Physica A 415 (2014) 544–556

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

A generalized theory of preferential linking

Haibo Hu [a,b,*], Jinli Guo [c], Xuan Liu [a], Xiaofan Wang [d]

[a] East China University of Science and Technology, Shanghai 200237, China
[b] City University of Hong Kong, Hong Kong SAR, China
[c] University of Shanghai for Science and Technology, Shanghai 200093, China
[d] Shanghai Jiao Tong University, Shanghai 200240, China

HIGHLIGHTS

- We model an online social network based on preferential linking.
- We propose a generalized model reproducing the evolution of online social networks.
- Unified analytical degree distributions are obtained for diverse preference cases.
- We study the mathematical structure of degree distributions using a novel approach.
- Degree distributions can only be the combinations of finite kinds of functions.

ARTICLE INFO

Article history:
Received 13 July 2013
Received in revised form 1 June 2014
Available online 20 August 2014

Keywords:
Online social network
Preferential linking
Network evolution
Model

ABSTRACT

There are diverse mechanisms driving the evolution of social networks. A key open question dealing with understanding their evolution is: How do various preferential linking mechanisms produce networks with different features? In this paper we first empirically study preferential linking phenomena in an evolving online social network, find and validate the linear preference. We propose an analyzable model which captures the real growth process of the network and reveals the underlying mechanism dominating its evolution. Furthermore based on preferential linking we propose a generalized model reproducing the evolution of online social networks, and present unified analytical results describing network characteristics for 27 preference scenarios. We study the mathematical structure of degree distributions and find that within the framework of preferential linking analytical degree distributions can only be the combinations of finite kinds of functions which are related to rational, logarithmic and inverse tangent functions, and extremely complex network structure will emerge even for very simple sublinear preferential linking. This work not only provides a verifiable origin for the emergence of various network characteristics in social networks, but bridges the micro individuals' behaviors and the global organization of social networks.

1. Introduction

In real life not everyone is equally popular, and in social networks also not everyone possesses the same status or position. Some individuals tend to be at the center of social networks while others remain on the periphery [1,2]. This realization gave rise to the concept of network centrality [3]. Centrality has important effects on the evolution of social networks. Degree centrality, i.e. the number of ties that an actor possesses, has received particular attention maybe due to its computational

* Corresponding author at: East China University of Science and Technology, Shanghai 200237, China. Tel.: +86 21 64252567.
E-mail address: sdhuzi@163.com (H. Hu).

http://dx.doi.org/10.1016/j.physa.2014.08.026
0378-4371/

simplicity. In many real-world social networks, researchers have found that most actors have only a few ties, while a small number have extraordinarily many. For instance it was found that degree distribution is highly skewed in sexual contact networks, where some super-connector actors acquire as many as 1000 partners [4]. Similar patterns also exist in movie co-appearance network, and numerous co-authorship networks in academia [5].

In the past few years, Web 2.0 which is characterized by social collaborative technologies, such as social networking site (SNS), blog, Wiki, video or photo sharing and folksonomy, has attracted much attention of researchers from diverse disciplines [6]. As a fast growing business, many SNSs of different scopes and purposes have emerged on the Web [7], many of which, such as *Facebook* [8], *Renren* [9], *MySpace* [10], *Orkut* [11] and newborn *Google+* [12], are among the most popular sites on the Web. Users of these sites, by establishing friendship relations with other users, can form online social networks (OSNs). Like real-world social networks, in OSNs individual degrees also show obvious heterogeneity. An analysis of the 721 million users on *Facebook* found that a few individuals have 5000 friends (a limit imposed by *Facebook*), more than 26 times as many as the average user's 190 [13].

One important reason social networks develop such a high variance in actors' degrees is that the number of ties an actor possesses affects processes of attachment. Social connections tend to accrue to those who already have them, the consequence of which is that small differences in actor degree compound over time into a distinct cumulative advantage [14]. In OSNs the creation of links between individual users has been studied in a number of contexts [15,16], and is believed to be driven by the principle of preferential attachment (PA), i.e. new users prefer to connect to old users with higher degree. PA is widely recognized as the principal driving force behind the evolution of many growing networks. Besides the PA hypothesis stands as the accepted explanation behind the prevalence of scale-free organization in diverse evolving networks.

The PA phenomena have been studied, qualitatively or quantitatively, in real-world and OSNs. However most of the researches are empirical and lack analyzable models. Besides in network evolution when new users establish friend relationship with old users, or new ties are established between old users, the old users with large degrees are all likely to be preferentially selected. However most previous researches either only focus on PA or combine the two cases into one, overlooking possible preference of varying degrees for link establishment under different scenarios. To date, there are few analytical studies that bridge the micro preferential linking (PL, considering link establishment not only between old users and new users but between old users) and macrostructure of OSNs. A key open question dealing with understanding the evolution of OSNs is: How will the combinations of linear PL, sublinear PL and randomized attachment generate networks with different characteristics? In this paper we exploit not only how linear PL leads to networks with power-law feature (which has been partly studied in the past), but also what network features will result from diverse PL mechanisms, which has not been previously studied.

In the reminder of this paper, after an overview of PA in social networks, we present a detailed case study based on real network dataset, following the procedure of network measurement, modeling, analysis, and model validation. Furthermore considering different forms of PL, we propose a generalized model for the evolution of OSNs, and present analytical results characterizing network features for diverse preference scenarios. At last from the perspective of sociology and economics we analyze the reasons why PL exists in OSNs. We discuss the limitation of the paper and a research framework for better understanding the evolution of OSNs is presented.

2. Preferential attachment

Many social networks have a measured degree distribution $P(k)$ that is either a power-law $P(k) \propto k^{-\gamma}$, or a power-law with an exponential cutoff. Growing models have been proposed to account for these features, most of them being based on some form of PA. Generally PA means that when new nodes join the network linking to the existing nodes, the probability of linking i is an increasing function of the degree k_i of i. Some models assume this function to be linear [17], while in other cases it has been assumed to depend on a different power of k_i [18–20]. In general, we have that the probability $\Pi(k_i)$ with which an edge belonging to a new node connects to an existing node i of degree k_i will be $\Pi(k_i) \propto k_i^{\beta}$, where $\beta \geq 0$. For $\beta = 1$ the rate is linear and the model reduces to the familiar BA model which yields a power-law degree distribution with $\gamma = 3$ [17]. For $\beta < 1$ the PA is sublinear and $P(k)$ is a stretched exponential $P(k) \propto k^{-\gamma} \exp\left[-(b(\gamma)/(1-\gamma))k^{1-\gamma}\right]$, where b is a constant depending on γ [18,19]. Recently some works extended the nonlinear PL model for collaboration networks [21] and rewiring model [22]. The absence of PA is attained in the limit $\beta = 0$, when the attachment rule is independent of degree. The resulting degree distribution in this case is given by $P(k) \propto \exp(-k/m)$ where m is a constant. For $\beta > 1$ a single node gets almost all the edges, with the rest having an exponential distribution of the degrees. Therefore, to know which kind of PA, if any, is at work in a particular growing network, one needs to study empirically networks for which the time at which new nodes entered the network and new edges formed is known.

In recent years some empirical researches have verified the existence of PA rule for social networks, including real-world and online, and exponent β has also been estimated for several networks. However there are some differences as for the functional form of $\Pi(k_i)$. In some cases it appears to be quite close to linear, while in other cases it has been found to be sublinear.

For real-world social networks, Newman studied scientific collaboration networks and found that researchers in physics and biology who already had a large number of collaborators are more likely to accumulate new collaborators in the future [23]. By fitting data he obtained $\beta = 1.04$ for Medline and $\beta = 0.89$ for the Los Alamos Archive. Jeong et al. explored the co-authorship network in the neuroscience field and the Hollywood co-cast actor network, and found that $\beta = 0.79$ for