Scale Free Network Growth By Ranking

Santo Fortunato, Alessandro Flammini, and Filippo Menczer

Motivation

- Network growth is usually explained through mechanisms that rely on node prestige measures, such as degree or fitness.
- In many real networks, those who create and connect nodes do not know the prestige values of existing nodes but only their ranking by prestige.
- They proposed a criterion of network growth that relies on the ranking of the nodes according to any prestige measure, be it topological or not.
- The resulting network has a scale-free degree distribution when the probability to link a target node is any power-law function of its rank, even when one has only partial information of node ranks.
- Their criterion may explain the frequency and robustness of scale-free degree distributions in real networks, such as the Web graph.

The Model

- First, a prestige ranking criterion is selected.
- At the (t + 1)th iteration, the new node t + 1 is created and new links are set from it to m pre-existing nodes.
- The previous t nodes are ranked according to prestige, and the linking probability p(t + 1 -> j) that node t + 1 be connected to node j depends only on the rank R_j of j:
- $p(t+1 \rightarrow j) = \frac{R_j^{-\alpha}}{\sum_{k=1}^{t} R_k^{-\alpha}}$
- Where α > 0 is a real-valued parameter.
- The linking probability decreases with increasing rank.
- The choice of prestige measure is arbitrary.

Ranking by Age

- If the nodes are sorted by age, from the oldest to the newest, the label of each node coincides with its rank,
- i.e. $R_t = t \forall t$
- In this case, the linking probability is the same as that of the static network model
- We can calculate the number of links that the Rth node will attract since its creation.
- Suppose that the evolution of the network stops when N nodes are created.
- The expected total number k_R^N of links that the Rth node has attracted at the end is
- $$k_R^N = \sum_{t=R+1}^{N} \frac{mR^{-\alpha}}{\sum_{j=1}^{t} j^{-\alpha}}$$

Ranking by Age

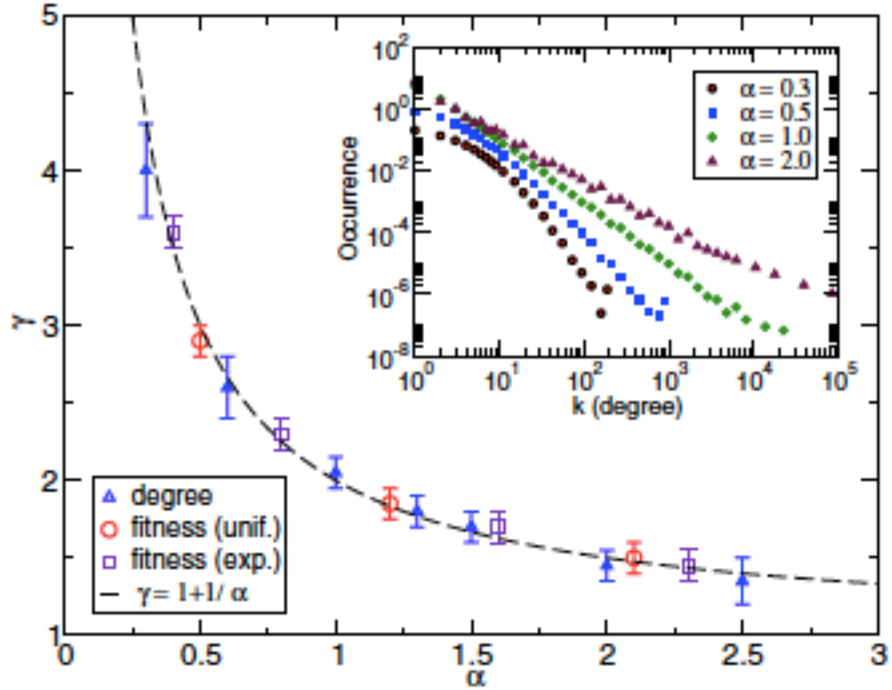
- Approximating with integrals gives that $k_R^N = AR^{-\alpha}$ where A is a function of N
- Knowing k_R^N it is possible to find how many nodes N_k have the same expected number of links k
- The probability p(k,N) that a node of the network has degree k is
- $p(k, N) \approx k^{-(1+1/\alpha)}$
- This shows that the degree distribution of the network follows a power law with exponent
- $\gamma = 1 + 1/\alpha$
- for any value of α

Ranking by Degree

- The number of incoming links of a node represents how many times the node has been selected by its peers.
- For undirected networks, we can equivalently use the degree.
- Nodes with (in-)degree zero, which if present are a problem for the extension of other growth models to directed networks, do not raise an issue here because they have ranks expressed by positive numbers, like all other nodes.
- To see what kind of networks emerge with this new prestige measure, we cannot apply the same derivation as for ranking by age because the degree-based ranking of a node can change over time.
- For a growing network there is a strong correlation between the age of a node and its degree, as older nodes have more chances to receive links.
- Also the ranking of nodes according to degree is quite stable.
- Therefore, we expect the same result as for the ranking by age.

Ranking by Degree

- To verify this expectation, they performed Monte Carlo simulations of the network growth process with the new degree-based ranking.
- The inset shows the degree distributions of four networks, corresponding to various values o the exponent .
- In the logarithmic scale of the plot the tails appear as straight lines, as one would expect for scale-free distributions.
- To verify that the relation between α and the exponent γ of the degree distribution is the one predicted by the model, in the main plot various pairs (α,γ) were compared with the hyperbola and the agreement was clear.



Advantages

- Compared to mechanisms proposed in the past to explain the emergence of scale-free networks, the rank-based model presents three main advantages.
- First, it assumes less information is available to nodes (or node creators); it seems more realistic in many real cases to imagine that the relative importance of items is easier to access than their absolute importance.
- Second, the link attractiveness of nodes is by no means restricted to topology; it can depend on exogenous attributes of the nodes, which makes our model suitable for applications in many different contexts.
- Third, the criterion is more robust in that:
- (i) it naturally extends to directed networks;
- (ii) it leads to long-tailed degree distributions for a broad class of linking probability functions—namely, power laws of rank with any exponent $\alpha>0$,
- (iii) the scale-free degree distribution generated by the model is not affected by limiting the information available to subsets of nodes.

Applications

- The rank-based model is directly applicable to the Web as a special case, if one considers the role of search engines in the discovery of pages.
- When a user submits a query, the search engine ranks the results by various criteria including a topological prestige measure, PageRank, closely correlated with in-degree.
- Users do not know the PageRank of the search hits but observe their ranking and, thus, are more likely to discover and link pages that are ranked near the top.
- Assuming that users tend to link pages that they discover by searching, and that they are aware only of the pages returned by search engines in response to their queries, their model predicts a scale-free distribution of in-degree with exponent $\gamma=2.1$
- This is in perfect agreement with established Web measurements