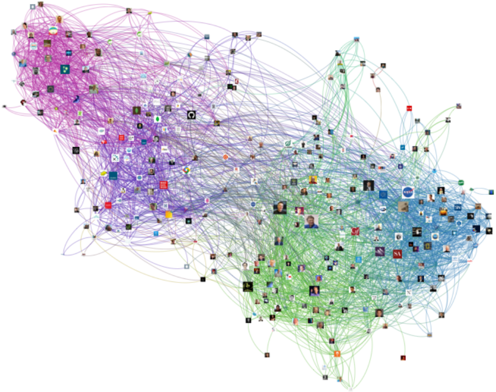
Graph Mining
SD212
1. Node and edge sampling

Thomas Bonald

2017 – 2018

The friendship paradox

"Your friends have more friends than you on average!"



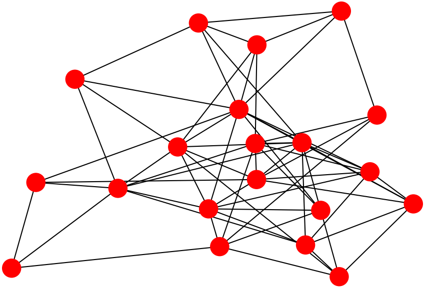
Outline

Setting

- Some **undirected** graph $G = (V, E)$ of n nodes and m edges

$$V = \{1, \dots, n\} \quad E \subset \{\{u, v\}, u, v \in V\}$$

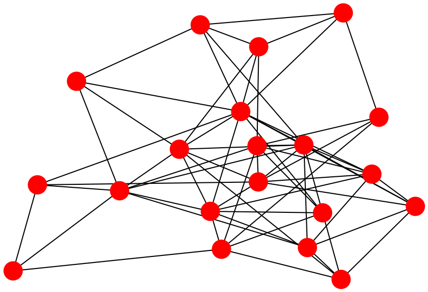
- No self-loops
- d_u, degree of node u



Setting

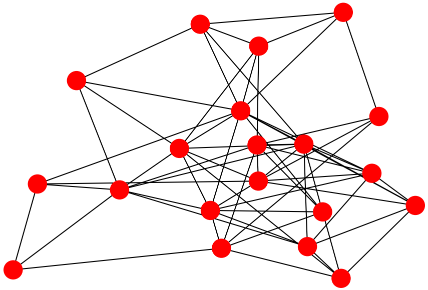
Proposition

$$\sum_u d_u = 2m \qquad \bar{d} \equiv \frac{1}{n} \sum_u d_u = \frac{2m}{n}$$



Random node

Sample a **node** uniformly at random



Random node

Sample a **node** uniformly at random

- Degree distribution

$$p_k = \frac{1}{n} \sum_u 1_{\{d_u = k\}}$$

Random node

Sample a **node** uniformly at random

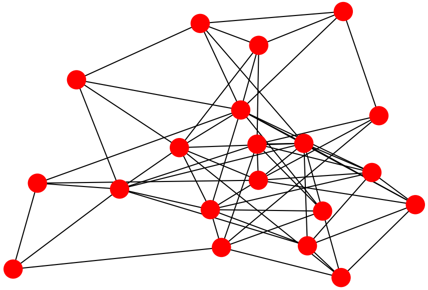
- Degree distribution

$$p_k = \frac{1}{n} \sum_u 1_{\{d_u = k\}}$$

- Expected degree

Random edge

Sample an **edge** uniformly at random and one of the two ends of this edge uniformly at random



Random edge

Sample an **edge** uniformly at random and one of the two ends of this edge uniformly at random

- Degree distribution

$$\hat{p}_k \propto k p_k$$

  This is the **size-biased** distribution

Random edge

Sample an **edge** uniformly at random and one of the two ends of this edge uniformly at random

- Degree distribution
$$\hat{p}_k \propto k p_k$$
  This is the **size-biased** distribution
- Expected degree

Example: power law distribution

- Typical degree distribution of real graphs

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

Example: power law distribution

- Typical degree distribution of real graphs

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

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

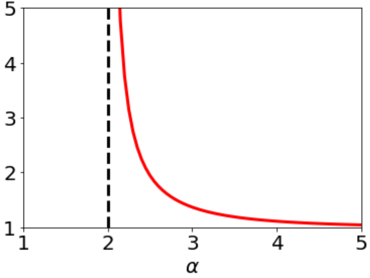
Example: power law distribution

- Typical degree distribution of real graphs

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

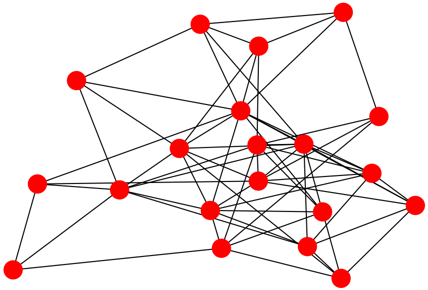
- Explained by the **"rich get richer"** phenomenon
  Barabasi & Albert 1999
- Sampling bias

$$\frac{\mathrm{E}(\hat{X})}{\mathrm{E}(X)}$$



Random neighbor

Sample a **neighbor** uniformly at random (from a random node)



Random neighbor

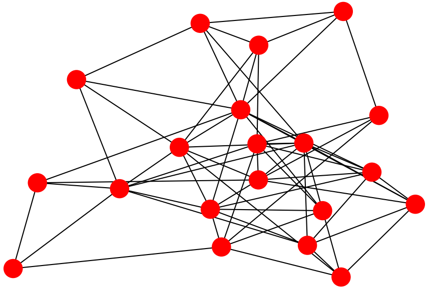
Sample a **neighbor** uniformly at random (from a random node)

Proposition

The expected degree is **larger** than $\bar{d}$

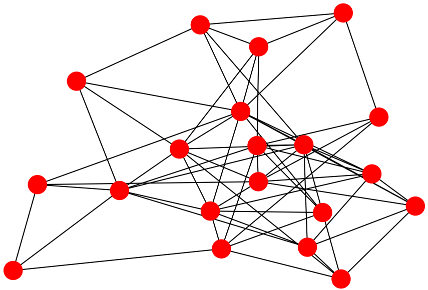
Degree correlation

- Take the degrees of two **neighbors** (random edge)

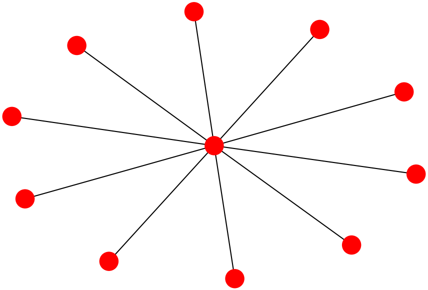


Degree correlation

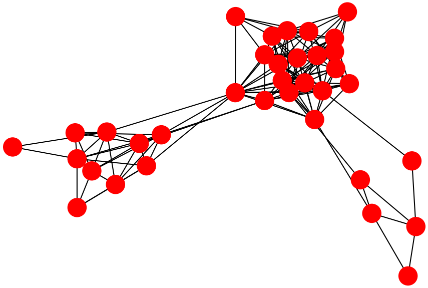
- Consider the degrees of two **neighbors** (random edge)
- Their correlation indicate the **assortativity** of the graph



A disassortative graph



An assortative graph



Outline

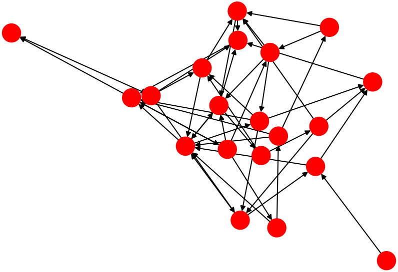
1. Undirected graphs
2. Directed graphs

Setting

- Some **directed** graph $G = (V, E)$ of n nodes and m edges

$$V = \{1, \dots, n\} \quad E \subset \{(u, v), u, v \in V\}$$

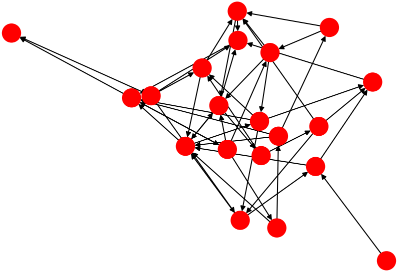
- d_u^+ / d_u^-, out-degree/in-degree of node u



Setting

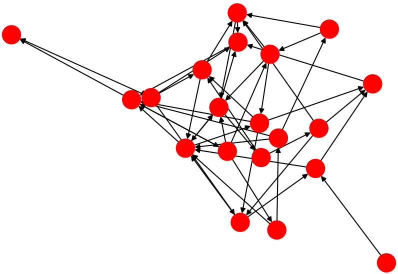
Proposition

$$\sum_u d_u^+ = \sum_u d_u^- = m \qquad \bar{d}^+ = \bar{d}^- = \frac{m}{n}$$



Random node

Sample a **node** uniformly at random



Random node

Sample a **node** uniformly at random

- Out/in degree distributions

$$p_k^+ = \frac{1}{n}\sum_u 1_{\{d_u^+=k\}} \quad p_k^- = \frac{1}{n}\sum_u 1_{\{d_u^-=k\}}$$

Random node

Sample a **node** uniformly at random

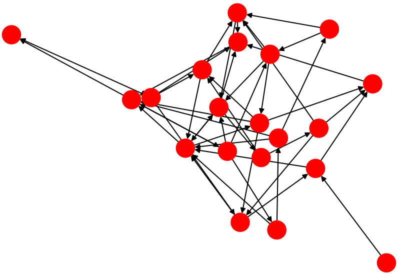
- Out/in degree distributions

$$p_k^+ = \frac{1}{n}\sum_u 1_{\{d_u^+ = k\}} \quad p_k^- = \frac{1}{n}\sum_u 1_{\{d_u^- = k\}}$$

- Expected out/in degrees

Random edge

Sample an **edge** uniformly at random



Random edge

Sample an **edge** uniformly at random

- Out-degree of the **origin** / in-degree of the **destination**

$$\hat{p}_k^+ \propto k p_k^+ \quad / \quad \hat{p}_k^- \propto k p_k^-$$

Random edge

Sample an **edge** uniformly at random

- Out-degree of the **origin** / in-degree of the **destination**

$$\hat{p}_k^+ \propto k p_k^+ \quad / \quad \hat{p}_k^- \propto k p_k^-$$

- Expected out/in-degrees

Random neighbor

Sample a **successor** / **predecessor** uniformly at random
(from a random node)

Summary

- Sampling from **nodes / edges / neighbors** is not the same!
- Something to think about

Are you **really** popular?

Hint: Look at your friends