Emergence of Scaling in Random Networks

By: Barabási, A., & Albert, R.

Summary and Review

By: Jeff Hildebrandt

Overview of the paper

- Takes advantage of new technologies and knowledge to model scale-free networks
- The networks follow power law distribution
- The networks are composed of vertices and edges that link the vertices
- The networks are constantly adding new vertices and new edges are added
- The model addresses expanding vertices and how new vertices prefer to link to already well established vertices

Examples of the networks described in this paper

- Genetic network
  - Vertices are proteins and genes
  - Edges are the chemical interactions between them
- Nervous system
  - Vertices are the nerve cells
  - Edges are axons
- Social science
  - Vertices are individuals or organizations
  - Edges are the social interactions between them
- Internet
  - Vertices are web pages
  - Edges are links

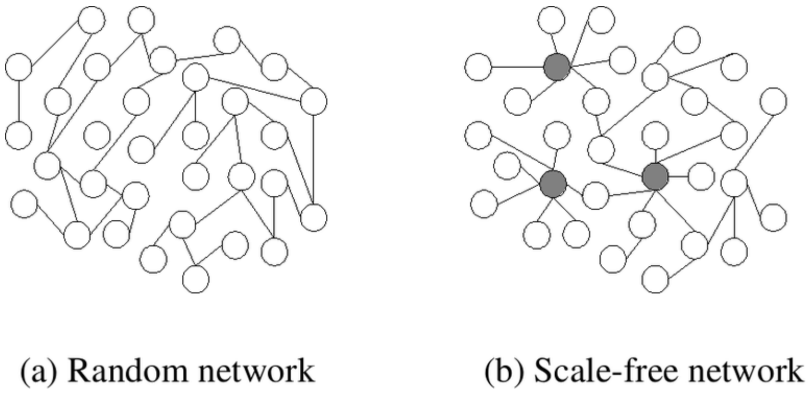
Motivation

- No good way to describe these systems
- Doing so would advance many disciplines
- Current models don’t consider adding new vertices or why some vertices have substantially more edges connecting to them
- Emergence of detailed topological data for some systems thanks to databases and the internet

Goal

- Make a model that simulates real networks that follow the scale free distribution
- Prove that networks can self-organize into a scale-free state via preferential attachment
- Show how current models fail when trying to model these networks

Scale-free networks



https://en.wikipedia.org/wiki/Scale-free_network

- scale-free power-law distribution
  - Represents two variables where when one variable changes it has a relative affect on the other variable
- Using the internet and bots combing data from sites to obtain the variables for the power law
- Power law
  - $P(k) \sim k^{-\gamma}$
    - P(k) is the probability that a vertex in the network interacts with k other vertices
  - K represents the number of edges adjacent to a given vertex
  - As numbers of edges increases the chance to have a large number of edges decreases
- Used three easily accessible examples

Scale-free networks

- Collaboration graph of movie actors
  - Vertices are actors
  - Edges are if two actors were cast in the same movie
  - $\gamma_{actor} = 2.3 \pm 0.1$
  - N = 212,250 vertices
- Internet
  - Vertices are web pages
  - Edges are links
  - Used bots to comb the web to collect hyperlinks to other web pages
  - $\gamma_{www} = 2.1 \pm 0.1$
  - N = 325,729 vertices
- Electric power grid in the US
  - Vertices are generators, transformers and substations
  - Edges are power lines
  - Due to small number of vertices there is a relatively high exponent variable
  - $\gamma_{power} \cong 4$
  - N = 4941 vertices

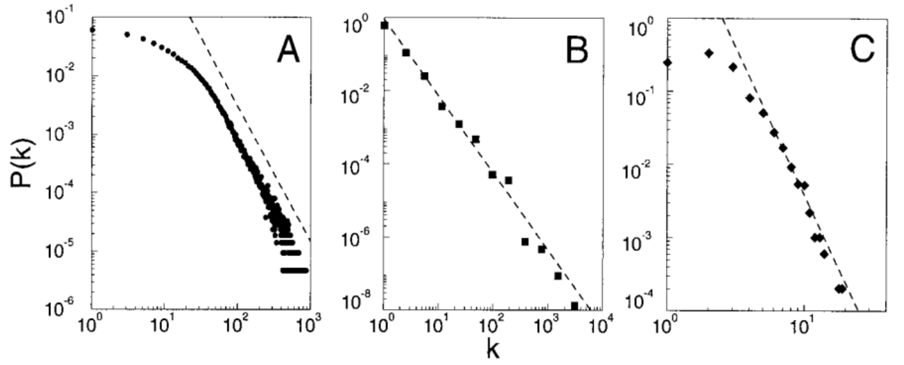
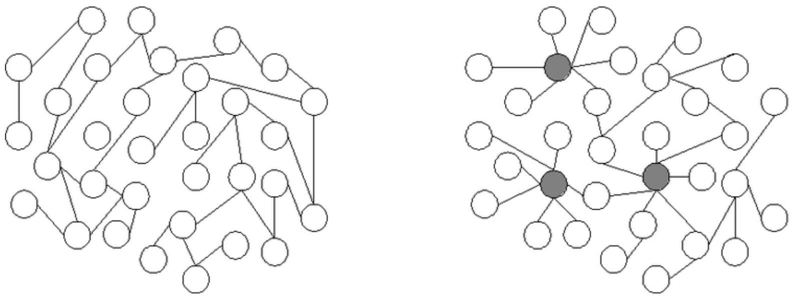


Fig. 1. The distribution function of connectivities for various large networks. **(A)** Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. **(B)** WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). **(C)** Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

Existing Models for complex topology

- Erdős–Rényi
  - Starts with N vertices and each vertex has an equal chance to attach to any other vertex with probablility p
  - Follows a Poisson distribution
    - Vertices have no affect on where other vertices connect
    - Does not follow the power-law in scale-free networks



(a) Random network (b) Scale-free network

https://en.wikipedia.org/wiki/Scale-free_network

- Watts-Strogatz
  - Creates graphs with small-world properties
    - Vertices are connected with their neighbors
    - Each edge connects to the next vertices with probablility p
  - Leading to six degrees of separation
    - "**Six degrees of separation** is the idea that all living things and everything else in the world are six or fewer steps away" (https://en.wikipedia.org/wiki/Six_degrees_of_separation)
  - Also follows a Poisson distribution

Fallacy of existing models for modeling scale-free networks

- In both of the models how vertices connect to other vertices is mostly random and any one vertex to have a significantly higher k than any other vertex is virtually non-existent
- In real world networks a high k do occur. Think of celebrities followed on Instagram or papers cited in academia
- Both assume new vertices aren't being added
  - Real systems the number of vertices N is constantly increasing

Create model based on preferential treatment of vertices

- Know the network connectivity of a vertex is based on $P(k) \sim k^{-\gamma}$
- Account for new vertices being added
- Account for preferential links to popular vertices

Model creation steps

- Starts with a small amount of m_0 vertices
- Each step add a vertex with an amount of edges smaller than previous vertices $m(\leq m_0)$
- Gives a higher chance for a new vertex to attach to a well established vertex
  - The % chance to attach to a vertex is determined by taking the edges for the given vertex and dividing it by the total number of edges
  - $\prod(k_i) = \frac{k_i}{\sum_j k_j}$
- Over time the model had a power law with exponent variable
  - $\gamma_{model} = 2.9 \pm 0.1$

Developing models to prove add new nodes and adding based on preferential treatment is necessary

- Model without preferential treatment
  - Changed the model to show how preferential attachment matters when a network is developing
  - Made the chance for a new vertex to any existing vertex a constant
    - $\prod(\mathrm{k})=\mathrm{const}=\frac{1}{m_0+t-1}$
  - Found that eventually the connectivity of nodes does not follow power-law
- Model without new nodes added
  - Start with N vertices and no edges
    - New vertices are not added
  - Connect vertices with $\prod(k_i) = \frac{k_i}{\sum_j k_j}$ (preferential treatment)
  - Find that the model follows power law at the beginning, but without new nodes eventually does not

Model results

- Top model is the correct power-law scaling of actual
- Bottom results are the results when:
  - A – no new vertices are added. Conforms at first but over time does not
  - B – constant probability to attach to a vertex
  - C – the connectivity of a vertex over time with power law scaling
    - Shows a “rich-get-richer” phenomenon
- Over time the connectivity of a network can be calculated as
  - $P(k) = \frac{2m^2}{k^3}$
  - For the model gives $\gamma = 3$
    - Fits the scale-free distribution
    - Can’t account for intricacies of real-networks
    - Doesn’t account for networks that also remove connections

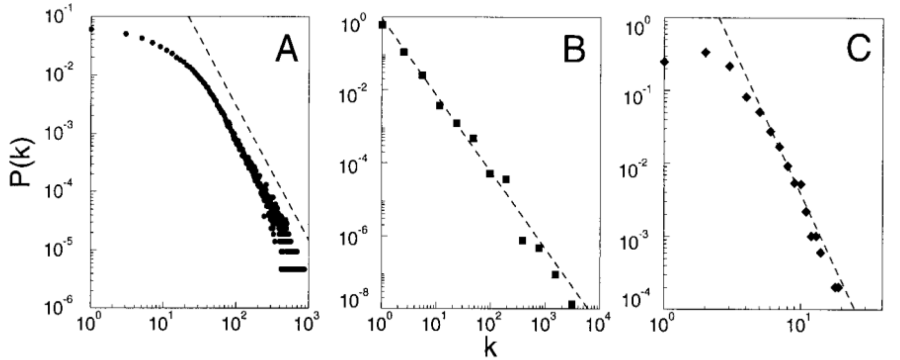


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. (**B**) WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). (**C**) Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

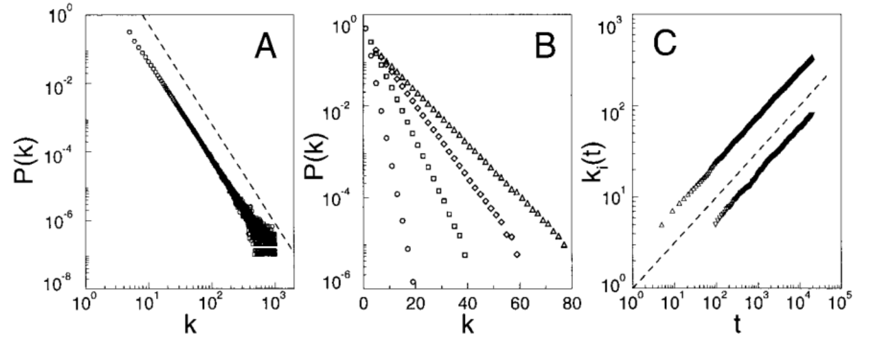


Fig. 2. (**A**) The power-law connectivity distribution at $t = 150{,}000$ (○) and $t = 200{,}000$ (□) as obtained from the model, using $m_0 = m = 5$. The slope of the dashed line is $\gamma = 2.9$. (**B**) The exponential connectivity distribution for model A, in the case of $m_0 = m = 1$ (○), $m_0 = m = 3$ (□), $m_0 = m = 5$ (◇), and $m_0 = m = 7$ (△). (**C**) Time evolution of the connectivity for two vertices added to the system at $t_1 = 5$ and $t_2 = 95$. The dashed line has slope 0.5.

Strength 1

- Gives examples of networks with lots of vertices and relatively small amount of vertices and did analysis of both
- Helped the reader see a good example of the range of the exponent variable as well give the reader a reference for estimating the exponent variable for a system they might be studying
- Also, helped the reader identify exactly which type of networks fall under the scale-free umbrella

Strength 2

- Takes the existing models and explains how they wouldn’t work for scale-free networks
- This allows readers already familiar with the field to quickly see the use of the additional research presented in this paper
- This also shows how their research is necessary for the field, since the scale-free networks at the time were largely not studied

Strength 3

- Creates models which counter their own to demonstrate their model wasn't a fluke
- They created a model to show that adding new vertices is necessary for scale-free networks
- They also created a model to show that preferential treatment is also necessary for scale-free networks
- Added credibility to the paper and the research contributed to the field as a whole

Weakness 1

- Came to the conclusion they set out to without factoring in disconnecting vertices
- They mentioned this in the paper, but did not explore it further
- This somewhat weakens their results since most of these networks have nodes constantly disconnecting and reconnecting
- With their research there’s no way to determine how much disconnecting vertices would factor into their results

Weakness 2

- In addition to running models without a growing N and without preferential treatment, they could have also run the Erdős–Rényi and Watts-Strogatz models with an increasing N.
- Both of the previous models were designed to run with a fixed number of vertices. If run with a growing N the author could further prove that preferential treatment is necessary for scale-free networks.
- Without doing this it is only assumed that the previous models would not work to model growing systems

Weakness 3

- Did not do enough analysis on real systems to show that they use preferential treatment and expanding nodes
- Could have done their own analysis to show definitively that real systems are scale-free because of the reasons that they think
- They prove that their methods will produce a scale-free network, but not that real networks are scale free because of preferential treatment
- Their model has older vertices as being more connected
  - I think they would find in real systems that this wouldn't necessarily be the case

Extension 1

- Analysis of Twitter users
- Analyze followers
- See trends of popular trends over time based off of new followers and links
- See the rise and fall of certain topics/people’s popularity
- Predict future trends based off of this data

Extension 2

- Redo the experiment, but this time include vertices disconnecting from other vertices
- To construct the test they would have to consider why vertices might disconnect to other vertices
- Consider disconnecting an edge from a vertex based on how many other vertices disconnect from that vertex over a period of time.
  - For instance in social networks a mass edge disconnect might occur after someone said an offensive statement or has become inactive for too long
    - It would follow that if lots of vertices disconnect in a short period of time, it would be likely another vertex would disconnect
  - Another example would be in the US electrical grid where disconnects would happen often if a transformer was starting to fail and couldn't support all of its edges
- In addition set up experiments where the disconnects happen randomly, over time, or if no new vertices have connected to a given vertex over a period of time
- This would create a more complete model which would result in more accurate results and a more complete study

References

- Barabási, A., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science,286*(5439), 509-512. doi:10.1126/science.286.5439.509
- Six degrees of separation. (2017, November 07). Retrieved November 19, 2017, from https://en.wikipedia.org/wiki/Six_degrees_of_separation
- Scale-free network. (2017, October 11). Retrieved November 19, 2017, from https://en.wikipedia.org/wiki/Scale-free_network