Random graphs

Erdos-Renyi random graph

Giant Component

SYSM 6302

CLASS 9

Random Graphs

→ networks in which certain parameters are fixed & the rest is random

#nodes, #edges, degree of nodes, etc.

→ Random network models are developed to study certain characteristics of real world networks in isolation

what effect does a power law distribution have on robustness of connectivity under edge failure?

→ Also to determine what are the generative processes that create networks of certain types

→ random network models create families of graphs that are similar

THE Random Graph: Erdos-Renyi Random Graph

#nodes

probability to connect any pair of nodes $p \in [0,1]$

$$G(n,p)$$

1. Add n nodes
2. For each (unique) pair i, j flip a biased coin:

$i = 1, 2, \dots, n$

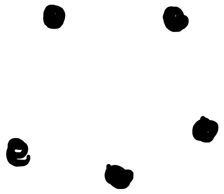
$j = i+1, \dots, n$

Heads has probability p $\Rightarrow$ create an edge between i & j

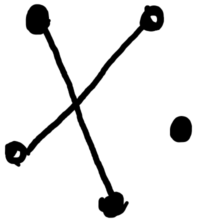
Tails has probability $1-p$ $\Rightarrow$ do not create an edge

$\rightarrow$ We will get a different graph each time, so $G(n,p)$ is a family of graphs

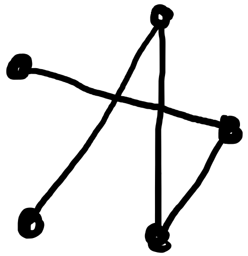
These are all possible ER graphs for $n=5$ and $p=0.2$



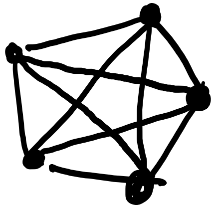
$|E|=0$



$|E|=2$



$|E|=4$



$|E|=10$

# of edges possible (without self-loops) $= \binom{n}{2} = \dfrac{n!}{2!\,(n-2)!} = \dfrac{n(n-1)(n-2)!}{2\,(n-2)!} = \dfrac{n(n-1)}{2} = 10$

→ What is the probability that $G(n,p)$ generated these graphs?

The different instances of $G(n,p)$ appear with different frequencies

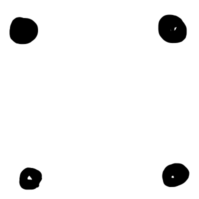
↳ so it is a distribution of graphs, not just a family of graphs

$$M = \#\text{ of possible edges} = \binom{n}{2} = \frac{n(n-1)}{2}$$

$$\underbrace{P(G)}_{\substack{\text{Probability of generating}\\ \text{a specific graph instance}\\ \text{with } m \text{ edges}}} = \underbrace{p^m}_{m \text{ "heads"}} \underbrace{(1-p)^{M-m}}_{M-m \text{ "tails"}}$$

getting an exact sequence: HTTH

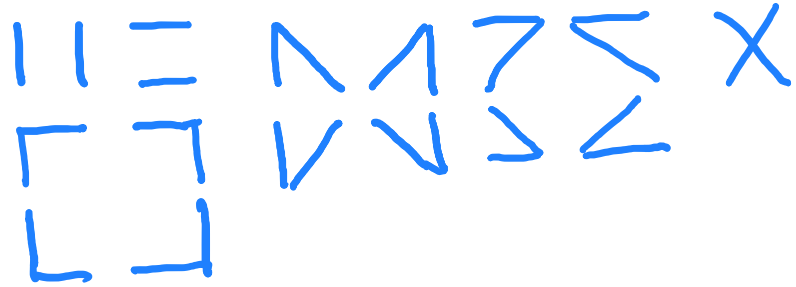
There is more than one way to generate a graph with m edges:



$M = \frac{4(3)}{2} = 6$ possible edges

HHTT
HTHT
HTTH
THHT
....

# of ways to have 2 edges:



$15 = \binom{6}{2}$

Binomial Distribution

$$P(m) = \binom{M}{m} p^m (1-p)^{M-m}$$

probability of generating a graph with m edges

# of ways of chosing which m edges out of M possible edges

probability of generating a specific graph with m edges

Binomial Distribution

Distribution that captures the number of "successes" in a sequence of independent trials.

two options:
True/False
Heads/Tails

the outcome of one success/failure doesn't impact another

→ models the number of heads from flipping a coin:

Flip a coin 3 times: $\{HHH, HHT, HTH, \underline{HTT}, THH, \underline{THT}, \underline{TTH}, TTT\}$

What is the probability of getting (exactly) 2 heads? $\frac{3}{8}$

$(n=3)$
$(m=2)$

$$P(2) = \binom{\frac{3(2)}{2}}{2}\left(\frac{1}{2}\right)^2\left(1-\frac{1}{2}\right)^{\frac{3(2)}{2}-2} = \binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1 = 3\cdot\frac{1}{4}\cdot\frac{1}{2} = \frac{3}{8}$$

Computing Statistics

$$\underbrace{\langle a \rangle}_{\text{expected value of this statistic}} = \underbrace{\sum_G}_{\text{sum over instances}} \underbrace{a(G)}_{\text{value of statistic for this instance}} \underbrace{P(G)}_{\text{probability of this instance}} = \sum_{m=0}^{M} a(m) P(m)$$

Mean (expected # of edges):

$$\langle m \rangle = \sum_{m=0}^{M} m P(m) = Mp$$

mean of Binomial distribution

not surprising: # of possible edges · probability of keeping them.

Expected average degree:

$$\langle c \rangle = \sum_{m=0}^{M} \frac{2m}{n} P(m) = \frac{2}{n} \cdot Mp = \frac{2}{n} \frac{n(n-1)}{2} p = \underbrace{(n-1)}_{\text{\# of possible connections each node can have.}} p$$

Degree Distribution

Probability of having degree k: $p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}$ ← Same structure as for the whole graph. Now locally around node, the max edges is $n-1$.

→ Value of network models → As $n \to \infty$ the distribution becomes more peaked and more narrow. → the expected values of network statistics are increasingly representative of all network instances.

→ Most real networks are sparse → average degree remains constant as n increases.
Most real networks are sparse → I will have the same number of friends if the world has 5B or 10B people.

↳ i.e., I will have the same number of friends if the world has 5B or 10B people.

$\Rightarrow c = (n-1)p \quad \Rightarrow \quad p = \frac{c}{n-1} \to 0$ as $n \to \infty$

③ $$\ln\left[(1-p)^{n-1-k}\right] = (n-1-k)\ln\left(1-\frac{c}{n-1}\right) \approx (n-1-k)\left[-\frac{c}{n-1}\right] \approx -c \Rightarrow (1-p)^{n-1-k} \approx e^{-c}$$

small

$$\ln(x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}(x-1)^k, \quad -1 < x \leq 1$$

① $$\binom{n-1}{k} = \frac{(n-1)!}{(n-1-k)!\,k!} = \frac{(n-1)(n-2)\cdots(n-k)}{k!} \approx \frac{(n-1)^k}{k!}$$

Could have chosen n^k or $(n-2)^k$, but this makes putting it together easier.

$$\Rightarrow p_k \approx \frac{(n-1)^k}{k!} p^k e^{-c} = \frac{(n-1)^k}{k!}\left(\frac{c}{n-1}\right)^k e^{-c} = e^{-c}\frac{c^k}{k!}$$

Poisson Distribution (i.e., not scale free!)

Poisson Distribution captures the number of events that occur in a fixed interval of observation, given an average arrival rate

Expected clustering coefficient

Suppose a node has degree $k \Rightarrow$ has k neighbors

↳ # of expected neighbor pairs is then: $\binom{k}{2} p$

number of neighbor pairs

$$C_i = \frac{\#\text{ connected neighbor pairs}}{\#\text{ neighbor pairs}} = \frac{\binom{k}{2} p}{\binom{k}{2}} = p$$

for sparse networks

$p = \frac{c}{n-1} \to 0$ as $n \to \infty$

Expected Diameter & Expected Average path length

$$d, \ell \sim \frac{\ln(n)}{\ln(c)}$$

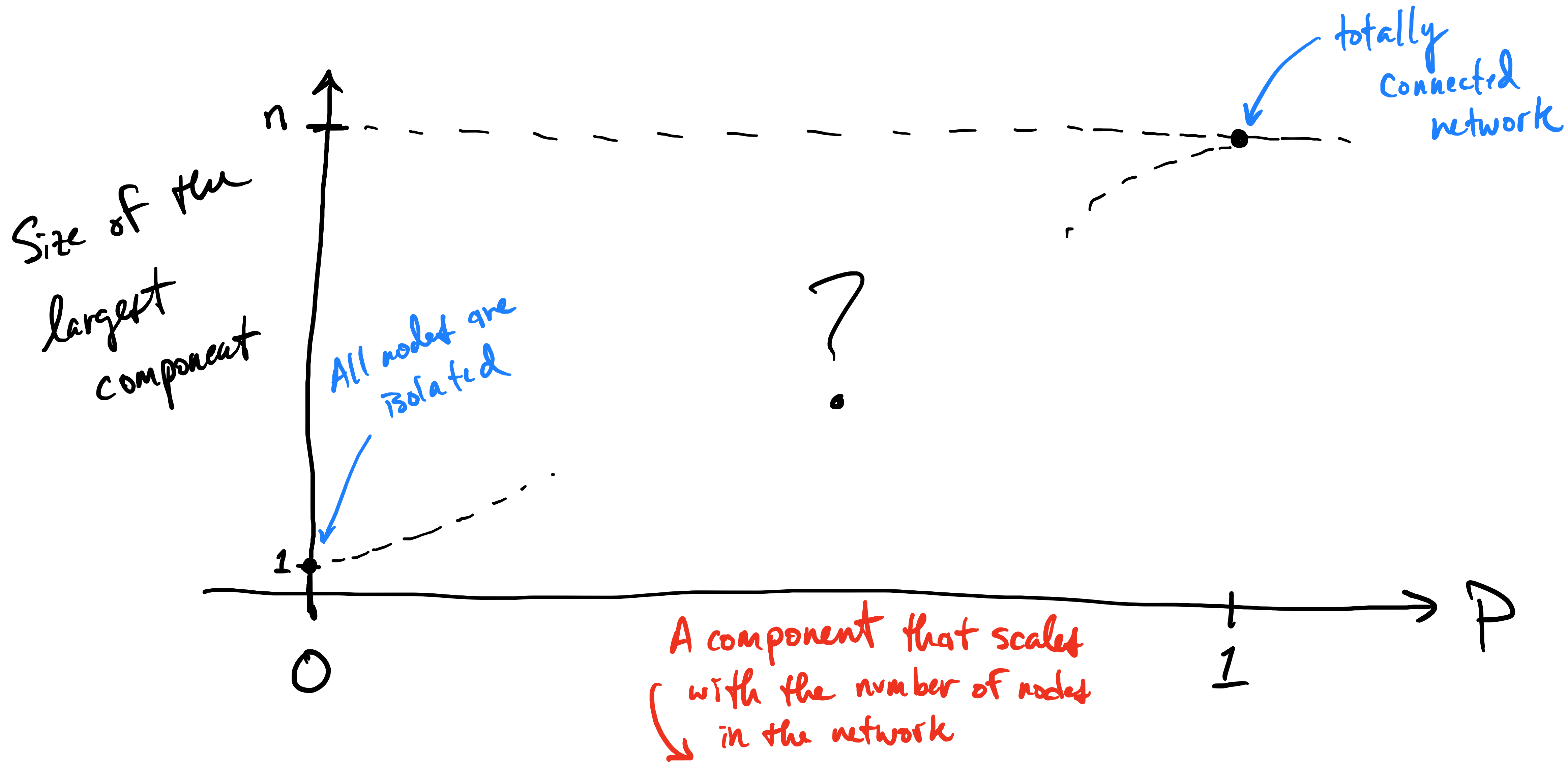
Erdos-Renyi: (compared to real networks)

✗ virtually no clustering

✗ poisson degree distribution

✓ short paths (small world)

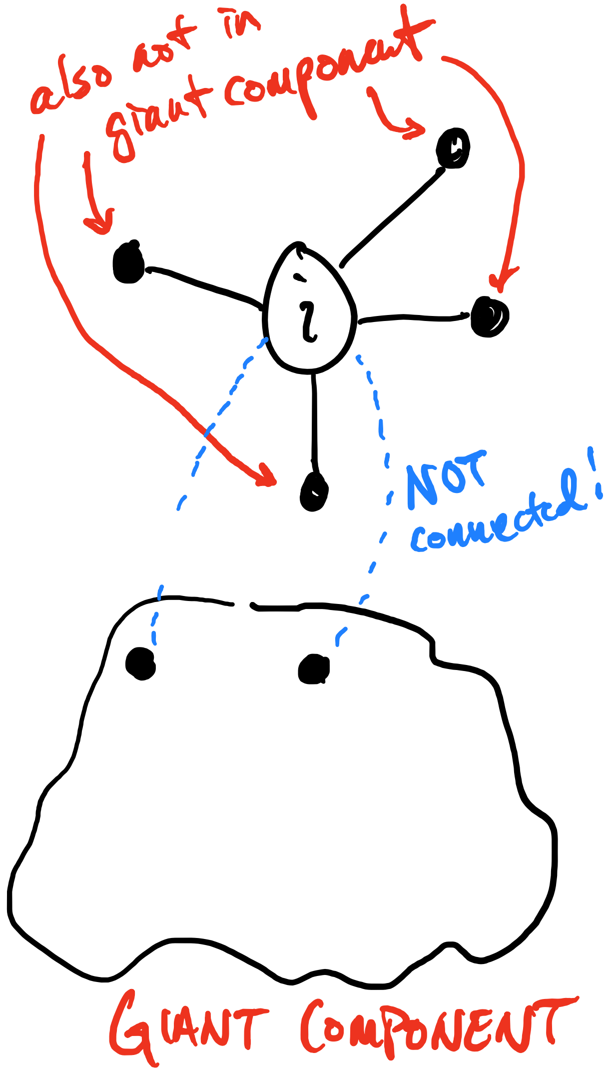
→ Easy to make a directed version: $n(n-1)$ edges instead of $\frac{n(n-1)}{2}$



How does the Giant Component emerge?

Probability that node i is not part of the giant component: u

(= fraction of nodes not in the giant component)



for a given node j: $\underbrace{1-p}_{\text{probability } i \,\&\, j \text{ are not connected}} + \underbrace{p}_{\text{probability } i \,\&\, j \text{ are connected}}\underbrace{u}_{\text{probability } j \text{ is not in giant component}}$

Over all nodes: $$u = (1-p+pu)^{n-1} = \left[1-\frac{c}{n-1}(1-u)\right]^{n-1}$$

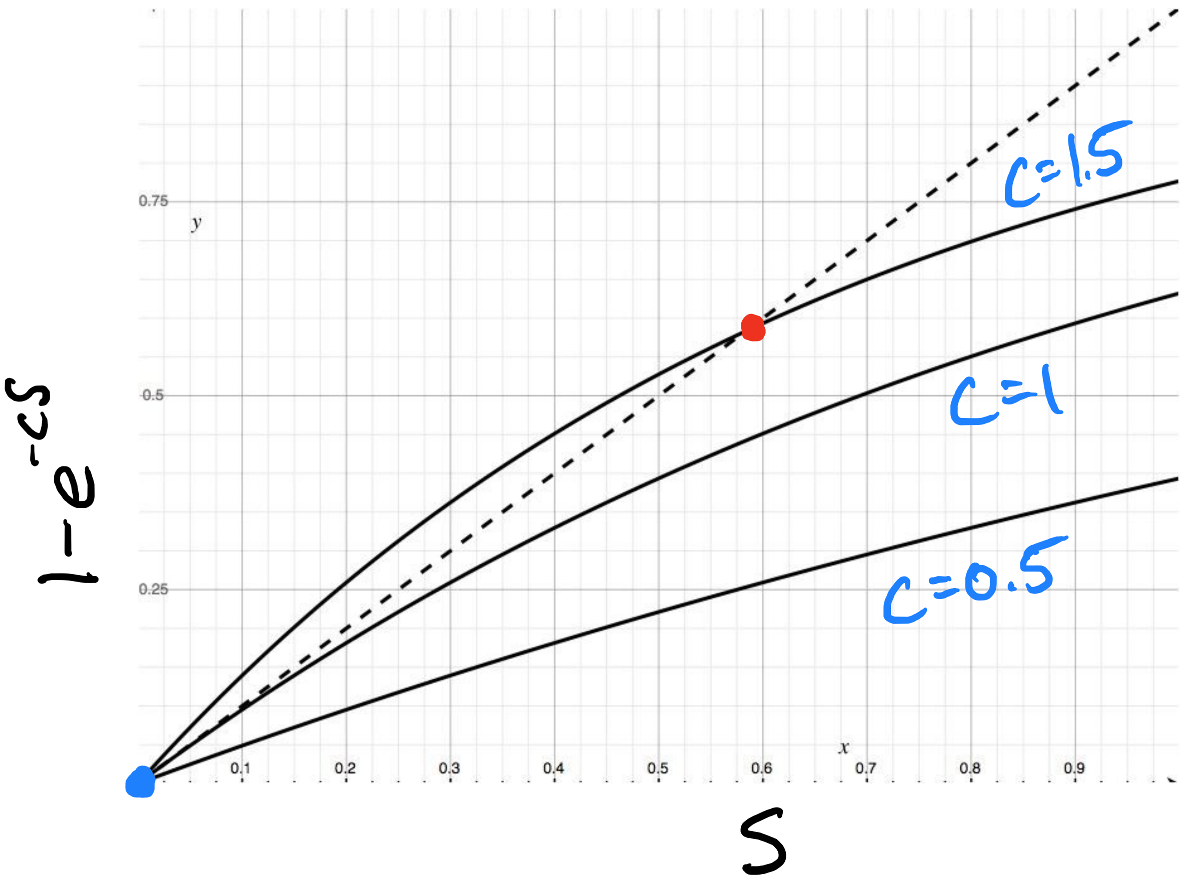
$$\Rightarrow \ln u = (n-1)\ln\left[1-\frac{c}{n-1}(1-u)\right] \approx n-1\left[-\frac{c}{n-1}(1-u)\right] = -c(1-u)$$

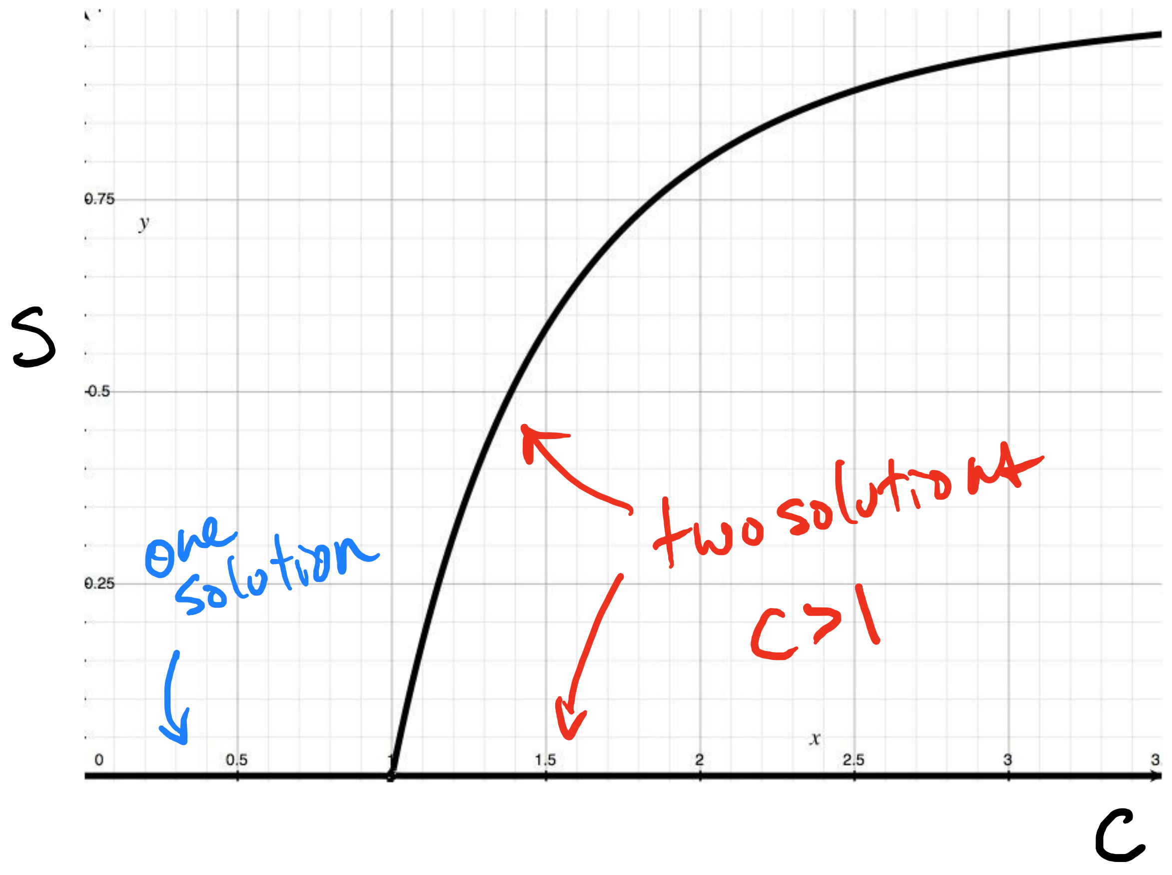
$$\ln u = -c(1-u) \quad \Rightarrow \quad u = e^{-c(1-u)}$$

$S = 1 - u$ the fraction of nodes <u>in</u> the giant component

$$S = 1 - e^{-cS}$$

$$S = 1 - e^{-cS}$$





The giant component exists when $c > 1$