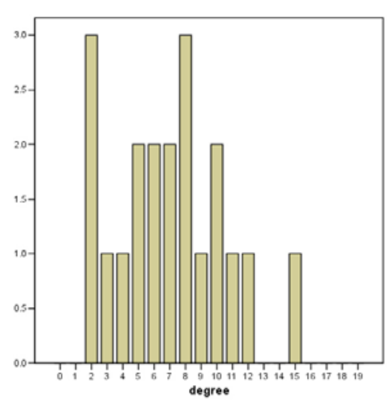
A brief discussion of scale free degree distributions and of the preferential attachment model.

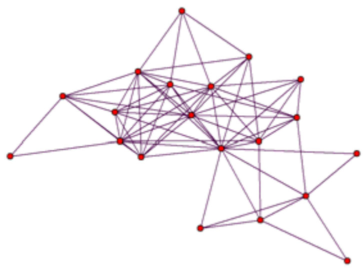
- Degree distributions
  - Scale free degree distributions
- Barabassi's preferential attachment model
- Is everything determined by the degree distribution?
- How well do scale-free degree distributions fit data?

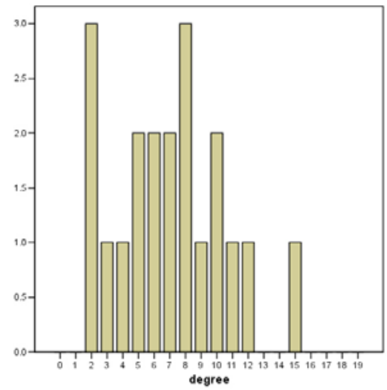
Degree distributions

- The degree distribution represents the frequencies of nodes with a given degree *k*.
- E.g. Degree distribution for the mutual student dataset.



Degree distributions





Probability models for the degree distributions

Let K be the degree of a randomly chosen person in the network. Then a statistical model for the degree distribution is represented by:

$$P(K=k) = f(k)$$

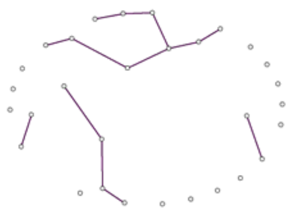
where $f(k)$ is a probability distribution (e.g. a normal distribution).

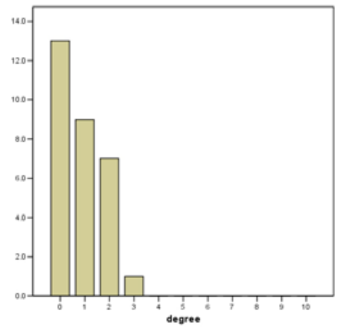
The question is: what is an appropriate $f(k)$?

Degree distributions for simple random graphs

Bernoulli graphs of low density tend to have degree distributions with some positively skew.

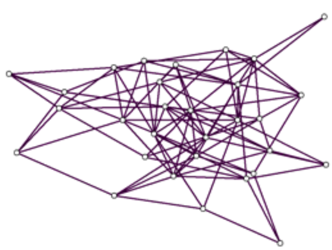
But without having very high degree nodes (“hubs”)

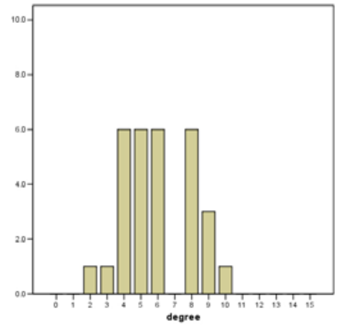




Degree distributions for simple random graphs

Bernoulli graphs of higher density tend to have symmetric degree distributions without skew.





Degree distributions for empirical social networks

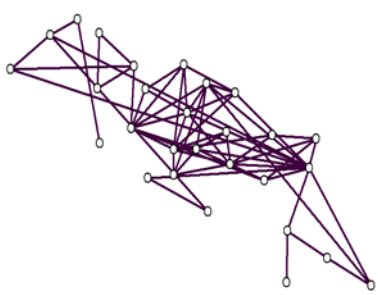
It is common for empirical social networks to have positively skewed degree distributions.

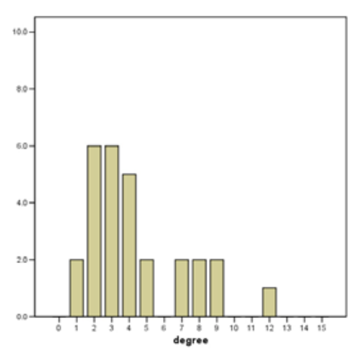
- Sometimes they exhibit outliers in the degree distribution with particularly high degree (Hubs)
- But not always ….

It is often argued that the internet exhibits some important hubs (eg Google, Acrobat).

Empirical example:

Training squad mutual trust network (Pane, 2003)





Skewed degree distribution

Inverse power law distributions: A possible probability model for degree distributions with hubs

Let K be the degree of a randomly chosen person in the network. Then an inverse power law degree distribution is represented by:

$P(K=k)$ is proportional to $k^{-\rho}$ (at least for large k)

where ρ is a scaling parameter (greater than 1).

Notice that log $P(K=k)$ is proportional to $-\rho \log k$

So that plotting an inverse power law degree distribution on a log-log scale results in a linear relationship with a negative slope.

Inverse power law distributions: A possible probability model for degree distributions with hubs

With an inverse power law degree distribution:

$P(K=k)$ proportional to $k^{-\rho}$

the probability of a small degree is high and the probability of a hub is small

- So we see many low degree nodes and few outlying hubs.

Inverse power-law degree distributions have often been called "scale-free" (somewhat loosely) in the physics network literature.

Preferential attachment model: A proposed mechanism for creating scale free degree distributions

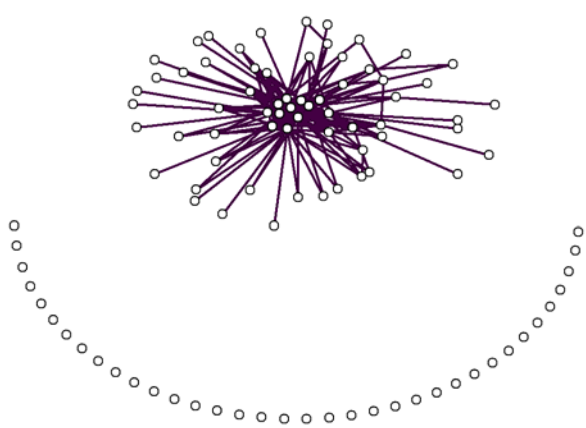
The probability of a tie occurring with a partner depends on the degree of the nodes.

- The popular become more popular (hubs).

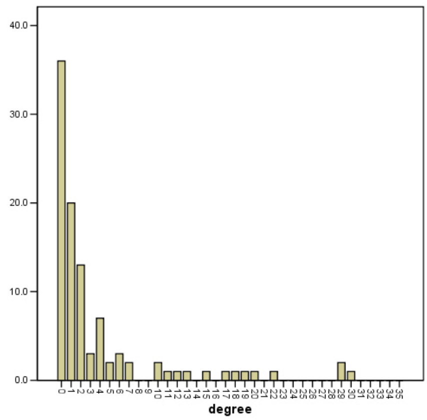
One possible algorithm: Start with a small graph and add a new node at each step. Add a new connection from the new node to an existing node. The probability of the new connection relates to the degree of the existing nodes. Continue until a desired number of steps is reached. (Barabási & Albert, 1999).

Often taken as a model for the growth of the internet.

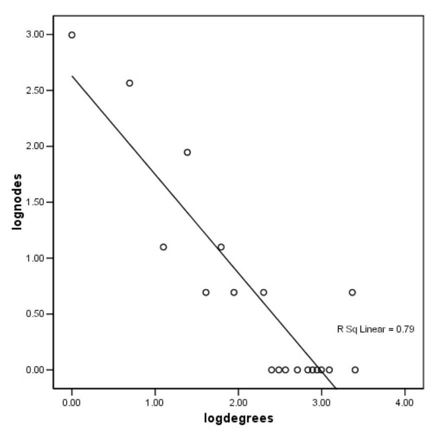
Simulating a scale-free network in Pajek



Simulating a scale-free network in Pajek



Simulating a scale-free network in Pajek



Preferential attachment models

Have been very popular in recent physics network literature (Barabási & Albert, 1999; Barabási, 2002)

- Used to characterise internet growth.

First proposed in relational contexts by Simon (1955)

- statistically, the preferential attachment model is represented by the Yule distribution (Yule, 1924)

Is network structure determined by the degree distribution?

No!

Counterexamples are provided by Snijders & van Duijn (2002) and Robins, Pattison & Woolcock (2005).

The degree distribution does not determine the number of triangles in the graph.

So not all structure is explicable by the degree distribution.

In a network analysis, need to consider whether the degree distribution is exogenous, or an outcome of network processes?

How well do preferential attachment models fit data?

CS

The research area where this has been most hotly debated is in the structure of sexual networks.

- Liljeros et al (2001) argued that many sexual networks were scale-free.

A common practice is to do a log-log plot of the degree distribution and fit a least squares regression line, estimating ρ from the slope of the line.

- Typically zero frequencies are excluded from the data.

"This is a very poor statistical approach … as the assumptions justifying least squares regression do not hold." (Jones & Handcock, 2002)

How well do preferential attachment models fit data?

Handcock & Jones (2003) examined the major sexual network data sets (both men and women), fitting preferential attachment models compared with other plausible statistical models for the degree distribution.

Results: A preferential attachment model was the best fit in only one of six networks.

Conclusions: 1. No single unitary process underlies the formation of sexual networks

2. Behavioural heterogeneity plays an essential role

3. Substantial model uncertainty exists for sexual network degree distributions.

References

Barabási, A-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286,* 509-512.

Barabási, A-L. (2002). *Linked: the new science of networks.* Cambridge, MA: Perseus.

Handcock, M., & Jones, J. (2003). An assessment of preferential attachment as a mechanism for human sexual network formation. *Proceedings of the Royal Society, B*, *270*, 1123-1128.

Jones, J., & Handcock, M. (2002). *An assessment of preferential attachment as a mechanism for human sexual network formation.* Working paper 23, Center for Statistics and the Social Sciences, University of Washington.

Liljeros, F., Edling, C.R., Amaral, L., Stanley, H., & Aberg, Y. (2001). *Nature, 411,* 907-908.

Robins, G.L., Pattison, P.E., & Woolcock, J. (2005). Social networks and small worlds*. American Journal of Sociology*. Forthcoming.

Simon, H. (1955). *Biometrika, 42,* 435-440.

Snijders, T.A.B., & van Duijn, M.A.J. (2002). Conditional maximum likelihood estimation under various specifications of exponential random graph models. In Jan Hagberg (Ed.), *Contributions to social network analysis, information theory, and other topics: A festschrift in honour of Ove Frank* (pp. 117-134). University of Stockholm: Department of Statistics.

Yule, G. (1924). *Philisophical transactions of the Royal Society of London Series B – Biological Sciences, 213,* 21.