On a Pyramid Structure in Social Networks and its Application

Kyoto University *LYU Wenruo
01991974 Kyoto University ZHAO Liang

1. Introduction

Social networks describe the interactions among individuals and other social existences. How to model social networks and use it to explain the observed phenomenon is a fundamental issue in network science. With regard to the average distance among pairs of individuals, Watts et al. ([1]) proposed the WS model and use it to explain the *small-world* phenomenon. Concerning the degree distribution, Barabási et al. ([2]) came up with the BA model to explain the *scale-free* phenomenon. Existing models ([1, 2]) do not consider the connections by spatial reasons. To improve the models, Krioukov et al. proposed a hyperbolic geometric model ([3]).

This study, observing a social *pyramid structure* (see Section 2), proposes a mixed model for social networks which concerns both virtual and spatial interconnections, With a parameter, this model can generate from a pure BA network to a pure spatial network. Empirical studies show that the proposed model can explain real-world social networks better.

2. Pyramid Structure

Let $G = (V, E)$ denote an undirected graph with a set V of n nodes and a set E of m edges. Without loss of generality, we assume that G is simple and connected. Let $\mathrm{dist}_G(u, v)$ denote the distance, i.e., the minimum number of edges needed to reach a node v from a node u in graph G and $\mathit{\Gamma}_G(v; i) = \{u \in V \mid \mathrm{dist}_G(v, u) = i\}$ denote the set of nodes of distance i from a node $v \in V$ in G. Notice that $\mathit{\Gamma}_G(v; 0) = \{v\}$. Let $|\cdot|$ denote the cardinality (size) of a set. The pyramid structure is defined as follows (see Figure 1 for an illustration).

Definition 1 *A rooted tree $T = (r, V_T, E_T)$ with a root node r, a node set $V_T \ni r$ and an edge set E_T is called a pyramid for some constants $\rho_{\min}$ and $\rho_{\max}$ if and only if for all $i = 0, \ldots, h-1$,*

$$1 < \rho_{\min} \leq \frac{|\mathit{\Gamma}_T(r; i+1)|}{|\mathit{\Gamma}_T(r; i)|} \leq \rho_{\max}, \qquad (1)$$

where $h = \max_{v \in V_T}\{\mathrm{dist}_T(r, v)\}$ denotes the height of T.

For any graph G and node r, a pyramid with root r can be found by a Breadth-First Search (BFS) in G and greedily include nodes of distance $0, 1, \cdots$, from r, as far as the (partial) BFS tree is a pyramid. This can be done in linear time. Let us denote it by $T_{\mathrm{BFS}}(r)$. Notice that $T_{\mathrm{BFS}}(r)$ is a maximal pyramid.

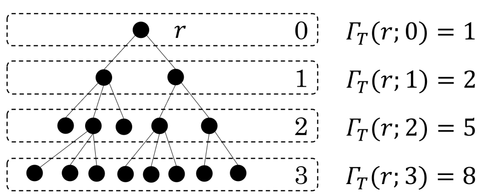
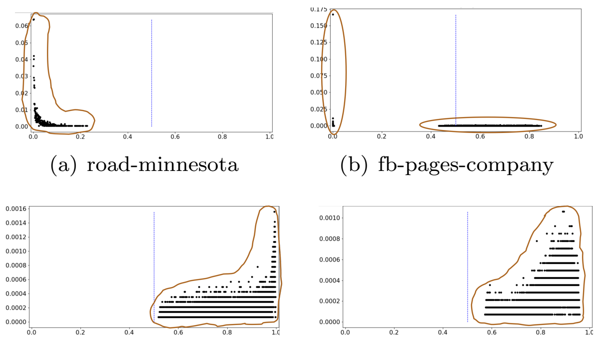


Figure 1: An illustration of a pyramid of height 3 with $\rho_{\min} = 1.6$ and $\rho_{\max} = 2.5$.

3. Observations

Let us focus only on the BFS pyramid structures. Define the normalized size of a pyramid T of size n_T to be $\frac{n_T}{n}$. Figure 2 shows the distributions of the normalized size of pyramids of four types of networks. X-axes show the normalized sizes of pyramids and y-axes show the proportion of pyramids with a certain size. Sub-figures 2(a) and 2(b) are real networks downloaded from [4] where sub-figure 2(a) is a road network and sub-figure 2(b) is a social network. Sub-figures 2(c) and 2(d) are generated by the author.



(a) road-minnesota (b) fb-pages-company (c) BA network (d) WS network

Figure 2: Distributions of normalized sizes of BFS pyramids of four networks.

It can be observed that in road networks, pyramids are usually small, and rarely can a large pyramid[1] appear, while in social networks, most pyramids are large, and there is a big gap between small and large pyramids. This phenomenon can be observed in most of the 127 real social networks on [4]. On the other hand, all pyramids in BA networks and WS networks are large. This observation inspires us to regard real

[1]A pyramid is said large if $n_T > \frac{n}{2}$.

social networks as a mix of road networks and BA/WS models and propose the following mixed generation model.

4. A Mixed Generation Model

We start with an initial graph that consists of several small structures, e.g., star graphs. At every step, we add a new node and link it to existing nodes according to the Preferential Attachment (PA) rule ([2]) or Semi-Random Attachment (SRA) rule iteratively. Let $\deg(v)$ denote the degree of a node v. At step t, these two rules are described as follows.

- PA: A new node v_t with k edges links to k different nodes in $G_{t-1} = (V_{t-1}, E_{t-1})$, where v_t links to $v_i \in V_{t-1}$ with probability $P(v_i) = \frac{\deg(v_i)}{\sum_{u \in V_{t-1}} \deg(u)}$.
- SRA: A new node v_t with 1 edge randomly links to some $v_i \in V_{t-1}$ of which $\deg(v_i) \leq 2$.

We remark that SRA rule is enlightened by observation that the degree of most nodes in road networks is 2.

Given four parameters, N for the desired number of nodes, k for the number of edges of new nodes attached according to PA rule, n_0 for the number of initial stars, and p for the proportion of nodes added according to PA rule except for initial nodes, we can generate a network in a mixed fashion. Note that when $n_0 = 1, p = 1.0$, this model degenerates to a pure BA model, and when $p = 0.0$, it generates a tree + initial stars.

5. Empirical Study

This study employs the proposed model, a pure BA model, and a pure WS model to simulate a real online social network fb-pages-company ([4]). To assess their performances, Hellinger distance (HD, [5]) is exploited to quantify the similarity between the distribution of normalized sizes of BFS pyramids of the real network and the generated network. Given two discrete probability distributions $P = (p_1, \ldots, p_k)$ and $Q = (q_1, \ldots, q_k)$, the Hellinger distance $HD(P, Q)$ is

$$HD(P,Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{k} (\sqrt{p_i} - \sqrt{q_i})^2}. \qquad (2)$$

Let M, BA, WS, and R be subscripts of networks generated by this model, BA network, WS network, and the real network respectively. We ensure n_M, n_{BA}, n_{WS}, and n_R are the same and m_{BA}, m_{WS}, and the maximum m_M are close to m_R. Due to the limited space, we ignore the parameter settings here and only show the result.

For BA networks, the values of HD are between 0.75 and 0.80 (see Sub-figure 2(a) for an example). For WS networks, the values of HD are between 0.60 and 1.00 for different rewiring probabilities ([1]) (see Sub-figure 2(a) for an example). However, by adjusting the parameters of this model, the values of HD for the generated networks can be lower than 0.45, which are distributed in the range of $[0.44, 0.86]$. And we can observe the aforementioned "gap" in the distributions of normalized sizes of BFS pyramids of most of these generated networks (see Figure 3 for an example, the meaning of x-axis and y-axis is the same as that in Figure 2). It indicates that the proposed model can explain the real-world social network better.

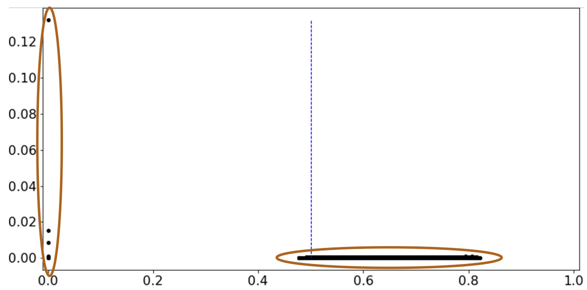


Figure 3: The distribution of normalized sizes of BFS pyramids of an example network generated by the proposed model, which possesses the smallest value of HD (0.4361). Parameters: $N = 14113, k = 3, n_0 = 565, p = 0.60$.

The emergence of pyramid structures in social networks will be analyzed in future work.

Reference

[1] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature, 393*(6684), 440-442.

[2] Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *science, 286*(5439), 509-512.

[3] Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., & Boguná, M. (2010). Hyperbolic geometry of complex networks. *Physical Review E, 82*(3), 036106.

[4] Rossi, R. A. & Ahmed, N. K. (2015). The Network Data Repository with Interactive Graph Analytics and Visualization. *AAAI*, URL: `https://networkrepository.com/index.php`.

[5] Nikulin, M.S. (2001) [1994]. Hellinger distance. *Encyclopedia of Mathematics*. EMS Press