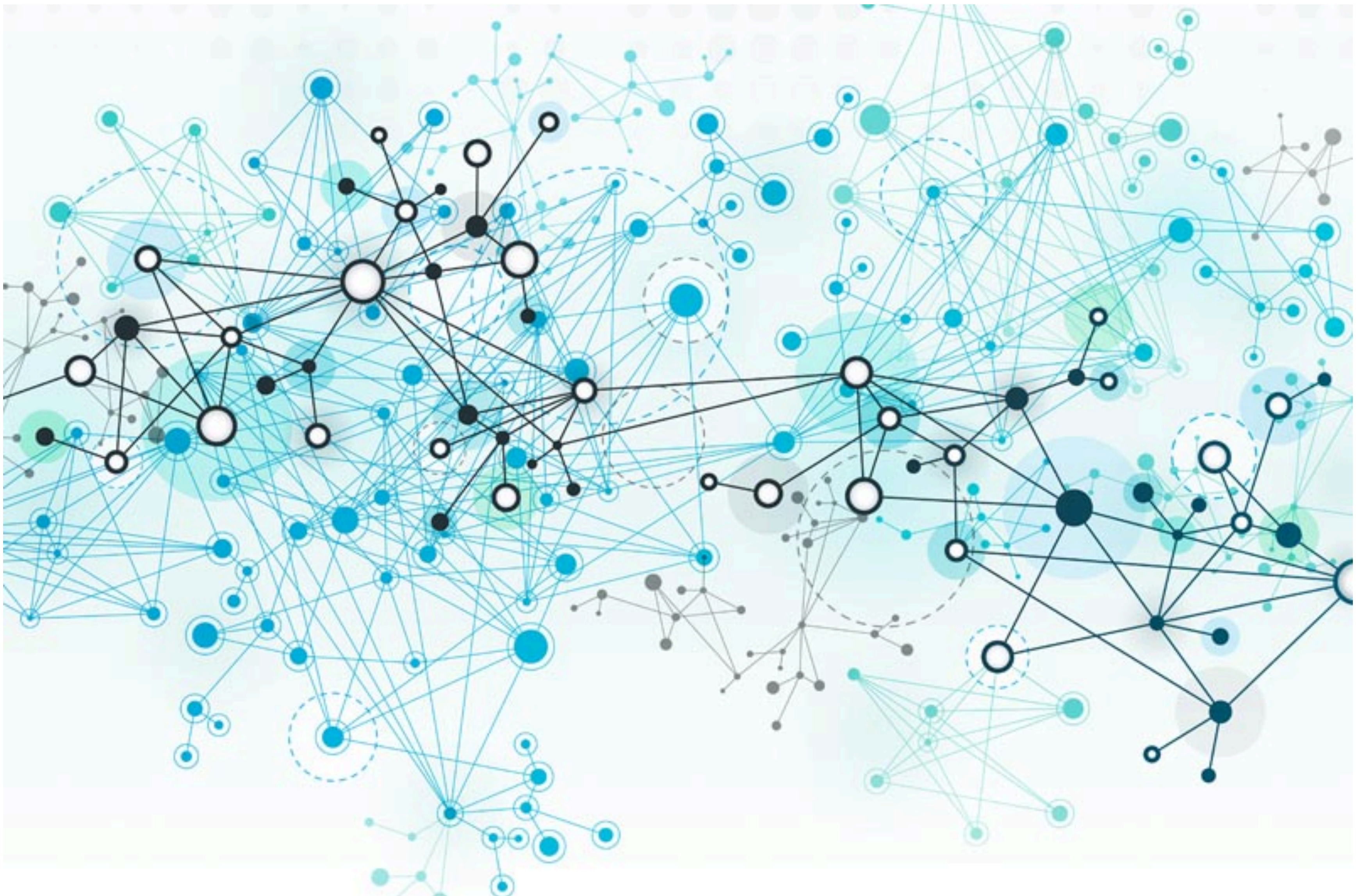
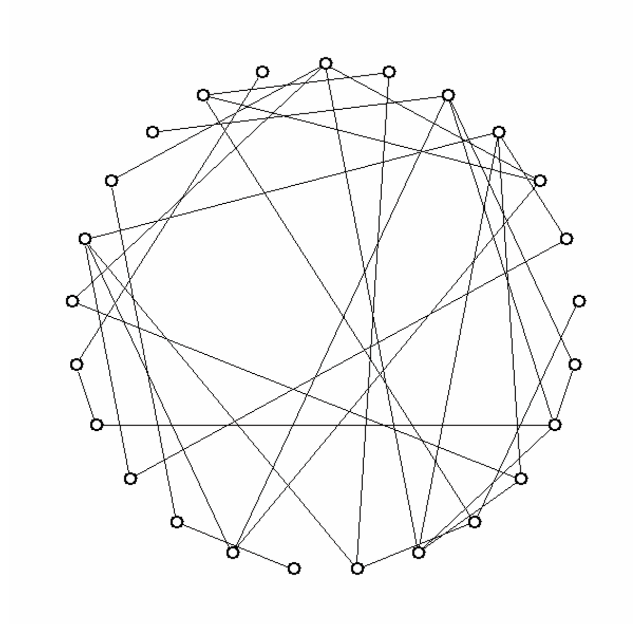
BIG NETWORK DATA
ECE 289 UC San Diego



Random Graphs

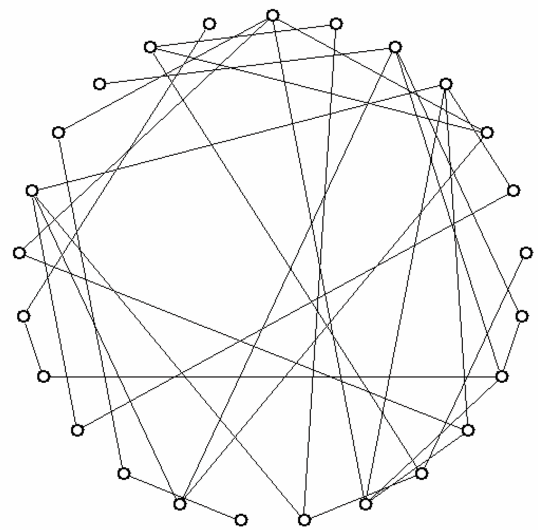
Erdos and Rényi (1959)



Threshold function for connectivity

For values of p(n) below the threshold all components are small, for large values there is a “giant” component

Giant component



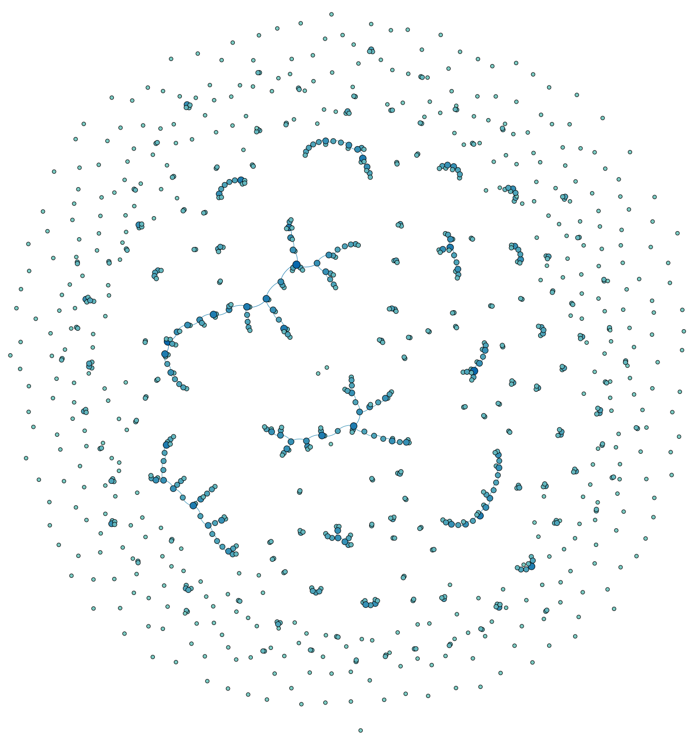
$p = c/n$

C_n largest connected component in G_n

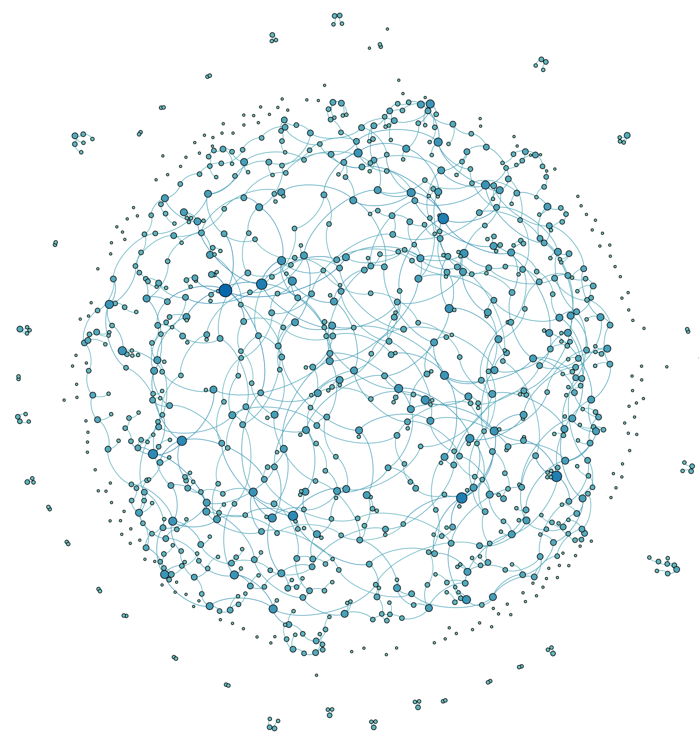
$$c < 1 \implies \exists \alpha(c) < \infty : \lim_{n \to \infty} P(|C_n| \leq \alpha \log n) = 1$$

$$c > 1 \implies \exists \alpha(c) > 0 : \lim_{n \to \infty} P(|C_n| \geq \alpha n) = 1$$

Giant component



$c < 1$



$c > 1$

Expected node degree

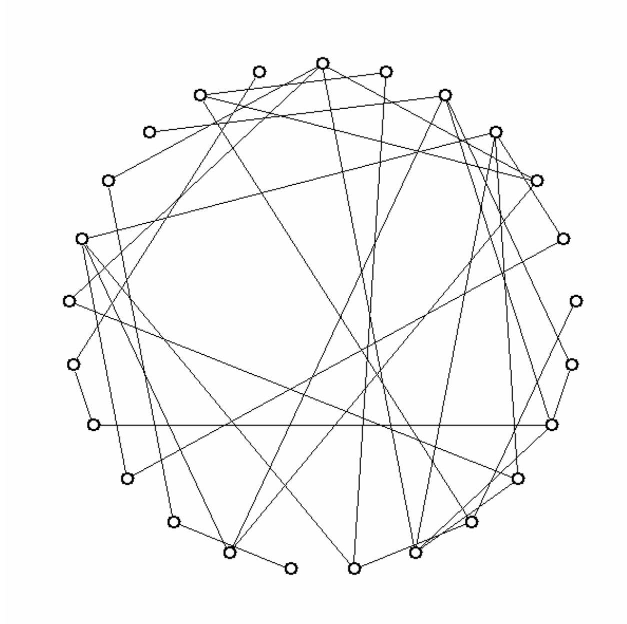
In terms of expected node degree, the threshold for a giant component corresponds to an expected node degree of one.

As the expected node degree is increased further, the fraction alpha of nodes in the giant component increases.

To achieve full connectivity (a single component), corresponding to alpha =1, the expected node degree must increase with *n*

At what rate?

Full connectivity



$$\forall \epsilon > 0 \text{ if } p > \frac{(1+\epsilon)\log n}{n} \implies \lim_{n\to\infty} P(G_n \text{ is fully connected}) = 1$$

$$\forall \epsilon > 0 \text{ if } p < \frac{(1-\epsilon)\log n}{n} \implies \lim_{n\to\infty} P(G_n \text{ is fully connected}) = 0$$

Expected node degree

In terms of expected node degree, the threshold for full connectivity corresponds to an expected node degree of log n.

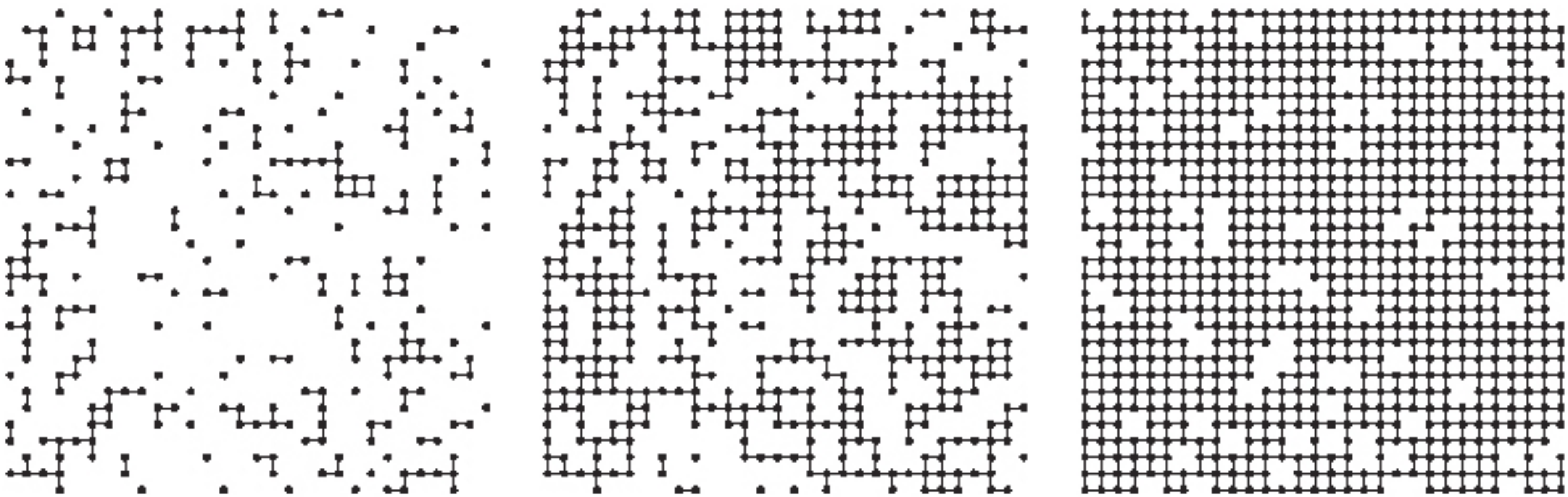
While a constant number of neighbors is enough to form a giant component, a logarithmic number achieves full connectivity

Random graphs are not a very good model of real (complex) networks, as they exhibit poor clustering and lack community structure

How about connectivity in highly clustered random networks?

Random Grid

Broadbent and Hammersley (1957)

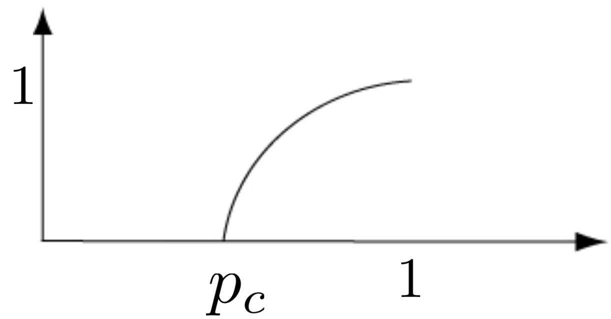


(a) (b) (c)

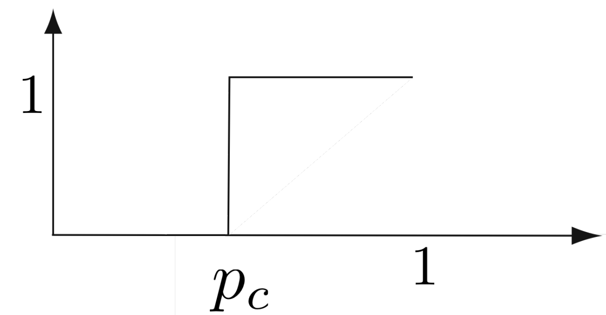
Percolation theory

Random grid on $\mathbb{Z}^2$

$\theta(p) = P_p(|C_0| = \infty)$



$\phi(p) = P_p(E_\infty)$

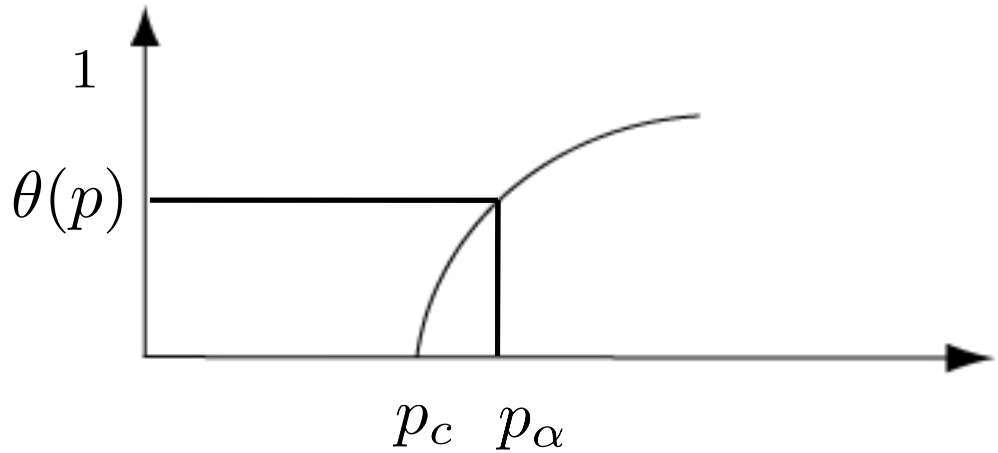


$\exists p_c \in [\frac{1}{3}, \frac{2}{3}]$ (Broadbent and Hammersley, 1957)

$p_c \geq 1/2$ (Harris, 1960)

$p_c = 1/2$ (Kesten, 1980)

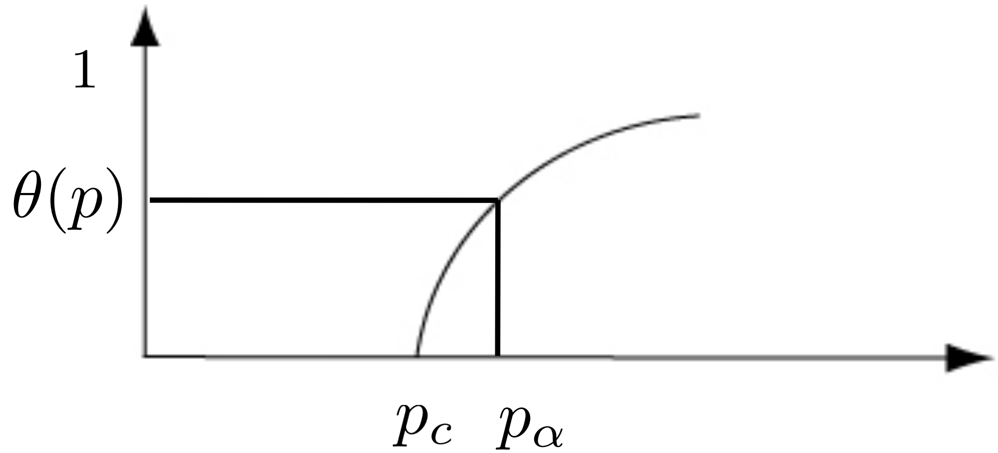
Giant component



The percolation function also provides the fraction of nodes that are connected with high probability in a grid of *n* nodes

$$p_\alpha = \inf\{p : \theta(p) < \alpha\}$$

Full connectivity



To achieve full connectivity we need $p \to 1$

At what rate?

Full connectivity

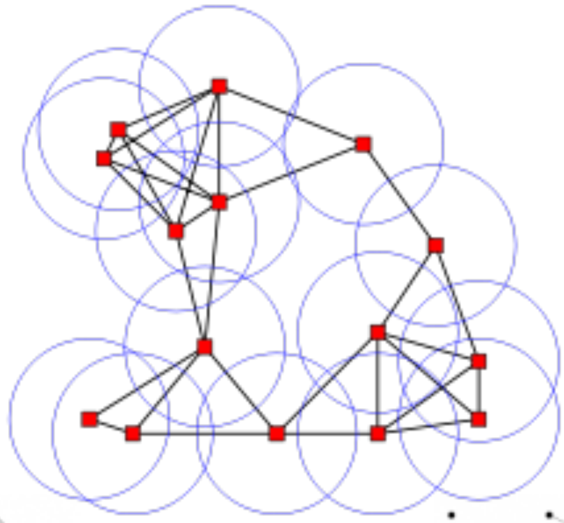
$$p_n = 1 - \frac{c_n}{n^{1/4}}$$

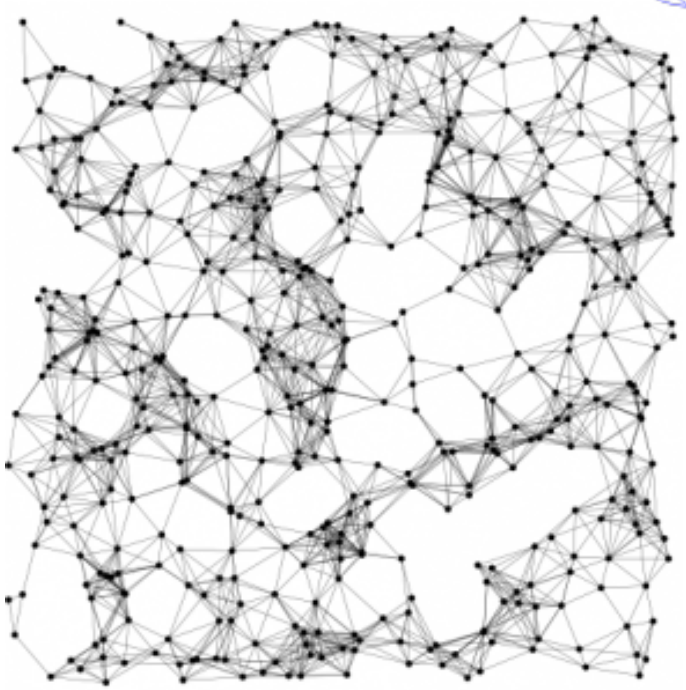
$$\lim_{n\to\infty} P(G_n \text{ is fully connected}) = 1 \iff c_n \to 0$$

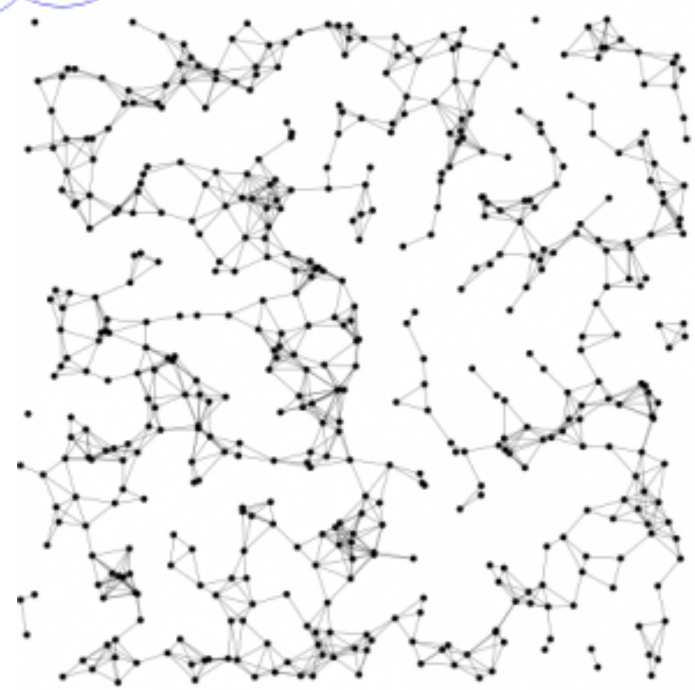
To achieve full connectivity the edge probability must tend to one at a rate that is slightly higher than the square-root of the side length of the box

Random geometric graphs

Gilbert (1961)

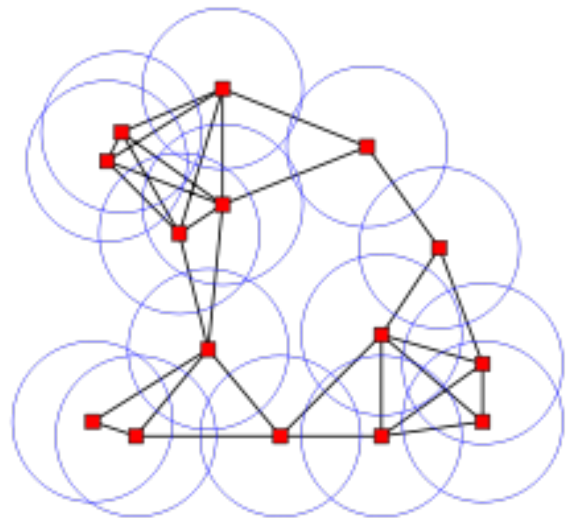






Random geometric graphs

Gilbert (1961)



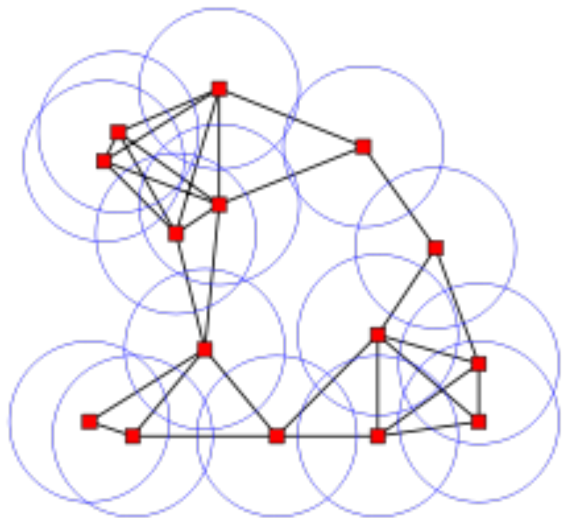
Random geometric graphs have high clustering (like the random grid)

The node degree is not uniformly bounded (like the random graph)

They provide a slightly more refined model for real complex networks

Random geometric graphs

Gilbert (1961)



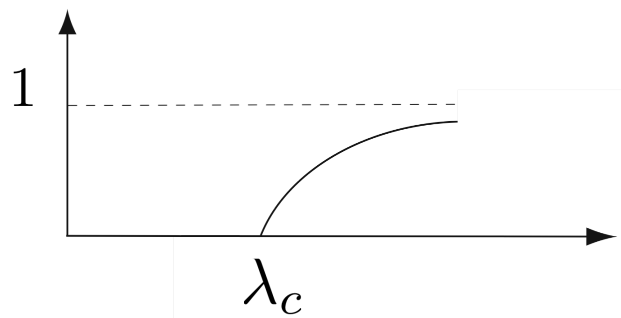
Points are uniformly distributed on the plane according to a Poisson point process of density λ

Points within unit range share an edge

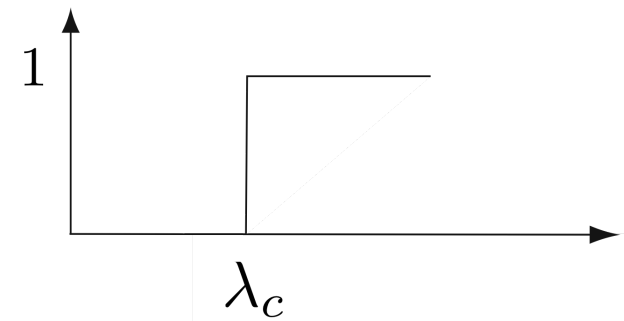
Continuum Percolation theory

Random geometric graph on $\mathbb{R}^2$

$\theta(\lambda) = P_\lambda(|C_0| = \infty)$



$\phi(\lambda) = P_\lambda(E_\infty)$



$\lambda_c \pi \approx 4.51$ (*Monte Carlo*, Balister, Bollobás, Walters, 2005)

$\mathbb{E}(\text{degree}) = \pi r^2 \lambda = \lambda\pi$

4.5 friends are enough to traverse the world !

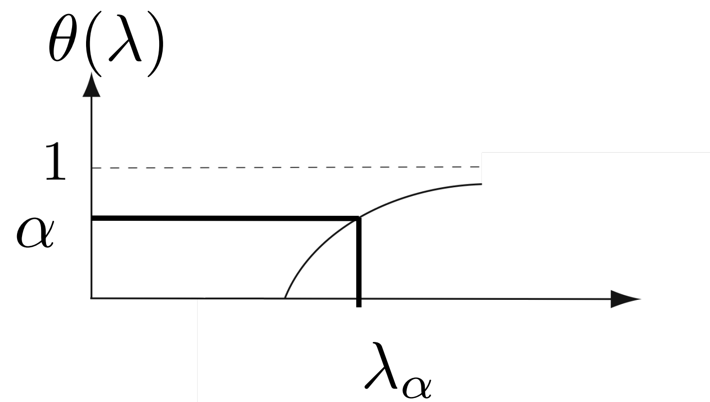
Giant component

$B_n \equiv \sqrt{n} \times \sqrt{n}$ box

There are on average n nodes inside the box

What fraction in the largest connected component?

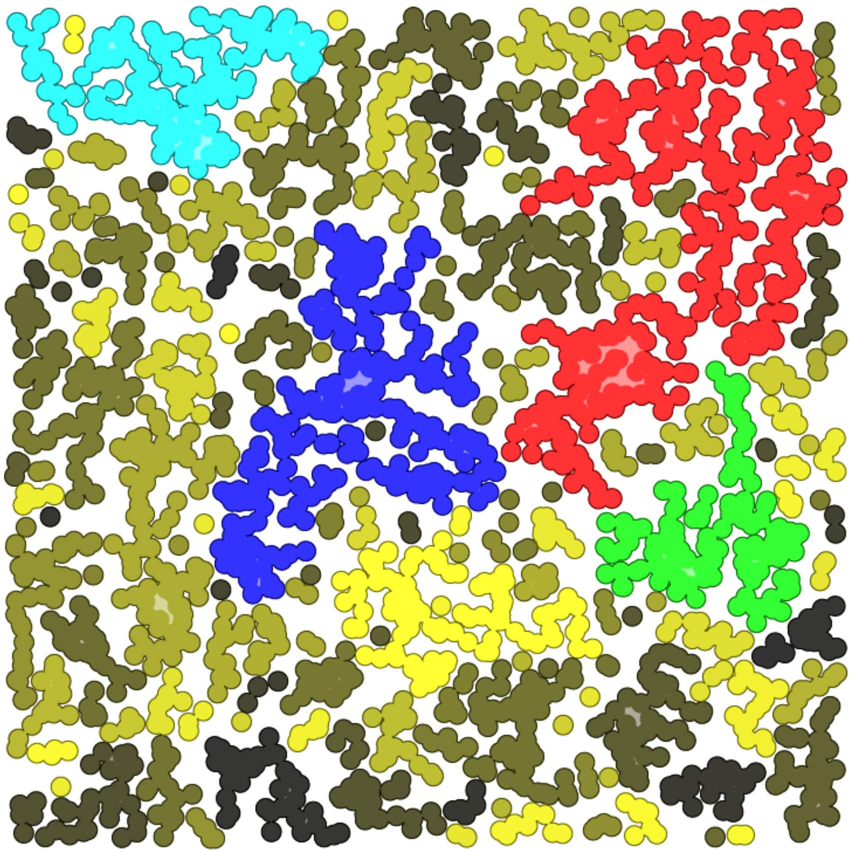
$$\lambda_\alpha = \inf\{\alpha : \theta(\lambda) > \alpha\}$$



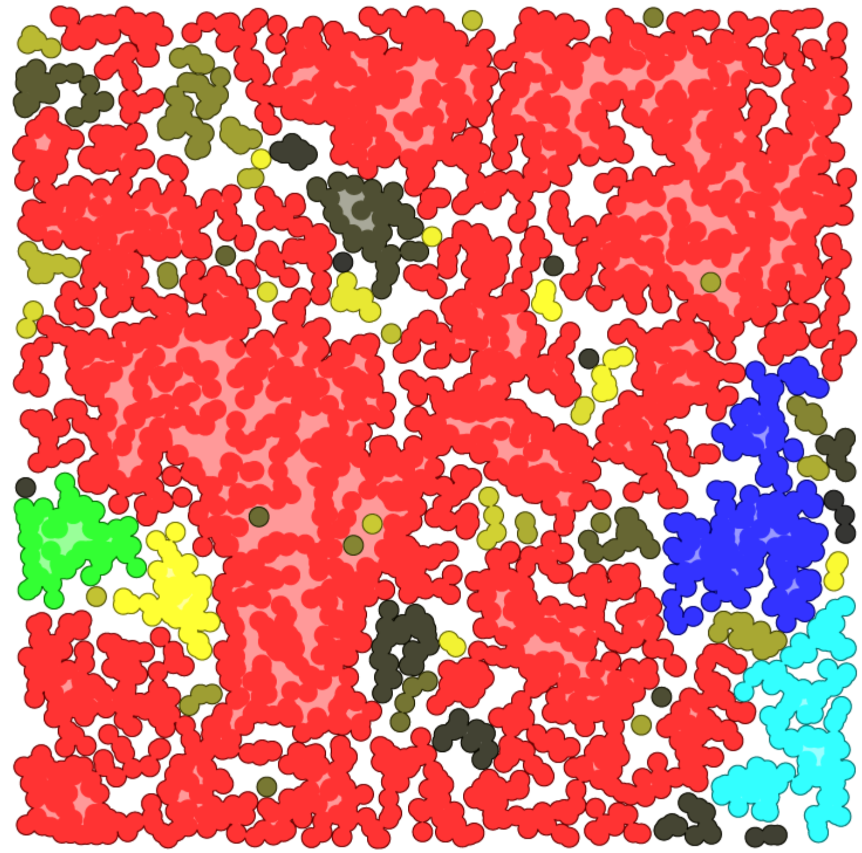
$$\lambda < \lambda_\alpha \implies \lim_{n\to\infty} P(|C_n| \geq \alpha n) = 0$$

$$\lambda > \lambda_\alpha \implies \lim_{n\to\infty} P(|C_n| \geq \alpha n) = 1$$

Giant component



$\lambda < \lambda_c$



$\lambda > \lambda_c$

Expected node degree

In terms of expected node degree, the threshold for a giant component corresponds to an expected node degree of 4.5

As the expected node degree is increased further, the fraction alpha of nodes in the giant component increases.

To achieve full connectivity, corresponding to alpha =1, the expected node degree must increase with n

At what rate?

Full connectivity

$$\pi\lambda = \log n + \beta_n$$

$$\lim_{n\to\infty} P(G_n \text{ is fully connected}) \iff \beta_n \to \infty$$

Penrose (1997)

The threshold for full connectivity corresponds to an expected node degree of log *n*

$$(1-\epsilon)\log n \implies \text{disconnected}$$

$$(1+\epsilon)\log n \implies \text{connected}$$

Summary

A constant expected number of neighbors is enough for the formation of a giant component

A logarithmically growing number is enough for full connectivity

Random geometric networks exhibiting both component properties and community structure

How about network diameter?

This is "small" for random graphs but "large" for random geometric graphs

And how about power law degree distributions?

Need more sophisticated models…