Scale Free Networks

Machine Learning / Complex Systems

Many structures in the real world can be mapped to graphs by describing the relations between entities in terms of nodes and edges

Stanford Network Analysis Project

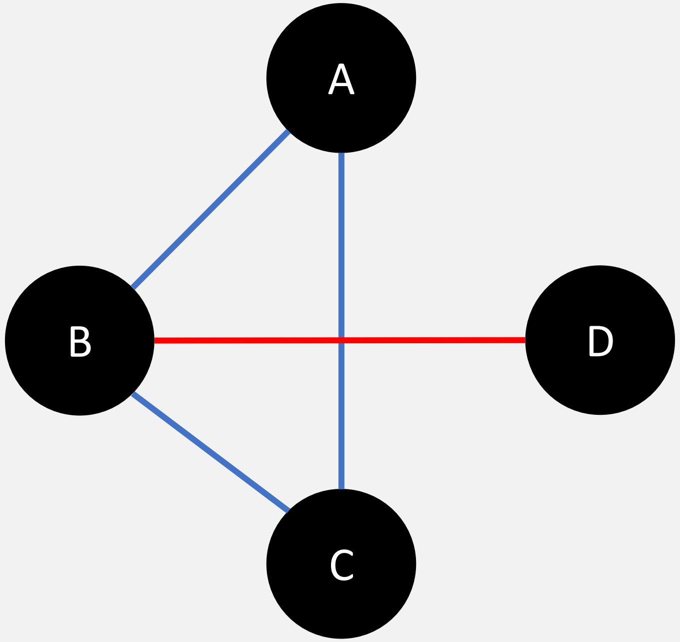
Physics Collaboration Graph

Twitter Higgs Dataset

Beeradvocate Reviews

Author A
Author B
Author C

Author B
Author D



Graph Evolution: Densification and Shrinking Diameters - J. Leskovec, J. Kleinberg and C. Faloutsos

Many structures in the real world can be mapped to graphs by describing the relations between entities in terms of nodes and edges

Stanford Network Analysis Project

Physics Collaboration Graph

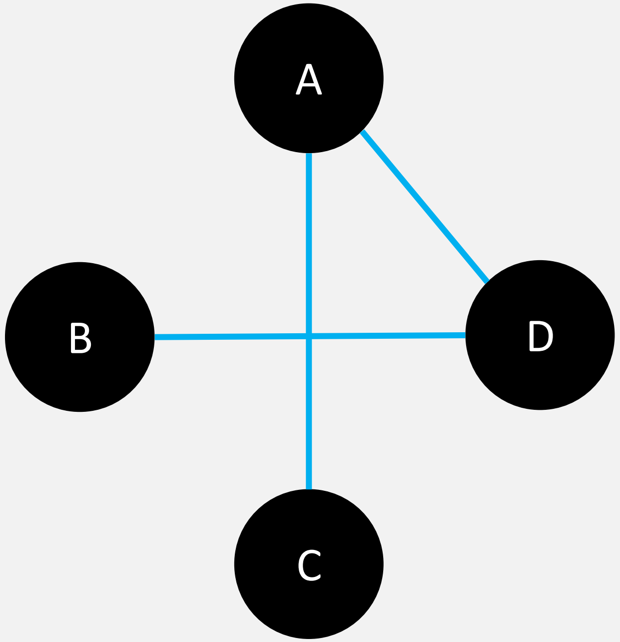
Twitter Higgs Dataset

Beeradvocate Reviews

A retweets C

B replies to D

D mentions A



The Anatomy of a Scientific Rumor – M. De Domenico, A. Lima, P. Mougel and M. Musolesi

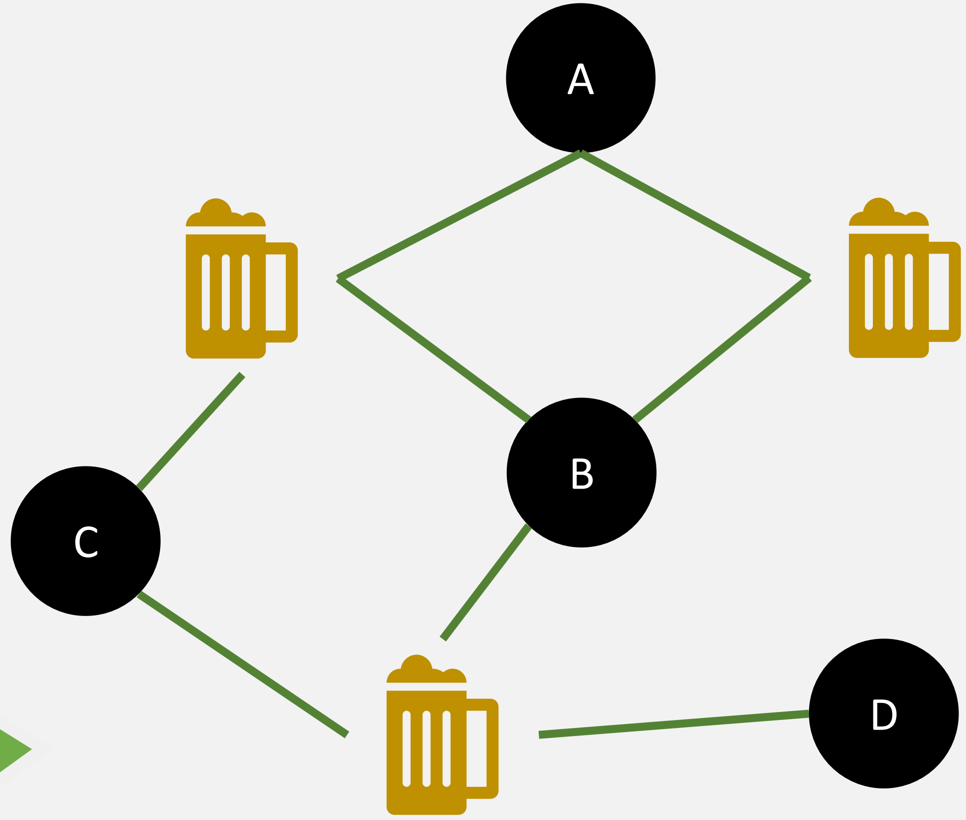
Many structures in the real world can be mapped to graphs by describing the relations between entities in terms of nodes and edges

Stanford Network Analysis Project

Physics Collaboration Graph

Twitter Higgs Dataset

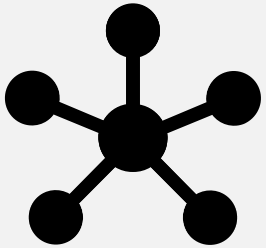
Beeradvocate Reviews



From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews.
- J. McAuley and J. Leskovec

Degree distributions are a way to characterize networks. They are given by the probability distribution for a node to have k edges

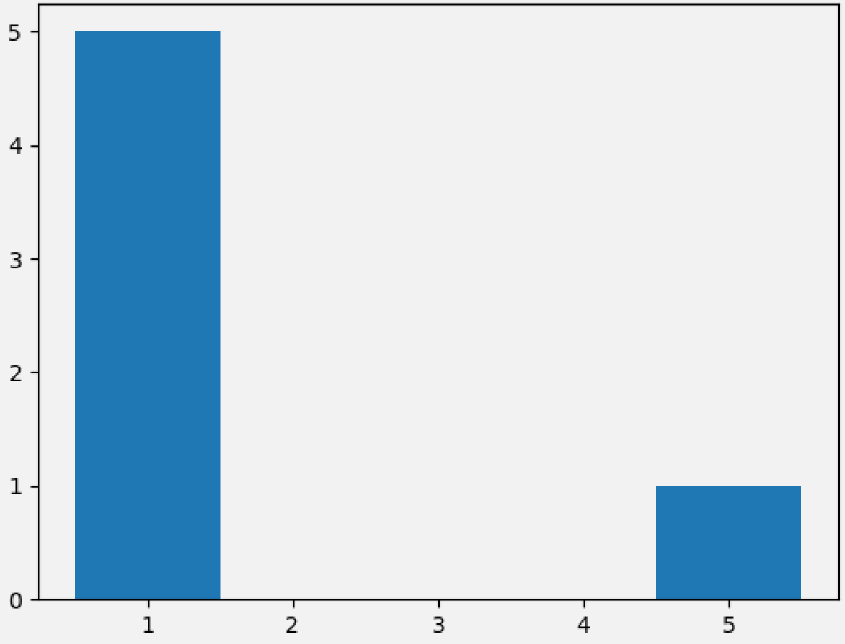
Graph



Adjacency matrix:

Node	1	2	3	4	5	6
1	0	0	0	0	0	1
2	0	0	0	0	0	1
3	0	0	0	0	0	1
4	0	0	0	0	0	1
5	0	0	0	0	0	1
6	1	1	1	1	1	0

Histogram



Degree distribution

k	p(k)
0	0
1	5/6
2	0
3	0
4	0
5	1/6

In this talk: Power law distributions $p(k) \sim k^{-y}$

Academic studies support the hypothesis that real world networks have a power law distribution

Word Wide Web

- Nodes are websites
- Edges are links

- Size: 325729
- Average Degree: 4.51
- $\gamma_{In} = 2.1$
- $\gamma_{out} = 2.45$

Movie Actor Collaboration

- Nodes are Actors
- Edge between actors if they appear in the same movie

- Size: 212250
- Average Degree: 28.78
- $\gamma \quad = 2.3$

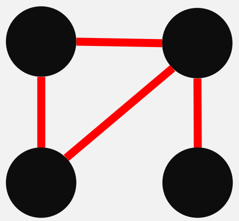
Cellular Networks

- Nodes are substrates
- Edges are chemical reactions

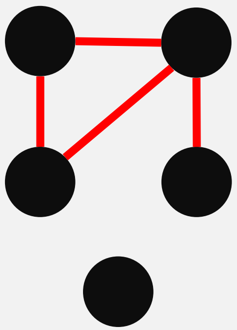
- $\gamma \quad \in [2.0, 2.4]$

A power law distribution can be achieved by modelling the dynamics of real world networks

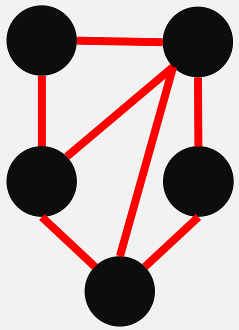
Start with random Network of size m_0



Network Grow



Preferential Attachment



1. Start with a random network of size m_0
2. Add a new node to the Network
3. Attach the new node to the already present nodes with $m \leq m_0$ edges
4. The probablility to attach to a node with degree k_i is given by $\Pi(k_i) = \frac{k_i}{\sum_{j=1} k_j}$
5. Repeat at step 2

Using the continuum approach assuming k is a continuous variable over time leads to an differential equation for k

$$\frac{\partial k_i}{\partial t} = m\,\Pi(k_i) = m\frac{k_i}{\underbrace{\sum_{j=1}^{N-1} k_j}} \approx \frac{k_i}{2t}$$

$$\sum_{j=1}^{N-1} k_j = 2mt - m$$

$\Rightarrow$ $k_i(t) = m\left(\frac{t}{t_i}\right)^{\frac{1}{2}}$ With $k_i(t_i) = m$

$$P(k_i(t) < k) = P\left(t_i > \frac{m^2 t}{k^2}\right) = 1 - P\left(t_i \leq \frac{m^2 t}{k^2}\right) = 1 - \sum_{l=1}^{\frac{m^2 t}{k^2}} P(t_i = t_l) = 1 - \frac{m^2 t}{k^2(t + m_0)}$$

Since $P(t_i) = \frac{1}{m_0 + t}$

$\Rightarrow$ $P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} \sim \frac{1}{k^3}$

Apart from the degree distribution there are several other properties which can be measured

Average Degree

- Average number of edges per node
- Barabsi-Albert: 2m

Average Path length

- Average path length from one node to all other nodes averaged over all nodes
- Barabasi-Albert: $\sim\frac{\ln(N)}{\ln(\ln(N))}$

Average Cluster coefficient

- $\frac{3 \times \#Triangles\ in\ the\ Graph}{\#Connected\ triples}$
- No analytical result for Barabsi-Albert graphs

Numerical calculation:

1. Create an ensemble of graphs
2. Compute the given property for every graph in the ensemble
3. Average over the ensemble

Error and attack tolerance are properties of networks strongly motivated by real world observations

Error Tolerance

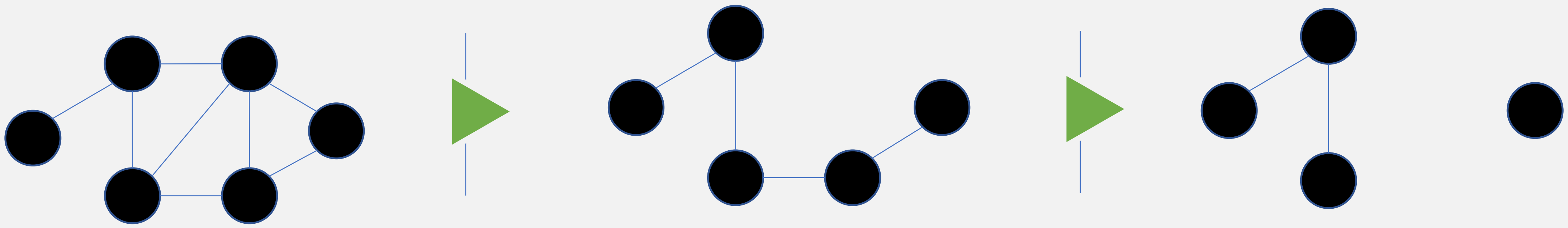
- How does the Network react to random failuers of nodes

Attack Tolerance

- How does the network react to targeted attacks

Use the average path length of the largest component to quantify the changes under errors/attacks

- Measures differences in the network when the network stays connected
- Stays finite even when the network contains more than one component



In the Barabási-Albert Model the oldest nodes will gain the most edges. This does not match always with reality

Idea: Give each node an additional constant fitness parameter to adjust the preferential attachment

$$\frac{\partial k_i}{\partial t} = m\,\Pi(k_i) = m\frac{\eta_i k_i}{\sum_{j=1}^{N-1} \eta_j k_j}$$

where $\eta \sim p(\eta)$

Bose Einstein Condensation

- Assign to each node an energy
- An edge between two nodes corresponds to two particles at the linked energy levels

Energy levels are given by: $\epsilon_i = -\frac{1}{\beta}\ln(\eta_i)$

The distribution of energy levels tranforms according to the probability density transformation law:
$g(\epsilon) = \beta\, p\left(e^{-\beta\epsilon}\right)e^{-\beta\epsilon}$

The fitness model is closely linked to statistical physics and Bose-Einstein Condensation

$$I(\beta, \mu) = \int d\epsilon \, g(\epsilon) \frac{1}{e^{\beta(\epsilon-\mu)} - 1} = 1$$

Integral has a solution

Integral has no solution

Fit get rich phase

- The fraction of particles on every energy level decays to zero as $t \to \infty$

Bose Einstein Condensation

- A finite fraction of particles will occupy the fittest node as $t \to \infty$

References

- Emergence of Scaling in Random networks – Barabasi, A.-L. and R. Albert (1999)
- Statistical Mechanics of complex – Albert R. and A.-L. Barabasi (2002)
- Competition and multiscaling in evolving networks – Bianconi, G. and A.-L. Barabasi (2000)
- Bose-Einstein condensation in complex networks – Bianconi, G. and A.-L- Barabasi (2000)

GitHub

- www.github.com/HOminus/ScaleFreeNet