Graph Mining
SD212
2. Graph models

Thomas Bonald

2018 – 2019

Motivation

Random graph = random instance of a graph with some specific statistical properties

Useful for:

- generating graphs "for free"
- testing algorithms (simulation)
- proving algorithms (analysis)

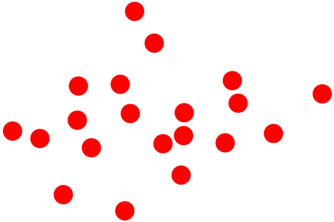
We focus on **undirected** graphs; the results naturally extend to directed graphs

Outline

1. Erdös-Rényi graphs
2. Preferential attachment
3. Configuration model

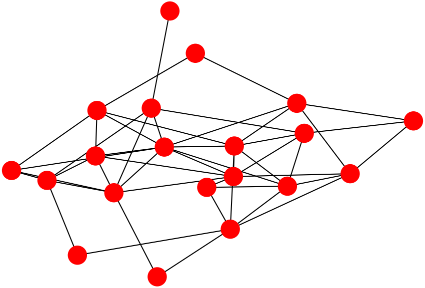
Erdös-Rényi graphs

- n nodes
- $p \in (0, 1)$
- An edge with probability p between any distinct nodes u, v



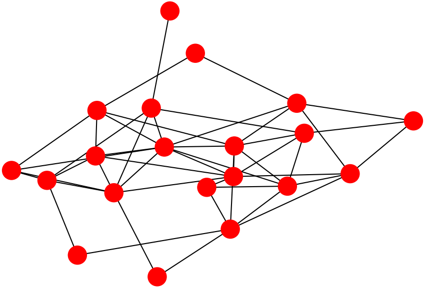
Erdös-Rényi graphs

- n nodes
- $p \in (0, 1)$
- An edge with probability p between any distinct nodes u, v



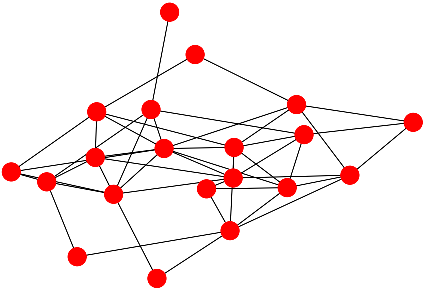
Erdös-Rényi graphs

- n nodes
- $p \in (0, 1)$
- An edge with probability p between any distinct nodes u, v
- Degree distribution



Erdös-Rényi graphs

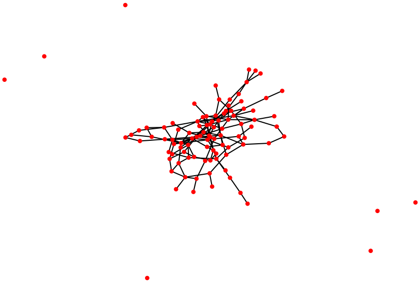
- n nodes
- $p \in (0, 1)$
- An edge with probability p between any distinct nodes u, v
- Degree distribution (all instances)
- Empirical degree distribution (one instance)



Large Erdös-Rényi graphs

- $n \to +\infty$
- $p \to 0$
- $np \to \lambda$

Example with $n = 100$, $p = 0.03$ ($\lambda = 3$)



Local structure

A tree!

Galton-Watson tree

Recursive definition:

- A root
- The offspring of each node has a Poisson distribution with parameter λ

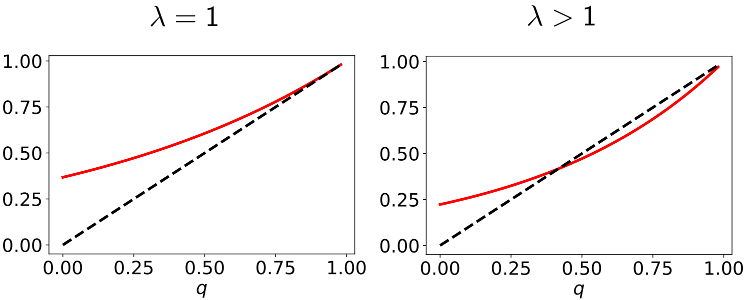
Three regimes

Extinction probability

Assume $\lambda \geq 1$

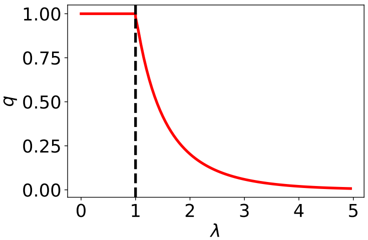
Fixed-point equation

$$q = e^{\lambda(q-1)}$$



Extinction probability

$$q = e^{\lambda(q-1)}$$



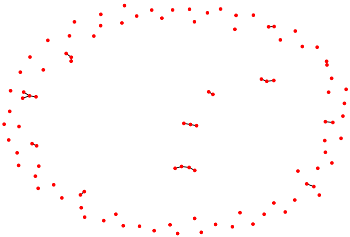
Back to Erdös-Rényi graphs

Three regimes:

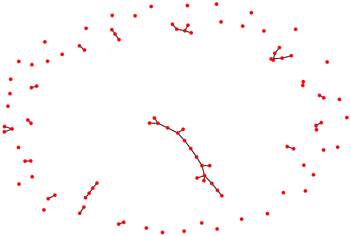
- **Subcritical ($\lambda < 1$):** finite tree $\rightarrow$ many small components
- **Supercritical ($\lambda > 1$):** infinite tree with probability $p = 1 - q$ $\rightarrow$ one **giant component** containing a fraction p of the nodes (the others in small components)
- **Critical ($\lambda = 1$):** finite tree (but infinite expectation) $\rightarrow$ one large component

Examples

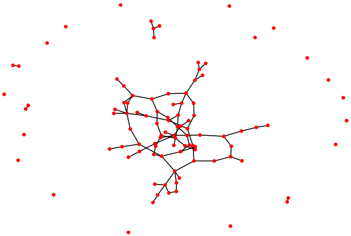
$\lambda = 0.5$



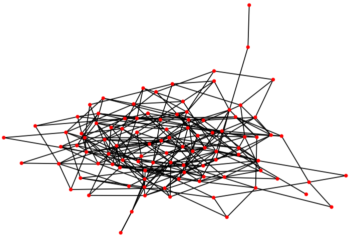
$\lambda = 1$



$\lambda = 2$



$\lambda = 5$

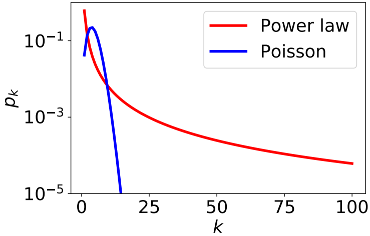


Outline

Power law

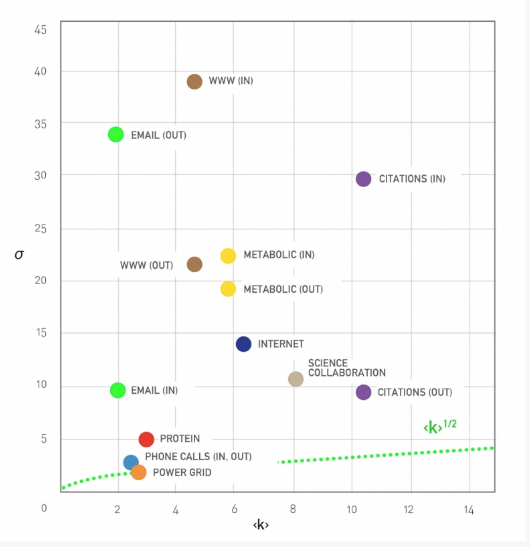
- The typical degree distribution of real graphs is of the form

$$p_k \propto \frac{1}{k^\alpha} \quad \alpha > 1$$



- Explained by the **"rich get richer"** phenomenon
  Barabasi & Albert 1999

Scale-free graphs



Source: Barabasi, Network Science, 2016

Barabasi-Albert model

Explicit construction (with $d \geq 1$):

- Start from a clique of d nodes
- Add new nodes one at a time, each of degree d and with **preferential attachment**

Outline

1. Erdös-Rényi graphs
2. Preferential attachment
3. **Configuration model**

Configuration model

Given some sequence of integers $d_1, \ldots, d_n$, can we generate a graph with this particular **degree sequence**?

Havel-Hakimi algorithm

Random configuration

Number of self-loops

Summary

Random graphs are generated from **models**

- Erdös-Rényi graphs $\rightarrow$ No structure
- Preferential attachment $\rightarrow$ Power law
- Configuration model $\rightarrow$ Degree sequence

These may be combined to get more realistic (but complex) models