University at Buffalo
The State University of New York
Department of Industrial and Systems Engineering
School of Engineering and Applied Sciences

From Small-World to Scale-Free Networks: What Do Real-World Networks Tell Us?

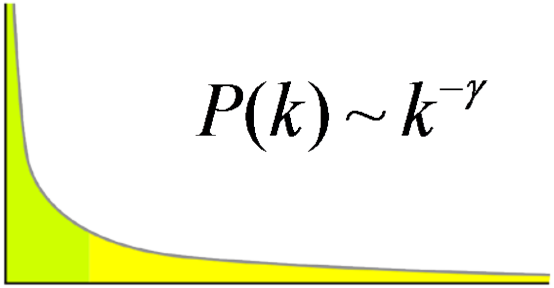
Sushant Khopkar, Alexander Nikolaev
University at Buffalo

Rakesh Nagi
University of Illinois at Urbana-Champaign

Large Scale Real Social Networks

- Real Social Networks, e.g., Facebook and Twitter, are known to exhibit the following three properties:

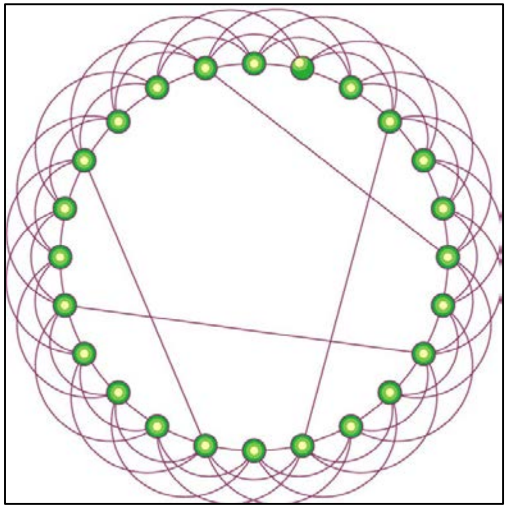
1. Small Average Path Length,
2. High Clustering Coefficient,
3. Power Law Degree Distribution.



$$P(k) \sim k^{-\gamma}$$

Small World Network Model

- A friendship network formation model by Watts & Strogatz (1998)



- The model features

1) small average path lengths and
2) high clustering coefficient.

- However, the degree distribution of typical W&S network does not follow a power law.

Scale-Free Model and Preferential Attachment

- The concept of “rich gets richer” is proposed (Price 1965, Barabasi & Albert 1999).
- Generated networks exhibit power laws in their degree distributions.
- Apart from “popular people tend to become more popular”, it does not provide any intuitive explanation for how people form links in real world.
- Another limitation – assumption about global visibility.

Open Questions

- Is there an alternative to preferential attachment that could explain emergence of power law based on local information spread, i.e., without assuming ‘full visibility’?

- What is the contribution of talent of an individual as opposed to the network position in gaining popularity (# of followers)?

Our Findings

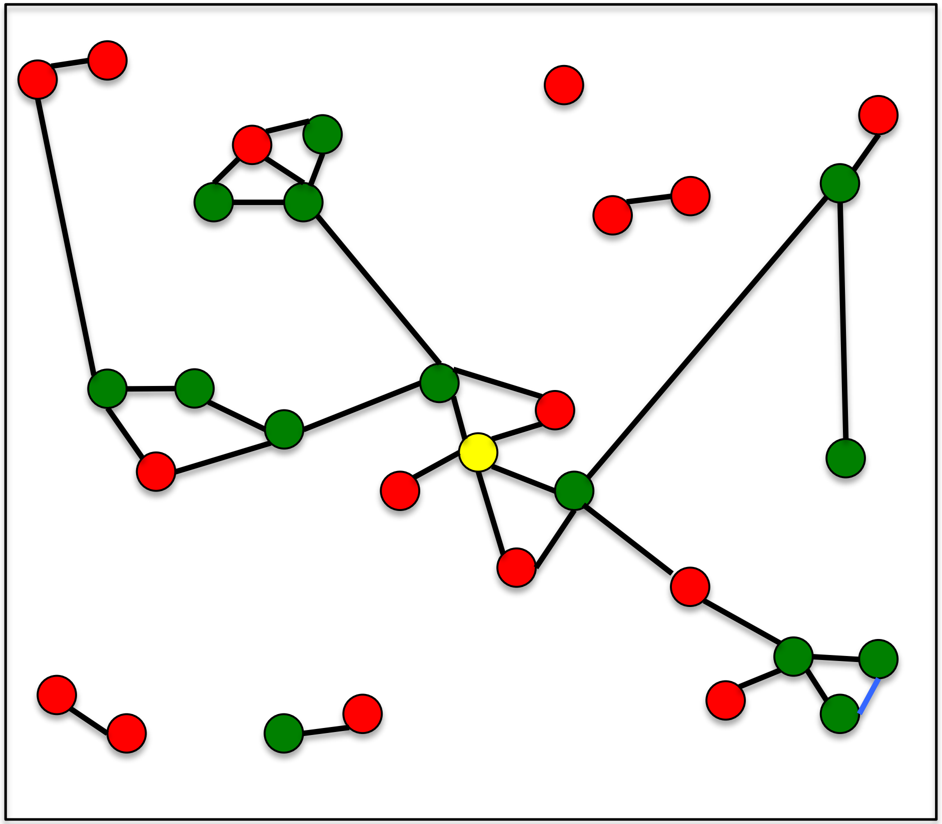
- On a lattice friendship network, a power law degree distribution emerges via follower type link accumulation.
- Few talented (or not so talented but lucky) individuals may become popular as word spreads about their talent. Rare events play a key role in this process.
- A highly intuitive spread model explains the connection between personal friendship networks and scale-free network formation.

Components of the Model

- Potential Popularity : each person has a potential number of followers in the world based on his / her talent. These potential followers may not be aware of the given person.

- This talent can be a vector of different talents (singing, sports, acting, IQ) or could be a scalar. These talent values (e.g., IQ) of people in population follow a normal distribution.

Potential Popularity and Visibility



Yellow node : A Talent

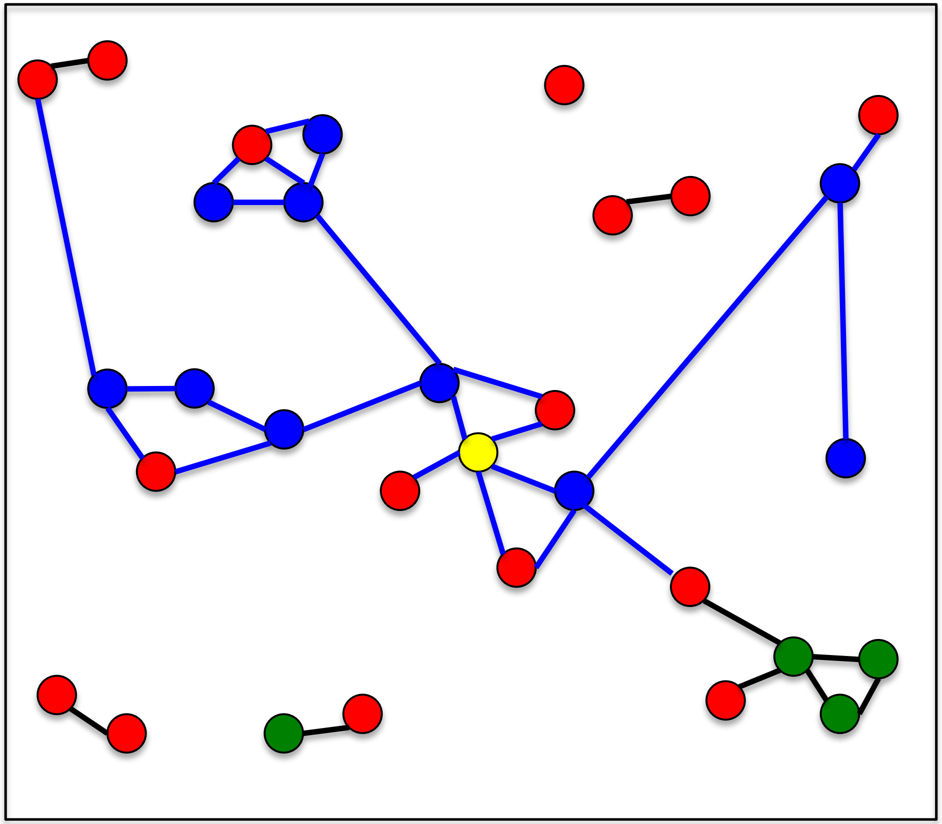
Green nodes :
Potential Followers

Red nodes :
Potential Non-Followers

A Watts-Strogatz lattice
is as underlying information
transfer (friendship) network.

The given node is initially
“visible” only to its
“first-circle” neighbors.

Phase 1: Spread of Popularity



Popularity spreads only through visible potential followers.

If word of the talented person reaches a potential follower node (green node), It will become an actual follower (blue node), and will spread the word further.

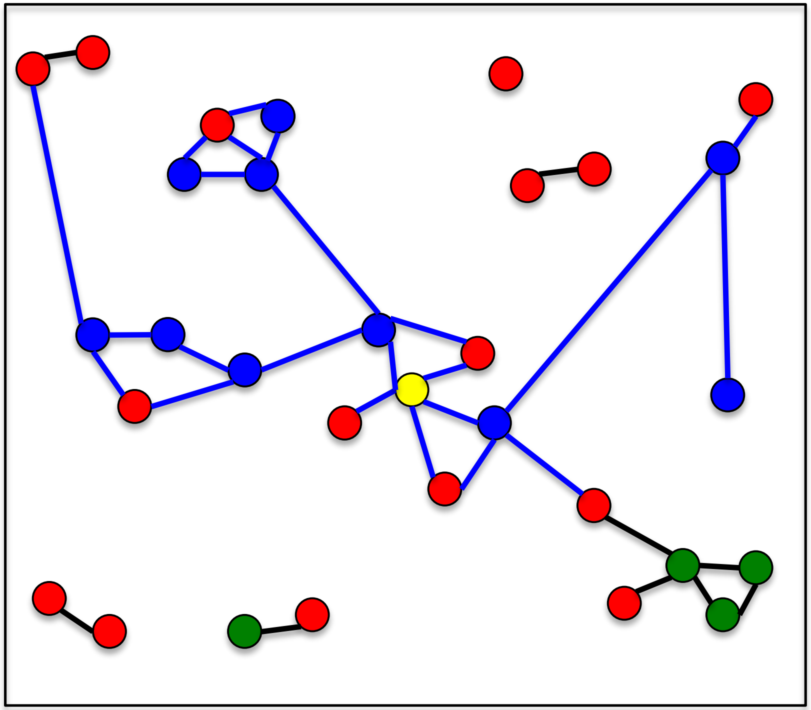
The red nodes do not spread the word: these actors do not appreciate the talent.

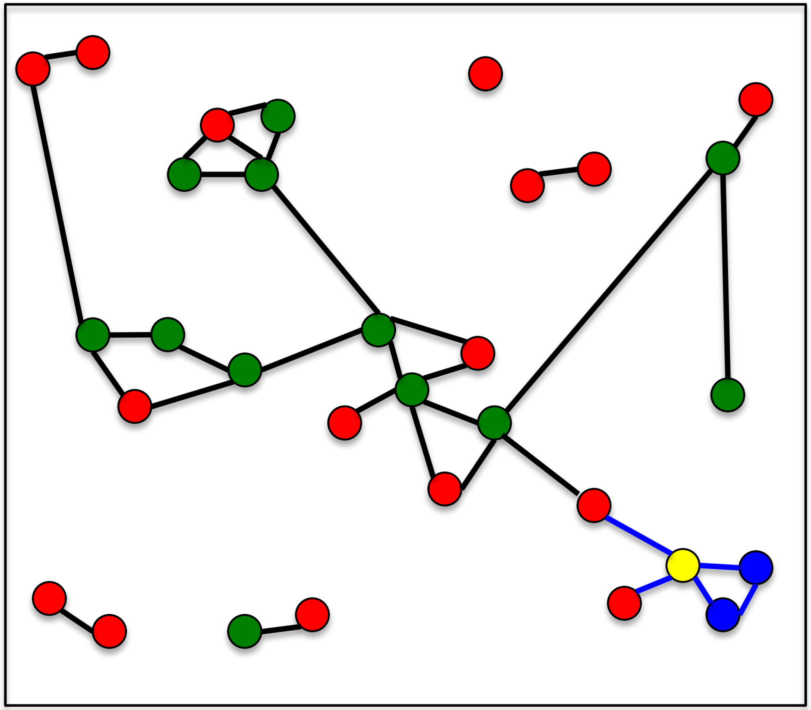
Probability of Impression

- Appreciation of talent is modeled as a Bernoulli R.V., with probability ‘p_g’ reflecting the level of talent.
- In viral spread models, probability ‘p_g’ is known as ‘Probability of Susceptibility’ (Newman, 2002).
- The distribution of ‘p_g’ over the entire population is assumed to be normal. It is known that IQ of humans is normally distributed (Jensen, 1998).

The "Luck" Factor: two equally talented persons can be unequally popular.

Graph Size = 30. Blue edges show spread of popularity of the yellow node.

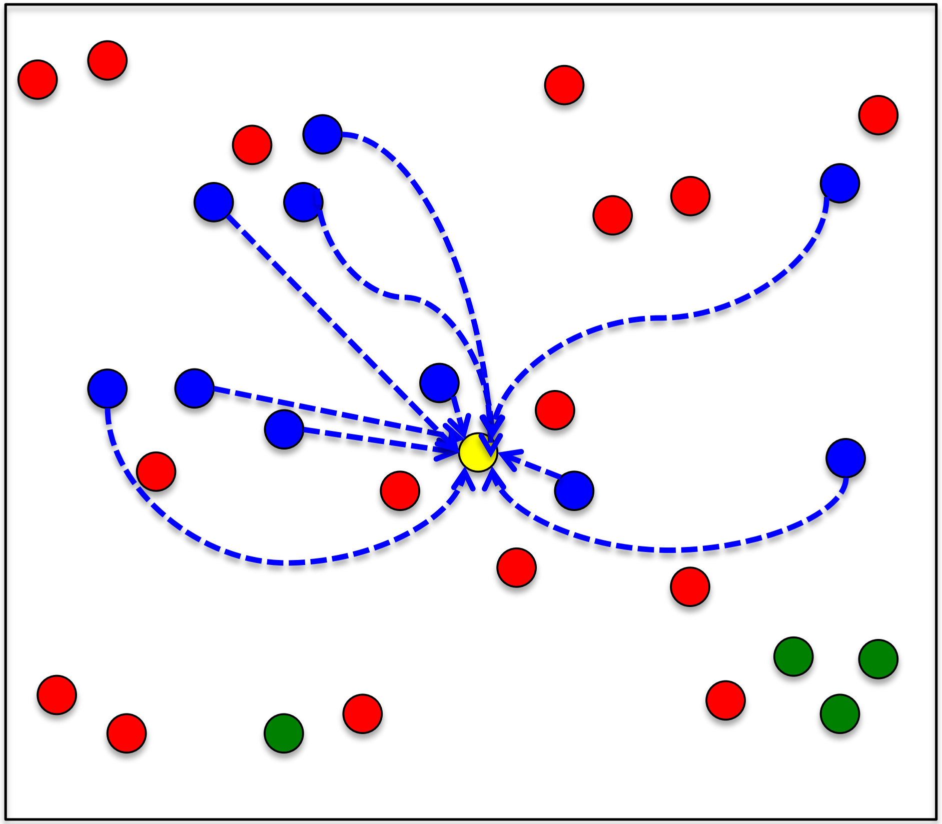




A
Potential Followers = 14
High Visibility
Potential Followers Reached = 10

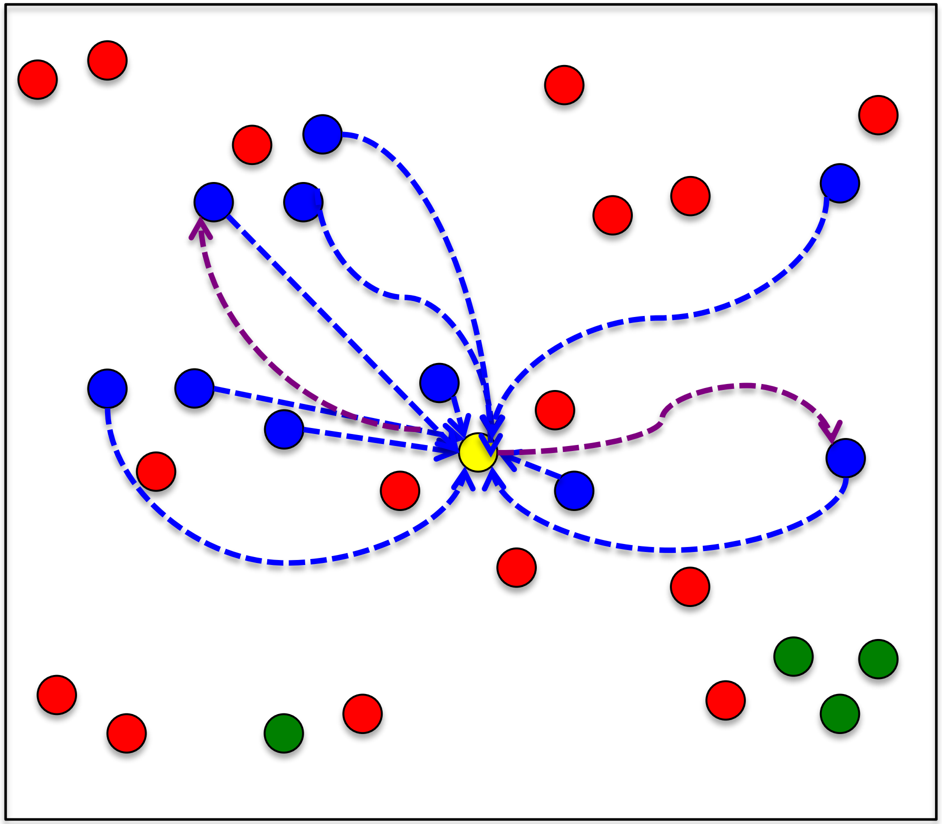
B
Potential Followers = 14
Low Visibility
Potential Followers Reached = 2

Follower Type Network



In thus formed follower type network, all friendship ties (undirected) become invisible / hidden and only directed, follower type links (blue dotted arrows) appear and remain visible.

Phase 2: Reciprocity



Another way to gain followers is by reciprocity.

If node ***A*** starts following node ***B***, ***A*** becomes visible to node ***B***. Now, ***B*** will follow back ***A*** with a probability = probability of impression of node ***A***.

Curved dotted violet colored edges show reciprocal edges.

Parameters of the Model

The proposed network formation model has 5 parameters:

1. Size of the small-world network (N)
2. Initial number of neighbors ($2k$)
3. Rewiring/shortcut probability (ϕ)
4. Mean of Probability of Impression (Mean_p_g)
5. Standard Deviation of Probability of Impression (SD_p_g)

Simulation Experiments

To perform a simulation experiment:

1. Select values of the 5 parameters
2. Using *2k* and ϕ generate a small-world network
3. Run Phase 1 – Spread of Popularity
4. Run Phase 2 – Reciprocity

A Sample Simulation Experiment

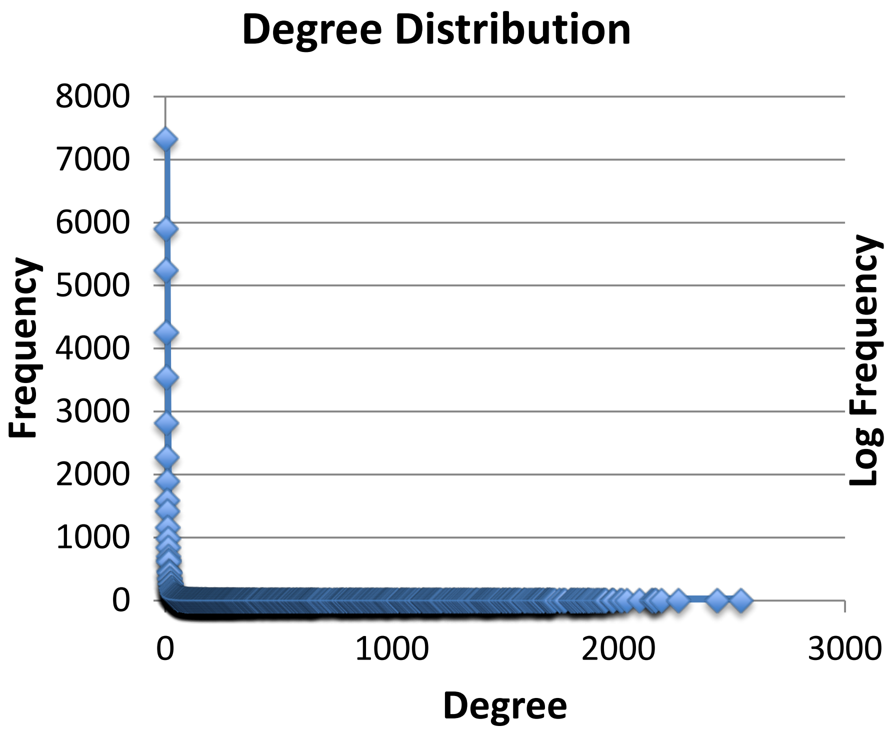
Select N = 50,000.

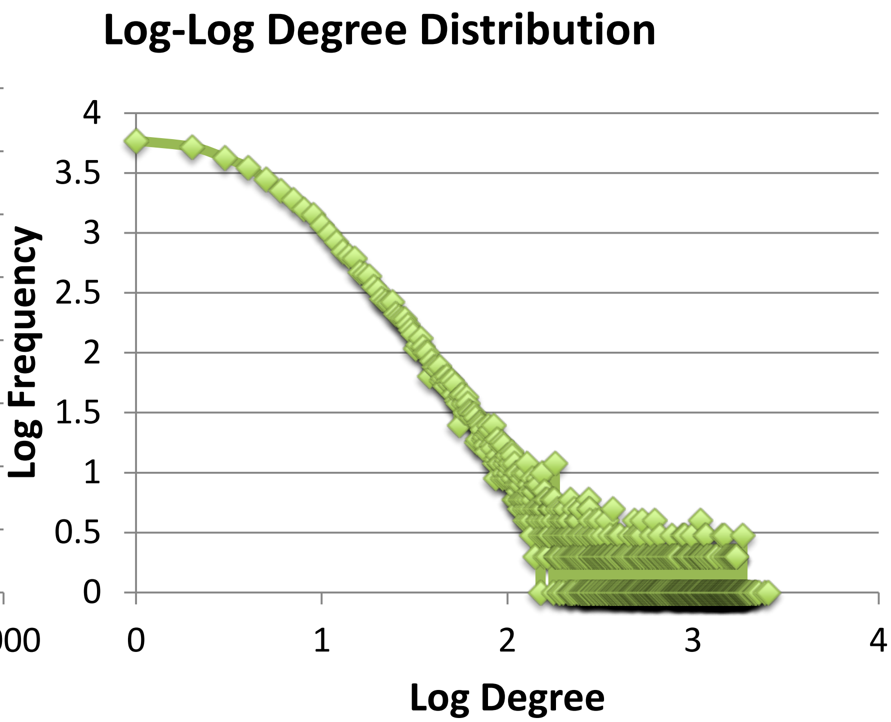
- Select any node as the first given (yellow) node.
- Draw its value of probability of impression from a normal distribution. It will be used in Bernoulli trials for each node, to decide potential followers (green nodes) and non followers (red nodes).
- Calculate the final spread of popularity (with *reciprocity*) as the in-degree of the given yellow node. This is just a single data point! Repeat it 50,000 times.

Sample Result 1

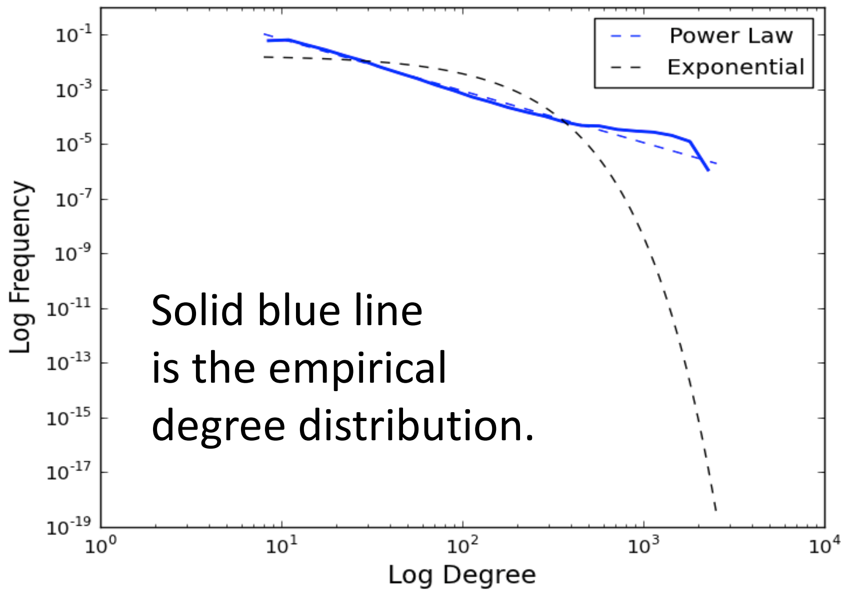
- $2k = 30$, $\phi = 0.01$,
- $Mean\_p_g = 0.08$, $SD\_p_g = 0.04$

Power Law!





Sample Result 1: (Confirming Power Law)



Ref: (Clauset et. al, 2009)

gamma (exponent)	1.892727905
X_min	8
D	0.019287281

D is Kolmogorov-Smirnov distance between the two distributions.

Comparison with Other Distributions			
First Distribution	Second Distribution	**Likelihood Ratio**	**p-value**
Power Law	Exponential	54.28980773	0.00E+00
Power Law	Lognormal	2.18006186	2.93E-02
Power Law	Stretched Exponential	22.12968161	1.64E-108

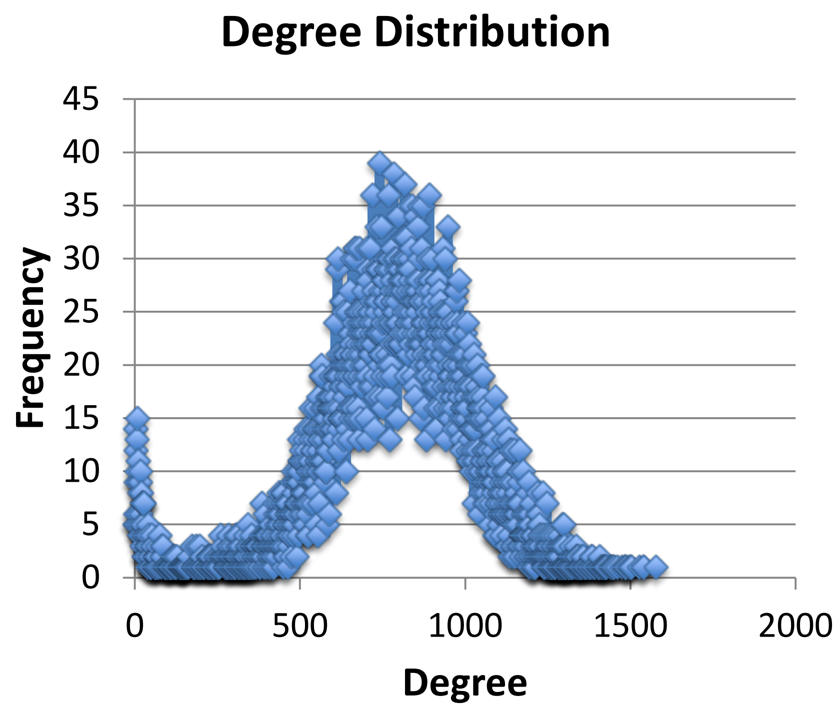
Positive and high value of likelihood ratio indicates that it's more likely to be the first distribution (a power law) than the second distribution.

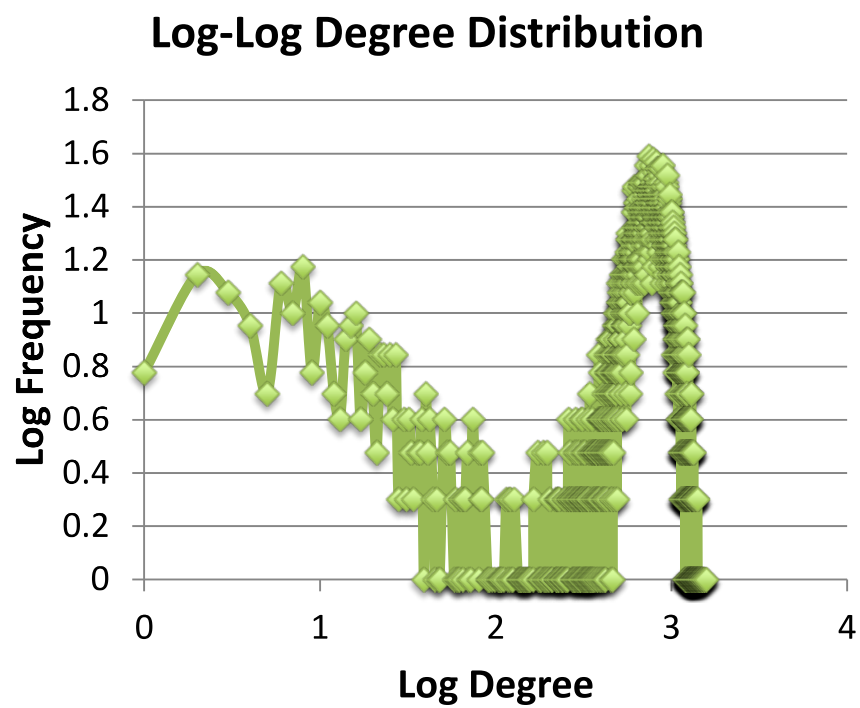
Sample Results 2 : We don't always get a power law!

- $2k = 150$, $\phi = 0.01$,
- Mean_p_g = 0.08, SD_p_g = 0.02

Parameter values affect the shape

Not a Power Law...





Percolation Threshold

- In-degree = # of potential followers reached + number of followers received by reciprocity (Phase 2).
- 'The number of potential followers reached' is governed by a threshold of probability of impression, called the 'percolation threshold'.
- Percolation threshold is defined as the value of the probability of susceptibility (impression), beyond which almost all of the potential followers become visible with very high probability.

Percolation Threshold

- The theoretical value of the percolation threshold (p_c) depends on the rewiring/shortcut probability (ϕ) and initial number of neighbors, ($2k$):

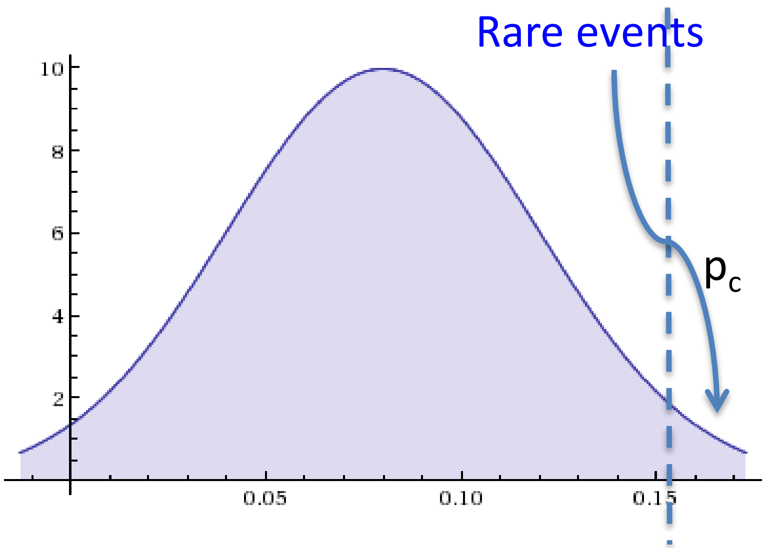
$$\phi = \frac{(1-p_c)^k}{2kp_c[2-(1-p_c)^k]}$$

(Moore, Newman 2000)

Sample Results

Rare events are responsible for power laws!

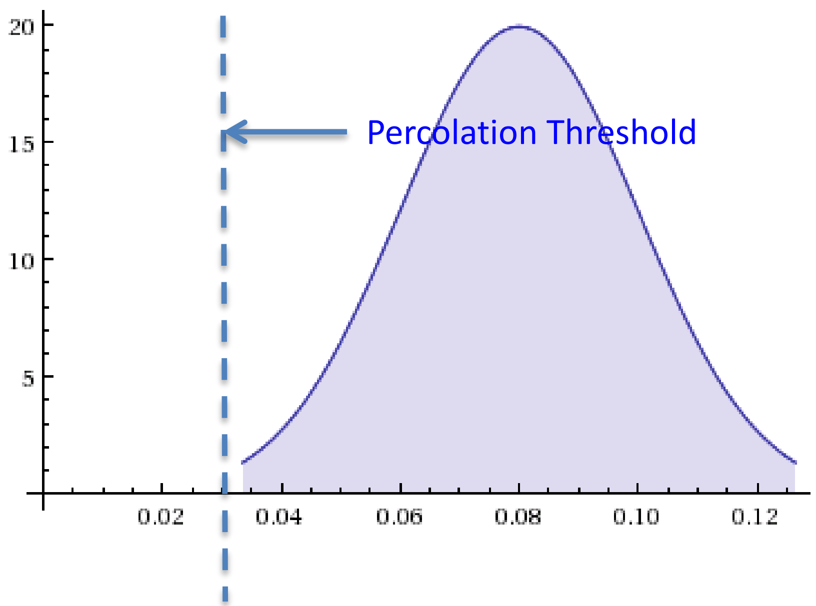
1) $2k$ =30, ϕ = 0.01, Mean_p_g = 0.08, SD_p_g = 0.04, p_c = 0.15050



Distribution of Probability of Impression

Results in a Power Law

2) $2k$ =150, ϕ = 0.01, Mean_p_g = 0.08, SD_p_g = 0.02, p_c = 0.0315447



Distribution of Probability of Impression

Does not result in a Power Law

Numerical Method: Percolation Process

- Apart from simulation, it is possible to calculate an in-degree distribution of a graph numerically, for a given set of parameters.

- In-degree distribution after Phase 1 is analogous to probability distribution of sizes of connected clusters for a fixed p_g, derived in (Moore, Newman 2000). It is done by calculating a Cauchy's integral of a function by FFT (Fast Fourier Transform) method.

Numerical Method: Percolation Process

- To get the in-degree distribution when p_g is normally distributed, the normal distribution is divided into *k* intervals and *k* values of probability of impression are used, taken as means of those intervals.

- Degree distribution generated using each value of p_g is then multiplied by the probability mass associated with that interval. All these distributions are then combined together to get the resulting in-degree distribution.

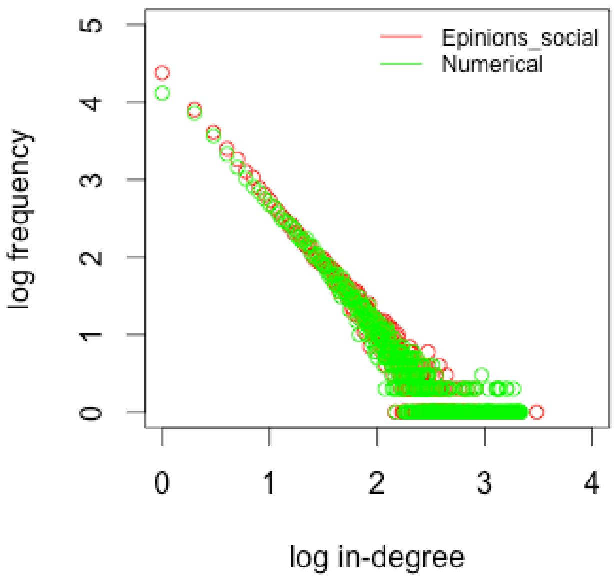
Learning Parameters for Real Graphs

- It is of interest to learn the model parameters for real world networks.
- For social, communication, citation networks…
  – What is the initial number of neighbors? What is the mean of talent distribution? Are these values similar across different types of networks?....
- Instead of adding phase 2 in numerical method, we remove reciprocal ties from a real world graph and apply phase 1 on it to learn the parameter values.

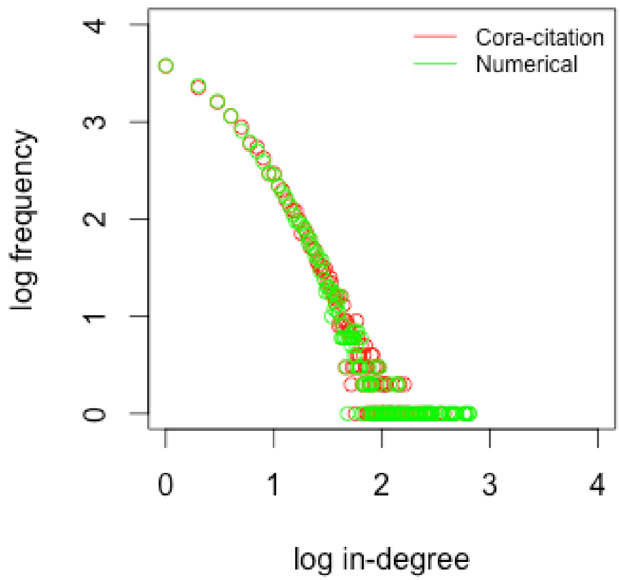
Nelder-Mead Method

- Also called “Downhill Simplex Method” - a numerical method used to find *min* or *max* of an objective function in a multi-dimensional space.
- Our objective function = log of absolute distance between frequencies of the true degree distribution and degree distribution of a generated graph.
- The method is applied to nonlinear optimization problems for which derivatives may not be known – our case!
- But it’s a heuristic method. Multiple starting points should be used while running the algorithm, to avoid local minima.

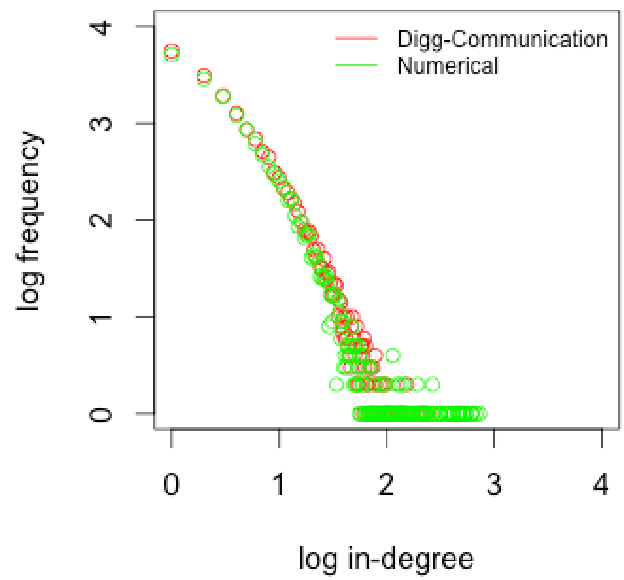
Sample Output of Parameter Estimation



Social



Citation



Communication

This work is in progress. The Nelder-Mead algorithm performance has now been validated in multiple synthetic experiments, promising insightful studies with real-world data.

Conclusion / Discussions

- It is follower-type directed links (and not friendship type undirected links) that are responsible for the emergence of power laws.
- Rare events, which are functions of individual "talents", network positions and luck, are responsible for the fat tails in the (power law) distributions.
- The proposed follower type link formation model does not only generate networks having power law degree distributions, but also gives a good intuitive explanation of the probable underlying processes behind them.
- Numerical method along with a Nelder-Mead algorithm are used to estimate parameter values of real world graphs and to gain insights.

Thank you!

- Khopkar, S., Nikolaev, A., & Nagi, R. (2016) “Towards Understanding the Laws Behind Small world and Scale Free Network Formation” (working paper)