



On statistical social behaviors from millions of cell phone users

Shu-Sen Zhang^a, Xun Liang^{a,*}, Yu-Dang Wei^a, Xiao-Ping Zhou^{a,b}, Ji-Chao Zhao^a

^a School of Information, Renmin University of China, Beijing 100872, China

^b School of Electrical & Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

HIGHLIGHTS

- We found two power-law distributions in the study of users' behaviors.
- We studied users' social patterns based on the users' age difference.
- We analyzed the cell phone users' call behaviors, gender differences, and time differences during the call.

ARTICLE INFO

Article history:

Received 29 November 2017

Received in revised form 31 March 2018

Available online xxxx

Keywords:

par Social networks

Cell phone user

Social law

Power-law distribution

Social behavior

ABSTRACT

Limited knowledge on the social laws for cell phone users is known due to the lack of public call data. In this paper, we collected data regarding more than 20 million users of the China Telecom Corporation and analyzed the power-law distribution and their social behavior laws. Two power-law distributions that exist in the distribution of degree and user contacts are found in the study of the users' behaviors, and we analyzed the conclusions drawn from the study. Also, we studied the users' social patterns based on the users' age difference, and analyzed user's social behaviors from the user's age, gender, social scope, time and other aspects.

© 2018 Published by Elsevier B.V.

1. Introduction

Social networking is a relational structure formed by the connection between different users, and an extension of the user's real-world social relationships through the network [1]. Amongst users, these communication networks break the limitations of traditional social networks in relation to individual time and space, and make available “online” interaction at any time. The rapid development and widespread existence of wireless technologies and mobile devices make the cell phone one of the most widely used communication tools possible. Telephone communication is a form of contact that people use ubiquitously. Social networks formed through communication networks will be referred to here as communication social networks. This includes the network formed by people's address books (or contact lists), the network of people they call, and their short messaging network. In fact, people in the social process will generate a lot of social data, which can serve as an important resource for understanding users' social behavior. For instance, the easily accessible digital records of behavior, Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious, intelligence and so on [2]. Compared to other social networks, such as Facebook, Twitter, microblogs and other online social networks, the communication social network more closely reflects people's

* Corresponding author.

E-mail addresses: zss2446@ruc.edu.cn (S.-S. Zhang), xliang@ruc.edu.cn (X. Liang), 1051927683@qq.com (Y.-D. Wei), zhouxiaoping@bucea.edu.cn (X.-P. Zhou), jichaozhao@126.com (J.-C. Zhao).

Table 1
Experimental dataset.

Datasets	Phone users (ten thousand)	Data records (one hundred million)	Networks
MPC	2290	1.7	MPC
CDRs	3955	48.9	CDRs
SMS	2420	5.6	SMS

social situation in real life. Therefore, from the perspective of social network research and human social behavior analysis, this social network has certain research value and significance.

In fact, the study of social networks is closely related to the complex networks [1], and it is the result of combining complex networks with real society. In 1957, Anatol Rapoport's mathematical model emphasized the importance of the distribution of degree in social networks [3]. In 1969, Milgram, a professor of sociology at Harvard University, found that the connection between any two people in the real world can be achieved by only six people on average, known as the theory of "Six Degrees of Separation" [4]. In the 1990s, Watts [5] and others published a "small world" network model which revealed the prevalence of small world properties in the network. In addition, in 1996, based on the study of ape's intelligence and related social networks, Robin Dunbar [6], a British anthropologist, inferred that the size of individual social networks that allow human beings to maintain for human cognitive constraint is 148, which is about 150 people. This is the famous "Dunbar number", also known as 150 rules. Analyzing the feature of social networks is an important way to study social networks. However, there is problem with having to use small datasets in previous studies, leading to limited conclusions [1].

Whilst many researchers try to study communication social network's features and social behaviors, information security, privacy protection and other restrictions, can make the data acquisition of cell phone difficult. Therefore, there have been few studies on communication social network's features and social behaviors. In the past, research about social networks using cell phone data has mainly focused on the network structure, for instance, its topological properties [7,8], or the prediction of social relationships between links or users [9,10]. At present, there is still a lack of social behavior research based on user attributes analysis in communication social networks. In fact, analyzing users' social behavior is also of potentially great practical and commercial value. For both telecommunication operators and mobile Internet operators, information about users derived from their social behavior can has important reference value for the realization of personalized services and direct marketing. Telecom operators, for instance, can precisely target different types of users to improve their experience of using cell phones, thereby improving customer retention and efficiency. A range of services can be envisaged here, including personalized recommendation, targeted advertising [11], etc. Current research regarding the social behavior of users is primarily focused on users of virtual social networks, such as Twitter, or microblogs. Research on users in real social networks typically draws on traditional methods, such as questionnaires.

The communication network dataset we shall be using comes from the China Telecom Corp. It consists of three parts: Mobile Phone Contacts (MPC) data, Call Data Records (CDRs) data and Short Messaging Service (SMS) data. The number of users is more than 20 million. Specific details are given in Table 1. These three datasets not only have call data records information, but also rich personal information. So, the communication social network can truly reflect people's social patterns and behaviors in real life. In addition, in order to maintain the privacy of this data and protect information security, the dataset we are using here has been fully anonymized.

In this paper, our focus is upon the power-law distribution features and user's social behavior laws in communication social networks. Among them, the study of user's social behavior laws includes the analysis of users' social patterns and social behaviors.

2. Analysis of power-law distribution

In 1999, Barabási [12] revealed the prevalence of power-law distribution in networks. Most nodes in networks have fewer links, and these nodes with a large number of links in networks usually accounted for a smaller proportion. Thus, the Poisson distribution, which is conventionally considered in networks, is changed. And they call this kind of network which follows the power-law distribution the scale-free network, as shown in Fig. 1.

2.1. Power-law degree distribution

In social networks, the degree distribution can be defined as the probability distribution of the k degrees. And the node's degree can directly reflect the position or influence of users in social networks. The greater the node's degree is, the more important the node is. According to the direction, we analyzed the communication social network from three aspects, the degree (no direction), in degree, and out degree, as shown in Fig. 2. We found that the degree distribution follows the power-law distribution and has a long tail. The power-law indices γ are 3.19, 3.26, and 1.75 respectively. This "long tail" distribution shows that the breadth (range) of the vast majority of users in social networks is constant. At the same time, there are very few users with a higher contact breadth.

Download English Version:

<https://daneshyari.com/en/article/7374654>

Download Persian Version:

<https://daneshyari.com/article/7374654>

[Daneshyari.com](https://daneshyari.com)