Graph Mining
SD212
2. Random graphs

Thomas Bonald

2017 – 2018

Motivation

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

Motivation

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

Useful for:

- generating graphs “for free”
- testing algorithms (simulation)
- proving algorithms / providing performance guarantees (analysis)

Motivation

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

Useful for:

- generating graphs "for free"
- testing algorithms (simulation)
- proving algorithms / providing performance guarantees (analysis)

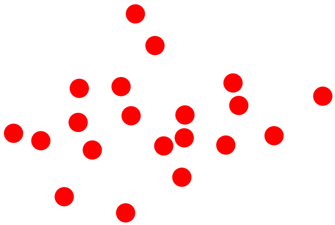
We focus on **undirected** graphs; the results naturally extend to directed graphs

Outline

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

Erdös-Rényi graphs

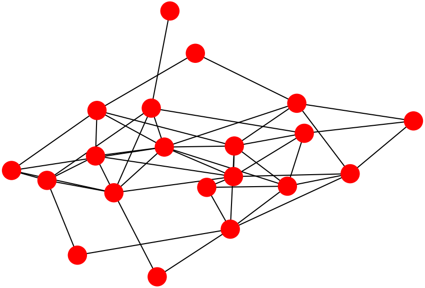
- n nodes
- $p \in (0, 1)$
- An edge with probability p between any distinct nodes u, v



~B(n,p)

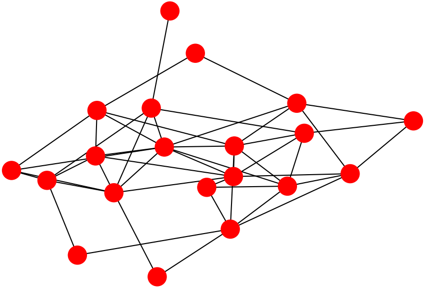
Erdös-Rényi graphs

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



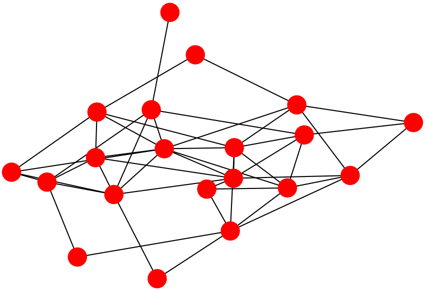
Erdös-Rényi graphs

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



Erdös-Rényi graphs

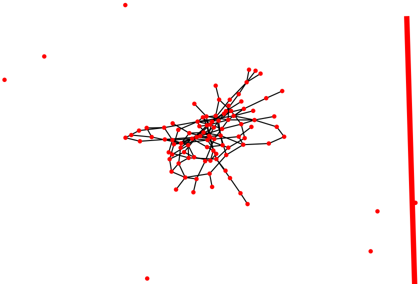
- n nodes
- $p \in (0, 1)$
- An edge with probability p between any distinct nodes u, v
- Degree distribution (all instances)
- Empirical degree distribution (one instance)

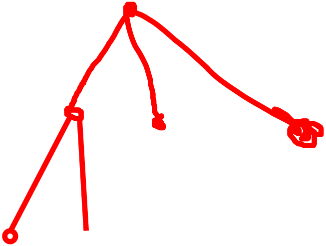


Large Erdös-Rényi graphs

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

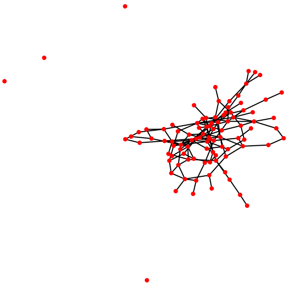
Example with $n = 100$, $p = 0.03$ ($\lambda = 3$)





Large Erdös-Rényi graphs

- $n \to +\infty$
- $p \to 0$
- $np \to \lambda$
- Degree distribution (all instances)
- Empirical degree distribution (one instance)



~P(\lambda)

d1,....,dn~ iid

p_k = 1/n sum_n i_{d_u = k}
= E(i_{d_u = k}) = e^{\lambda} *
\lambda^k /k!

Galton-Watson tree

Recursive definition:

- A root
- The offspring of each node has a Poisson distribution with parameter λ n-> +\infy, p->0, (n-n_new)p = \lambda

X ~ P(\lambda)

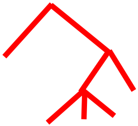
G(t) = E(t^*)
=\sum \frac{\lambda}{k!} e^{-\lambda}t^k
= e^{\lambda(t-1)}
G_k(t) = E(k^{2k})

Z_{k+1}= \sum^z2_{i=2}X_i
G_{k+1}(t) = E(t^{Z_{k+1}) = E(t^{Z_{\sum^z2_{i=2}X_i})
= \sum_{b>=0} P(Z_k = n}E(t^{\sum_1^n X_i}) = G_k(G(t))
G_{k+1} = G_k(G(E)) = G \o ... \o G(t) (k+1 个G复合)

E(Z_{k+1}) = \lambdaE(Z_k)

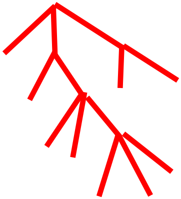
Three regimes

\lambda < 1:



\frac{1}
{1-\lambda}

\lambda > 1:

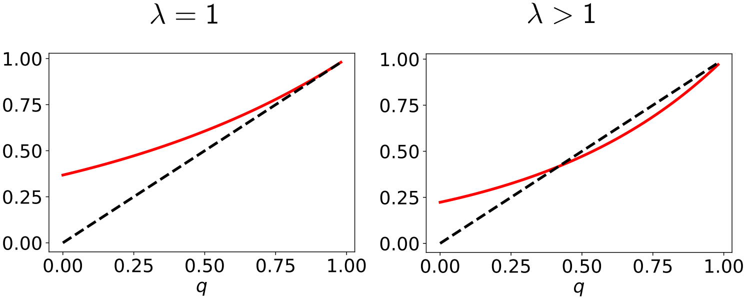


Extinction probability

Assume $\lambda \geq 1$ `q = \lim_{k -> + \infinity} P(Z_{k=0})`

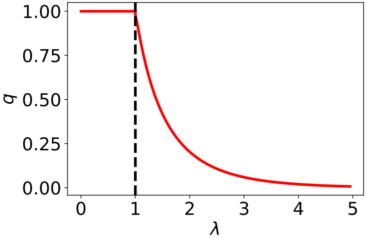
Fixed-point equation

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



Extinction probability

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



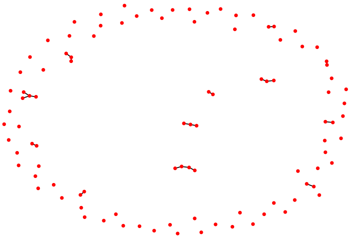
Back to Erdös-Rényi graphs

Three regimes:

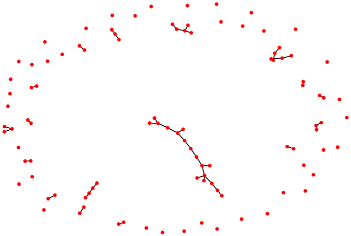
- **Subcritical ($\lambda < 1$):** finite tree $\rightarrow$ many small components
- **Supercritical ($\lambda > 1$):** infinite tree with probability $p = 1 - q$ $\rightarrow$ one **giant component** containing a fraction p of the nodes (the others in small components)
- **Critical ($\lambda = 1$):**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

Examples

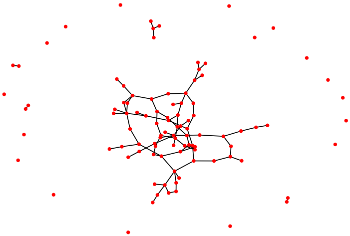
$\lambda = 0.5$



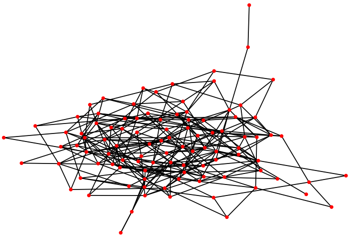
$\lambda = 1$



$\lambda = 2$



$\lambda = 5$



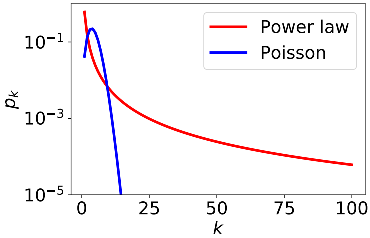
Outline

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

Power law

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

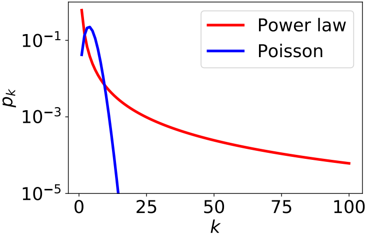
$$p_k \propto \frac{1}{k^\alpha} \quad \alpha > 1$$



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

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**
4. Stochastic block 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

http://blog.51cto.com/sbp810050504/883904

Random configuration

Number of self-loops

Outline

Stochastic block model

Random graphs with some underlying **community** structure

- A partition of $\{1, \ldots, n\}$ in k blocks
- A block connectivity matrix B

Summary

Random graphs are generated from **models**

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

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