日本オペレーションズ・リサーチ学会
2022年 春季研究発表会

A Study on a Pyramid Structure in Social Networks

Kyoto University *LYU Wenruo
01991974 Kyoto University ZHAO Liang

1. Introduction

Since the late 1990s, driven by the quick emergence of large and reliable network maps [1], quantitative studies on social networks have emerged and rapidly attracted the interest of an increasing number of researchers from many fields. Many models and tools have been developed and applied to study social phenomena. Watts et al. [2] proposed a famous WS model to explain the *small-world* phenomenon, i.e., the average distance is proportional to the logarithm of the number of nodes. Notice that this definition is with respect to *a type of networks*, not for *a given network*.

On the other hand, Barabási et al. [3] declaired that the distribution $P(k)$ of degree k of social networks follows a power law, i.e., $P(k) \sim k^{-\gamma}$ for some constant γ. This is often called the *scale-free* phenomenon, and [3] came up with another famous BA model to explain it.

This study, inspired by the social phenomenon with the same name, proposes a simple *pyramid structure* and uses it to provide a unified and more general view on the aforementioned phenomena of social networks. It is first shown that the size distribution of height-1 pyramids can be used to replace the degree distribution. As an extension, the size distribution of height-2 pyramids shows much different phenomena among networks even if they are with similar scale-free phenomenon. This implies pyramid structure is a more feature-rich framework for studying networks. Furthermore, theoretical and empirical studies on the size distributions of pyramids show that BA and WS models are not sufficient to explain social networks by the size distribution of pyramids and the proportion of large pyramids, where a pyramid is said *large* if its size is greater than the half of the network. It also provides an evidence to the hypothesis on the existence of large pyramids in social networks given by Zhao and Peng 2020 [4] in studying an appropriate size of a parliament. The third contribution of this study is a novel small-world definition for a given network by the existence of large pyramid with high clustering coefficient. Empirical studies support the proposed model.

2. Model

Let $G = (V, E)$ denote a 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, connected, and undirected. Let $\deg(v)$ denote the degree of a node v, and $\mathrm{dist}_G(u, v)$ denote the distance, i.e., the minimum number of edges needed to reach a node v from a node u in graph G. Let $\Gamma_G(v; i) = \{u \in V \mid \mathrm{dist}_G(v, u) = i\}$ denote the set of nodes of distance i from a node $v \in V$ in G. Notice that $\Gamma_G(v; 0) = \{v\}$. Let $|\cdot|$ denote the number of elements in a set, i.e., the size of a set. The proposed model of 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 r, a set V_T of nodes and a set E_T of edges is called a (ρ, ε)-pyramid for some constants ρ and ε iff $\forall i = 0, 1, \ldots, h-1$,*

$$1 < \rho - \varepsilon \leq \frac{|\Gamma_T(r; i+1)|}{|\Gamma_T(r; i)|} \leq \rho + \varepsilon, \tag{1}$$

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

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

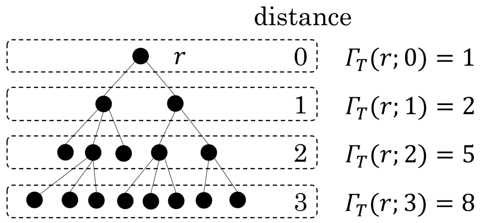


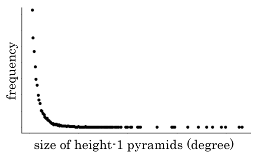
Figure 1: An illustration of a pyramid of height 3 and size 16 with $\rho = 2.05$ and $\varepsilon = 0.45$.

3. Empirical Studies

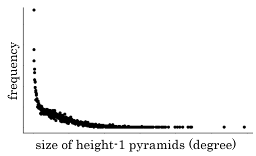
Empirical studies are conducted for 163 networks, 155 of them were downloaded from network repository (`https://networkrepository.com/index.php`) and the rest networks were generated based on BA model and WS model.

Since $\deg(v)$ is one less than the size of the maximal height-1 pyramid with root v if $\deg(v) \geq 2$, degree distribution can be replaced by the size distribution of the maximal height-1 pyramids. As an extension, the size distribution of maximal height-2 pyramids is considered. Comparing two social networks extracted from Facebook, fb-pages-company ($n = 14,114$, $m = 52,126$) and socfb-Harvard1 ($n =$

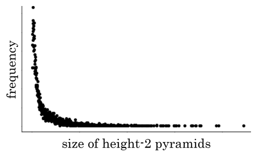
15,086, $m = 824,595$), while both of their size distribution of height-1 pyramids show a power law with $\gamma = 2.30$ and 1.39 respectively, the size distribution of height-2 pyramids for fb-pages-company shows a power law ($\gamma = 1.32$), but socfb-Harvard1 does not. See Figure 2. It indicates that degree distribution is too simple to study social networks, and pyramid structure is more feature-rich.



(a) fb-pages-company, $h = 1$, $\gamma = 2.30$



(b) socfb-Harvard1, $h = 1$, $\gamma = 1.39$



(c) fb-pages-company, $h = 2$, $\gamma = 1.32$

(d) socfb-Harvard1, $h = 2$, not power law

Figure 2: Size distributions of height-1 and 2 pyramids in two social networks.

This study also uses pyramid structure to study the average distance of networks. Noting that for a pyramid T, $h = \Theta(\log n_T)$. Hence $\forall u, v \in V_T, \mathrm{dist}_G(u, v) = O(\log n)$. Therefore, if n_T is large enough ($> n/2$), the small-world phenomenon can be expected.

On the other hand, the sizes of T_{BFS} in social networks are either large or small (≈ 0) while in road networks, pyramids are usually small ($< 0.5n$). However, pyramids in BA networks and WS networks are all large, no small, which show different patterns from social networks. This difference indicates that BA and WS models are not enough to model real social networks.

Finally, we observed the ratio P_{Large} of large T_{BFS} in social networks is between 0.5 and 1 with a 90% confidence interval with an average 0.9374. In contrast, in power networks, $0 \leq P_{\mathrm{Large}} \leq 0.3$, and in road networks, $P_{\mathrm{Large}} \approx 0$. See Table 1. This observation indicates that influence a majority is much easier than expected in a social network.

Based on the above observations, we consider a large pyramid acts as a core structure in a small-world network. Given a pyramid $T = (V_T, E_T)$, we decompose G into two subgraphs G^* and G' induced by V_T and $V \setminus V_T$, respectively. Empirical study results on 3 social networks (see Table 2) show that the difference between $\mathrm{ACC}(G^*)$ and $\mathrm{ACC}(G)$ is not significant, where the *average clustering coefficient*

$$\mathrm{ACC}(G) = \frac{\sum_{v \in V} \mathrm{C}(v)}{n}, \tag{2}$$

and *local clustering coefficient*

$$\mathrm{C}(v) = \frac{|\{(u, w) \in E \mid (v, u) \in E, (v, w) \in E\}|}{\deg(v) \times (\deg(v) - 1)/2} \tag{3}$$

if $\deg(v) \geq 2$, otherwise $\mathrm{C}(v)$ is defined as 0. This observation reveals that large pyramid reflects core structural features of the original network.

Table 1: Comparison of P_{Large} for various networks.

	networks (# samples)	P_{Large}
highly possible	social (127)	$0.50 \sim 1.00$ (90% confidence interval), average ≈ 0.9374
	ecology (2), economic (3), ER (4), BA (4), WS(4)	≈ 1.00
	infrastructure (2)	$0.70 \sim 0.80$
	animal social (1)	$= 1.00$
lowly possible	animal social (2)	≈ 0.25
	power (8)	$0 \sim 0.30$
	brain (1), road (2), cheminformatics (3)	≈ 0

Table 2: Comparison of G and G^* for three social networks.

name	n	$\mathrm{ACC}(G)$	n_T	$\mathrm{ACC}(G^*)$
fb-pages-tvshow	3,892	0.3737	2,456	0.4084
socfb-Haverford76	1,446	0.3230	1,067	0.3088
soc-wiki-Vote	889	0.1528	620	0.2109

Based on the above results, a definition of *generalized small-world network* is given as follows.

Definition 2 *$G = (V, E)$ is said a small-world network if there exists a large pyramid $T = (V_T, E_T)$ in G, and $\mathrm{ACC}(G^*) > c$, where G^* is the subgraph of G induced by V_T and $c > 0$ is a constant.*

We consider this definition is useful in studying social networks.

Reference

[1] Barabási, A. L. (2009). Scale-free networks: a decade and beyond. *science, 325*(5939), 412-413.

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

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

[4] Zhao, L., & Peng, T. (2020). An Allometric Scaling for the Number of Representative Nodes in Social Networks. In: Masuda et al. (eds) *Proceedings of NetSci-X 2020*. Springer.