Commun Nonlinear Sci Numer Simulat 44 (2017) 482–494

Contents lists available at ScienceDirect

Commun Nonlinear Sci Numer Simulat

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

Research paper

Effect of homophily on network formation

Kibae Kim[a], Jörn Altmann[b,*]

[a] Moon Soul Graduate School of Future Strategy, Korea Advanced Institute of Science and Technology, Daejeon, South Korea
[b] Technology Management Economics and Policy Program, College of Engineering, Seoul National University, 1 Gwanak-Ro, Gwanak-Gu, Seoul 08826, South Korea

ARTICLE INFO

Article history:
Received 29 November 2014
Revised 24 February 2016
Accepted 15 August 2016
Available online 8 September 2016

Keywords:
Complex social network evolution
Cumulative degree distribution
Preferential attachment rule and homophily
Empirical data and simulation

ABSTRACT

Although there is much research on network formation based on the preferential attachment rule, the research did not come up with a formula that, on the one hand, can reproduce shapes of cumulative degree distributions of empirical complex networks and, on the other hand, can represent intuitively theories on individual behavior. In this paper, we propose a formula that closes this gap by integrating into the formula for the preferential attachment rule (i.e., a node with higher degree is more likely to gain a new link) a representation of the theory of individual behavior with respect to nodes preferring to connect to other nodes with similar attributes (i.e., homophily). Based on this formula, we simulate the shapes of cumulative degree distributions for different levels of homophily and five different seed networks. Our simulation results suggest that homophily and the preferential attachment rule interact for all five types of seed networks. Surprisingly, the resulting cumulative degree distribution in log-log scale always shifts from a concave shape to a convex shape, as the level of homophily gets larger. Therefore, our formula can explain intuitively why some of the empirical complex networks show a linear cumulative degree distribution in log-log scale while others show either a concave or convex shape. Furthermore, another major finding indicates that homophily makes people of a group richer than people outside this group, which is a surprising and significant finding.

1. Introduction

During the recent decade, empirical research has shown that large complex networks, including the World Wide Web [1], online social networks [2] and Wikipedia [3], can have significantly different network topologies. These topological characteristics are of major interest as they determine the diffusion of information [2,4–6] and the robustness of systems [7,8]. These network topologies have been characterized to have concave, convex, and linear cumulative degree distributions [9,10].

For the formation of these networks, the preferential attachment rule has commonly been accepted. This rule assumes that the probability of adding a new link to a network is proportional to the number of links that a node already has [1]. The resulting cumulative degree distribution has a linear shape, representing a power law function. For reflecting empirical networks more precisely and achieve convex and concave shapes of cumulative degree distributions, prior research has

* Corresponding author. Fax: +1 610 956 4404.
E-mail address: jorn.altmann@acm.org (J. Altmann).

http://dx.doi.org/10.1016/j.cnsns.2016.08.011

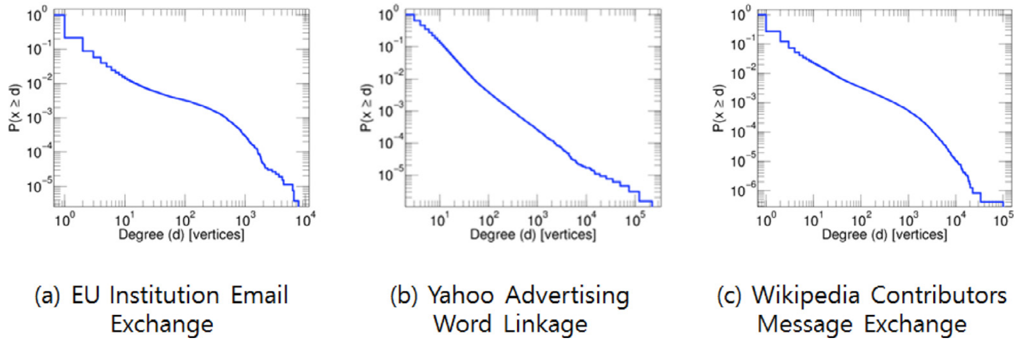

Fig. 1. Empirical data based cumulative degree distributions of EU institution employees email exchanges, Yahoo advertising word linkages, and Wikipedia contributors message exchanges (taken from http://konect.uni-koblenz.de/).

modified the preferential attachment rule into several versions by introducing links generated among existing nodes [11] and by considering preferences according to the geographical distance between nodes [12,13].

The theoretical relationship between actual behavior of agents and the formulas given only covers linear and concave cumulative degree distributions. It is widely accepted that the basic preferential attachment rule represents the effect that the popularity of a node increases with its node degree. The cumulative degree distribution in log-log scale results in a linear curve. It has also been accepted that the addition of links between existing nodes through friends-of-friends effects results in concavely shaped cumulative degree distributions in log-log scale.

However, the theoretical connection between realistic individual behavior of agents in those networks and the concavely shaped cumulative degree distributions in log-log scale is hard to make with the existing formula. The existing formula introduces an exponent to the degree of a node, which is not intuitive, rather a mathematical tool to generate convexly shape curves [12]. Beside this formula, there is no other mathematical formulation that achieves a convex cumulative degree distribution and is an intuitive representation of a theory on individual behavior. Therefore, the following examples of convex cumulative degree distributions in log-log scale, which are shown in Fig. 1, cannot be explained by a formula based on theory on individual's behavior. The examples include the cumulative degree distribution of words used by advertises in their phrases at Yahoo (Yahoo advertising word linkages), which shows a light convex shape, and the curves of the email exchange between employees of a large EU institution (EU institution employees email exchanges) and the message exchange between Wikipedia users through discussion pages (Wikipedia contributors message exchanges). Those degree distributions depict a convex shape at the beginning of the curve and a concave shape at the tail of the curve [10]. Consequently, it can be stated that a formula is needed that can link the convex shape of a cumulative degree distribution with theory on individual behavior intuitively.

Further empirical studies showed that the network evolution is also impacted by attributes of nodes [14,15]. For example, software services belonging to the same company are likely to be combined to develop new services compared to those software services of different companies [16,17]. People in a virtual community are likely to gather according to their gender, age, and proximity [18,19]. This preference of agents to be connected with other agents that share common attributes is called "homophily" [20]. Theoretical research proved that homophily promotes a group of nodes with common properties to be integrated more densely than groups without that property [21]. However, this effect of homophily on the curvature of cumulative degree distributions has not been investigated yet.

Our research objective is to bring together the research on cumulative degree distributions and the research work on homophily. With respect to this research objective, the following research questions rise: How does the formula look like that combines the preferential attachment rule and homophily? What is the cumulative degree distribution of networks that evolve based on homophily and on the preferential attachment rule? Does a seed network topology influence the evolution of the network?

To answer these research questions, we introduce the modified preferential attachment model that is based on [22]. This model combines the homophily preference with the preferential attachment model [11] within a single formula. For the analysis, we investigate the topology of networks evolved through our modified preferential attachment model for different levels of homophily and five different seed networks (i.e., a single dipole network, a multiple dipole network, a ring network, a star network, and a random network).

The result of the comparison is threefold. First, homophily affects the network evolution based on the preferential attachment rule for all types of seed networks that we used. Surprisingly, the resulting cumulative degree distribution always shifts from a concave shape to a convex shape, as the level of homophily gets larger. Second, there is a level of homophily that can compensate concave-shaping effects, making the cumulative degree distribution linear in log-log scale. Third, the link density of the seed networks can cause local distortions on the network evolution, as our results for high-density, random networks show.