vertex v, how many neighbors does v have?
    - In expectation, v has μ neighbors
    - In expectation, each of v's neighbors has μ neighbors, so v has μ^2 second order neighbors
    - Continuing this reasoning, μ has μ^k neighbors of order k
- When will μ^k grow to cover a large chunk of the graph as k grows large?
- This will happen if and only if $\mu > 1$!

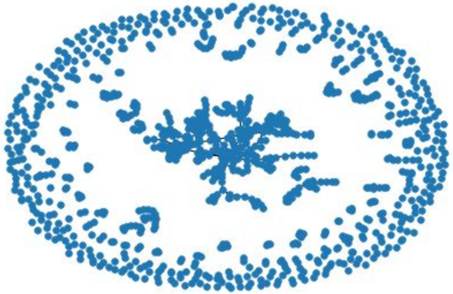
Some Examples With n = 1000



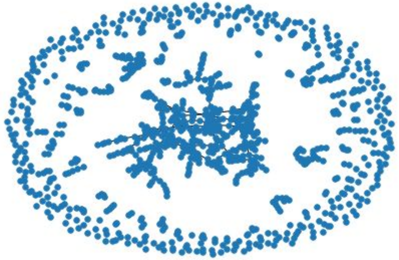
μ = .5



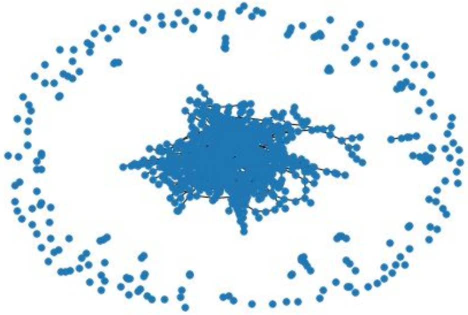
μ = .9



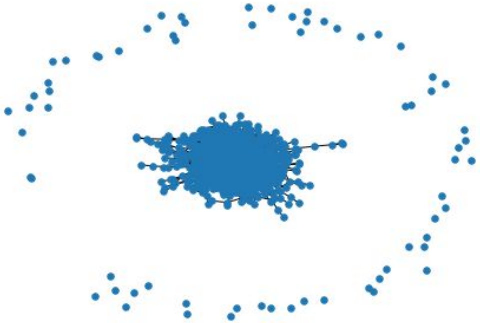
μ = 1



μ = 1.2



μ = 2



μ = 3

Can Multiple GC's Form?

- What if we had 2 GC's, each with about n/2 nodes?
- There are (n/2) vertices in the first GC, and (n/2) vertices in the second GC, so there are $(n/2)^2 = n^2/4$ edge possibilities to cross this gap.
- The expected number of edges crossing the gap is $pn^2/4 = (pn)n/4 = \mu n/4$.
- If $\mu > 1$ (necessary for the GC to form), and n is large, we have many edges crossing this gap
- Only one GC can exist!

Your Friends Have More Friends Than You Do

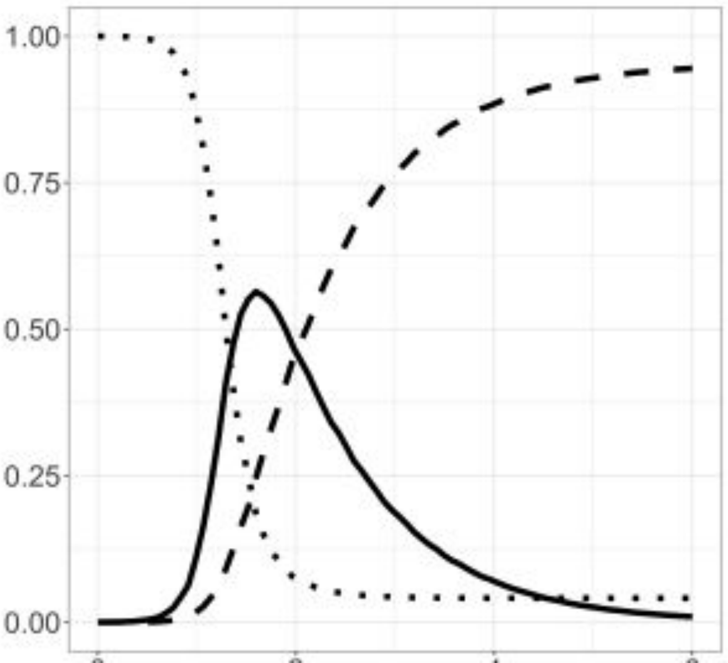
- Take a random node v, it has μ neighbors in expectation
- Now let us examine the neighbors of v.
- Each of neighbor of v, call it v', certainly has an edge to v by assumption
- In addition, v will have an edge to each other node in the graph with probability p
- We see that there are n-1 vertices other than v, so v' has $p(n-1) \approx pn = \mu$ neighbors among these
- Therefore, in expectation v' has μ+1 neighbors!
- The neighbors of v have more neighbors than v in expectation

Epidemic Modeling Using The Contact Process

- Say we have a population of susceptible nodes S
- Say we have a population of infected nodes I
- Say we have a population of removed nodes R (Recovered or dead in practice)
- Say we have an I node, it will move to R in any time step with probability r
- Say we have an I node, it will infect each of its S neighbors with probability i
- Informally, we will assume that r and i are both large (or the timesteps are very fine), so we only care about the ratio i/(i+r) = λ, which we will call the infection rate
- Start out with all susceptible and 1 infected, and observe

What Does This Look Like In Practice?

- In this figure, the population fractions of S,I,R are shown
- S is dotted, R is dashed, I is solid
- S will drop over time with infections
- R will increase over time as infections recover
- I will initially grow rapidly
- When few susceptibles remain, I will drop
- Eventually all infected will be removed



When Will an Epidemic Occur?

- In order to trigger the rapid growth phase, rapid infections must occur in the fully susceptible population
- This means λ must be quite large
- If λ is small, the infection will likely not get much farther than patient 0
- How can we calculate the critical λ, which we will call λ_c?
- Note that if $\lambda > \lambda_c$, we will not always have an epidemic, but an epidemic is possible
- It is always possible that patient 0 will be removed too soon, and the infection will fail

How Can We Calculate λ_c?

- This is similar to the idea of a giant component
- If the infection is spreading from a point, it will start out like a tree
- It will grow on each step if each infected node infects at least one new susceptible node
- A node on average has μ susceptible neighbors (recall the friends have more friends result)
- Therefore, we will have $\lambda\mu$ infections at the next level
- In order for the infection to survive, we must have $\lambda\mu>1$
- Therefore, $\lambda_c\mu = 1$, and $\lambda_c = 1/\mu$

Can We Calculate λ_c on Other Graph Topologies?

- We will assume these graphs have n vertices as n grows very large
- What about a path graph?
  - Here, each vertex will have 2 neighbors.
  - If any vertex does not pass on the infection, the infection will die.
  - Therefore, every vertex must pass on the infection, so λ_c=1
- What about a star graph?
  - Here, one vertex will be neighbors with every other vertex
  - If the central vertex is infected, then a large fraction of the leaves will be infected
  - If λ>0, this will occur with positive probability
  - Therefore λ_c=0

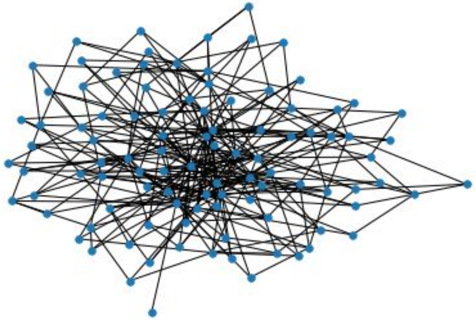
Can We Find λ_c In General?

- Not at full generality, but for a large class of networks, we have a nice tool
- $\lambda_c = \langle k \rangle / \langle k^2 \rangle$ where $\langle k \rangle$ is the average degree of the network and $\langle k^2 \rangle$ is the average squared degree
- This means that a graph with a low epidemic threshold can either have a high mean degree (because $\langle k^2 \rangle$ grows faster than $\langle k \rangle$) or a high variance in degree (because $\langle k^2 \rangle$ will grow, but $\langle k \rangle$ will stay the same)
- Note that this is why superspreaders are so concerning, they significantly increase $\langle k^2 \rangle$ in a social network
- This is related to the friends have more friends result, one infection to a well connected friend can cause infections to explode

Quick Reminder About Barabasi Albert Networks

- Nodes are iteratively attached to nodes already present in the network
- This leads to a much less uniform degree distribution than ER graphs
- Because the variance on degree is so high, in large BA networks, λ_c=0
- This means that a disease will always cause an epidemic on a large BA network, no matter how small the infection rate
- Unfortunately, many believe that the BA degree distribution is a good model for our society

Barabasi Albert network with 100 nodes and 3 connections per node



SIS Epidemics

- What if people can get infected with the disease multiple times?
- We can delete the R case, and have all of the infecteds transfer back to S when they recover
- We will still examine λ, but the questions we ask will be slightly different
- Now we will ask whether the infection can live forever, and what that survival will look like

Some Strange Behavior on Trees

- Let us say that we have a tree where every vertex has degree μ
- If the infection spreads, it is much more likely for it to spread to one of the μ-1 child branches than the 1 parent branch
- That means that it is much easier for an infection to spread downstream than upstream
- As a result, we end up with 2 critical λ values. One of which indicates an infection strong enough to survive (by spreading downstream) and one which indicates an infection strong enough to infect the whole tree (by spreading upstream)

Endemic Infections on BA graphs

- In the real world we see infections which do not infect a large percentage of the population at a given time, but manage to live forever (think computer viruses, the common cold)
- This is very unlikely on an ER graph, as there is a very narrow window above λ_c where a small infection can survive, and its survival is very tenuous in that window
- In a BA graph, because λ_c=0, an infection with a small λ value can still live forever (This is currently a hot area, the paper where this result was discovered was published in 2015)

Key Takeaways

- Large random graphs like to form one large component
- Your friends probably have more friends than you, but that is true for most people
- Network processes often have “critical values” where behavior changes drastically and rapidly when a threshold is crossed
- During a global pandemic, we want to decrease both the average degree and variance of degrees in order to stop the spread (so avoid superspreaders)
- If you want to get into a hot area of applied math with a lot of exciting results coming out, check out random graphs!