The Role of Trends in Evolving Networks

Osnat Mokryn  Marcel Blattner  Yuval Shavitt

Modeling complex networks has been the focus of much research for over a decade \cite{1,2}. Preferential attachment (PA) \cite{4} is considered a common explanation to the self organization of evolving networks, suggesting that new nodes prefer to attach to more popular nodes. The PA model results in broad degree distributions, found in many networks, but cannot explain other common properties such as: The growth of nodes arriving late \cite{5} and Clustering (community structure). Here we show that when the tendency of networks to adhere to trends is incorporated into the PA model, it can produce networks with such properties. Namely, in trending networks, newly arriving nodes may become central at random, forming new clusters. In particular, we show that when the network is young it is more susceptible to trends, but even older networks may have trendy new nodes that become central in their structure. Alternatively, networks can be seen as composed of two parts: static, governed by a power law degree distribution, and a dynamic part governed by trends, as we show on Wiki pages. Our results also show that the arrival of trending new nodes not only creates new clusters, but also has an effect on the relative importance and centrality of all other nodes in the network. This can explain a variety of real world networks in economics, social and online networks, and cultural networks. Products popularity, formed by the network of people’s opinions, exhibit these properties. Some lines of products are increasingly susceptible to trends and hence to shifts in popularity, while others are less trendy and hence more stable. We believe that our findings have a big impact on our understanding of real networks.

Trends are a driving force in our lives. From economy to online engagement and popular research subjects, trends govern many aspects \cite{6,7,9}. Trends start locally \cite{10,11}, but only some of them, with time, create enough buzz to accumulate a global effect. We can think, for example, on a giant search engine that has started locally in the academic community at first. Alternatively, there is a popular recent Korean clip, that managed to become the number one hit on YouTube of all times.

We suggest that the preferential attachment process needs to be expanded to include the tendency of new and existing nodes to connect not only to rich nodes, but also to trending nodes. Hence, a node is attractive if it is either rich or trendy, and the relative weight of the trending factor is a characteristic of the network. We would like to accommodate our assumptions that networks have a tendency to trends, yet these trends may start randomly at any point. To this end, we extend the PA mechanism to include this tendency.

**Trending Preferential Attachment (TPA):**

We assume an evolving network where in each step a node is added with $m$ links. Let a network’s tendency to adhere to trends be denoted by $r$. Then, a node with degree $k$ that has acquired $\Delta k$ links in the last step will acquire new links in a monotonically increasing rate that is a function of

$$k + r \cdot \Delta k$$  \hspace{1cm} (1)

The basic PA algorithm relies on random patterns. Once a node collects more incoming links it is more likely to become more popular. TPA builds on a similar idea: it looks at the change in the degree of a node, and amplifies this change by $r$, i.e., by the network’s reaction to trends. The more trendy is the network, the bigger is the effect of small changes. Hence, new nodes are more likely to attach to nodes that either have a high degree or are gaining a momentum in the growth of the number of new links, and hence are trendy. By relying on random local changes and amplifying them, TPA is close in spirit to the PA model, unlike the static popularity model suggested in \cite{5}.

We next introduce a basic Trending Preferential Attachment growth algorithm. Like in the PA model we start with $m_0$ nodes. Then, at each step, a new node with $m \leq m_0$ links is added. The other ends of the links are chosen with a probability that correlates with the node’s importance, denoted by its relative weight $W_i$:

$$\Pi(W_i) = \frac{W_i}{\sum_j W_j}$$  \hspace{1cm} (2)

Where $W_i = k_i + r \cdot \Delta k_i$, and $\Delta k_i$ is the recent growth in the degree of node $i$ degree. In the most simplistic way, $\Delta k_i = \Delta k_i(t) = k_i(t) - k_i(t - 1)$. At each step a node with $m$ links is added to the network, hence, the total weight at time $t$ is $2mt + mr$. The rate at which a node acquires edges is $\partial W_i/\partial t = W_i/(2t + r)$, which gives

$$W_i(t) = m \cdot \left(\frac{r + 2t}{t_i}\right)^{0.5}$$  \hspace{1cm} (3)

Where $t_i$ is the time in which node $i$ was added.

According to the TPA model, the $r$ parameter governs the trendiness of the network. For large $r$, some nodes may incur a sudden speedy growth, while for other nodes the fast growth stops. However, in slow changing networks, where $r$ is very small, the growth depends indeed much more on the degree of the node rather than on current trends. Clearly, when $r \to 0$ then $W_i(t) = k_i(t)$.
In the thermo-dynamic limit, namely when $t \to \infty$, TPA differs from the PA model only by an additive factor:

$$P(W_i < W) = P(t_i > \frac{m^2}{W^2}(r+2t)) = 1-\frac{m^2(r+2t)}{W^2(m_0+t)}$$

Thus, $P(W)$ is given by

$$P(W) = \frac{\partial P(W_i(t) < W)}{\partial W} = \frac{2m^2(r+2t)}{(m_0+t)W^3} \Rightarrow \frac{4m^2}{W^3}$$

Thus, we get $\gamma = -3$, as $W$ is equivalent to $k$ in the PA model.

What we are interested in is the dynamics of young trending networks, namely, when $r >> N$, where $N$ is the number of nodes in the network. When the growth model allows for the addition of one node at each step, $N \sim t$. Consequently, for very trending young networks, we get:

$$\Pi(t) = \frac{W_i(t)}{\sum_j W_j} \rightarrow r^{-0.5}$$

Showing that newly arriving nodes have a similar probability of becoming important as older nodes in the network.

Figure 1 shows the degree of nodes as the function of their arrival time. In the PA model, there is a strong correlation between the arrival time and the degree of a node, yielding the rich gets richer phenomenon. However, as the network’s trendiness $r$ grows, the correlation weakens. For the case of $r = 900000 >> N$, there is no correlation and high degree nodes can arrive almost at every time.

The TPA model also reveals clustering behavior found in many real networks, unlike the PA model. Figure 2 reveals the clustering behavior of networks as a function of $r$. Trending networks reveal high complexity relative to networks with a lower tendency to adhere to trends. Particularly, with increasing $r$ the structure becomes more complex and the number of clusters increases as well.

In real networks with appearing trends, one would expect a dense connected region (group) attached to each hub, whereas the trending hubs themselves are not tightly connected. This is another feature which is not observed in PA model. The growing process does not allow for separated groups or clusters, reflecting fans of a particular trend. However, the TPA model has the ability to generate such patterns. Figure 3 reveals the structure of a network with a very high tendency to trends.

Obviously, while some networks have high trendiness, others may have almost none ($r \to 0$). Let us examine, for example, the word co-occurrence network. The network is formed by the frequency of co-occurrences of words in sentences. If we consider this network in the pre-web era, it is clearly a very slow evolving network, as languages evolved very slowly. New words appear seldom, and words’ popularity is almost stable over periods of decades. Thus, a book written over a period of a few years is a static word network, namely with $r = 0$, and its word degree distribution is close to a power law. Bi et al. made an interesting observation [12] that unlike many other books, the Bible word distribution deviates from a pure Zipf distribution. Indeed, the Bible, which was written over a few centuries, captures the language dynamics that is responsible for the deviation.

Thus, we can see the degree distribution of a network as a superposition of its static behavior, governed by a power-law, and the trending dynamics. To demonstrate this, we investigated the trends in weekly Wiki edit pages created at a single day. The small deviation of the distribution from a power law suggests a low trendiness. We filtered out pages in which there were edits in the last 4 weeks. The results, denoted in Figure 4 show that without the trending pages the edit distribution is closer to a power law.

References

[1] D. Watts and S. Strogatz, “Collective dynamics of small-world networks,” Nature, vol. 393, pp. 440–442, 1998.
[2] R. Albert and A. Barabási, “Statistical mechanics of complex networks,” Reviews of modern physics, vol. 74, no. 1, p. 47, 2002.
[3] M. Newman, A. Barabasi, and D. Watts, The structure and dynamics of networks. Princeton University Press, 2011.
[4] A. Barabási and R. Albert, “Emergence of scaling in random networks,” science, vol. 286, no. 5439, pp. 509–512, 1999.
[5] S. Dorogovtsev, J. Mendes, and A. Samukhin, “Structure of growing networks with preferential linking,” Physical Review Letters, vol. 85, no. 21, pp. 4633–4636, 2000.
[6] E. Denison, Trends in American economic growth, 1929-1982. Brookings Institution Press, 1985.
[7] F. Ng and A. Yeats, “Major trade trends in east asia: what are their implications for regional cooperation and growth?” World Bank Policy Research Working Paper, no. 3084, 2003.
[8] M. Mathioudakis and N. Koudas, “Twittermonitor: trend detection over the twitter stream,” in Proceedings of the 2010 international conference on Management of data. ACM, 2010, pp. 1155–1158.
[9] S. Asur, B. Huberman, G. Szabo, and C. Wang, “Trends in social media: Persistence and decay,” in 5th International AAAI Conference on Weblogs and Social Media, 2011.
[10] N. Koenigstein and Y. Shavitt, “Talent scouting in p2p networks,” Computer Networks, vol. 56, no. 3, pp. 970–982, 23 Feb. 2012.
[11] Y. Dover, J. Goldenberg, and D. Shapira, “Network traces on penetration: Uncovering degree distribution from adoption data,” Marketing Science, vol. 31, no. 4, pp. 689–712, 2012.
[12] Z. Bi, C. Faloutsos, and F. Korn, “The dgx distribution for mining massive, skewed data,” in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2001, pp. 17–26.
Fig. 3: A network with 2000 nodes and \( r = 2 \cdot 10^5 \)
Fig. 1: Node degree as a function of the network trendiness $r$

Fig. 2: A heat map of the block diagonalized Laplacian matrix. The graph produced by the corresponding Laplacian matrix $L = D - A$, where $D$ is the degree diagonal matrix and $A$ is the adjacency matrix. In the diagonalized Laplacian matrix the eigenvalues are ordered in increasing order and hence also the eigenvectors (columns) accordingly.
Fig. 4: Wiki edits per week distribution for pages created at Feb 25th 2002: (a) features the full distribution of all 8752 pages; (b) features the static part containing the non-trending 6024 pages distribution, where the trending pages are filtered out.