An Experimental Study of Large-Scale Mobile Social Network

Zheng-Bin Dong[1] Guo-Jie Song[1] Kun-Qing Xie[1] Jing-Yao Wang[2]
[1] Key Laboratory of Machine Perception, Ministry of Education, Peking University
{dongzhengbin,gjsong}@cis.pku.edu.cn
[2] China Mobile Group HeiLongJiang Co., Ltd
{kunqing}@cis.pku.edu.cn

ABSTRACT

Mobile social network is a typical social network where one or more individuals of similar interests or commonalities, conversing and connecting with one another using the mobile phone. Our works in this paper focus on the experimental study for this kind of social network with the support of large-scale real mobile call data. The main contributions can be summarized as three-fold: *firstly*, a large-scale real mobile phone call log of one city has been extracted from a mobile phone carrier in China to construct mobile social network; *secondly*, common features of traditional social networks, such as power law distribution and small diameter etc, have been experimented, with which we confirm that the mobile social network is a typical scale-free network and has small-world phenomenon; *lastly*, different from traditional analytical methods, important properties of the actors, such as gender and age, have been introduced into our experiments with some interesting findings about human behavior, for example, the middle-age people are more active than the young and old people, and the female is unusual more active than the male while in the old age.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences–*Sociology*; G.2.2 [**Mathematics of Computing**]: Graph Theory–*Graph algorithms, Network problems*.

General Terms

Algorithms, Experimentation, Performance, Human Factors

Keywords

Degree distribution, shortest path, diameter, clustering coefficient, betweenness centrality.

1. INTRODUCTION

Mobile phones are becoming increasingly ubiquitous throughout large portions of the world, especially in highly populated urban areas, where mobile phone penetration is almost 100%. Mobile phone call networks extracted from call logs is a social network naturally, where the phone user is the node and there is an edge between them if they have communication at least once. Mobile social network reflects people's interaction in social lives and has some advantages compared to other social network, e.g. co-authorship network, email network and world wide web, to uncover the properties and structure of large-scale social network and human behavior in social lives. However, because of the privacy problem, it is very difficult to obtain the mobile phone call data and then there is no much scientific research on this valuable data.

In this paper we construct the mobile social network of a city in China and analyze the statistic properties and structure of the network by using some important network metrics.

2. MOBILE SOCIAL NETWORK

We extract the mobile social network from one month call logs of a city in China. We get 2,590,361 phone users who have two attributes: *age* (from 15 to 70) and $gender$, which are computed from the Chinese Resident Identification Number. We use an undirected binary graph to represent the mobile social network, where the nodes are users and the edges are relationships. A show of the mobile social network by using Pajek [1] is in Figure 1(f), in which the triangle represents the male while ellipse represents the female and different colors represent the different age brackets.

The mobile social network with 2,590,361 nodes is not a connected graph, which has 56,601 connected subgraph and the largest connected subgraph has 2,532,298 nodes – the percentage is 97.76%. The rest analysis of this paper will focus on the largest connected subgraph.

3. EXPERIMENTS AND RESULTS

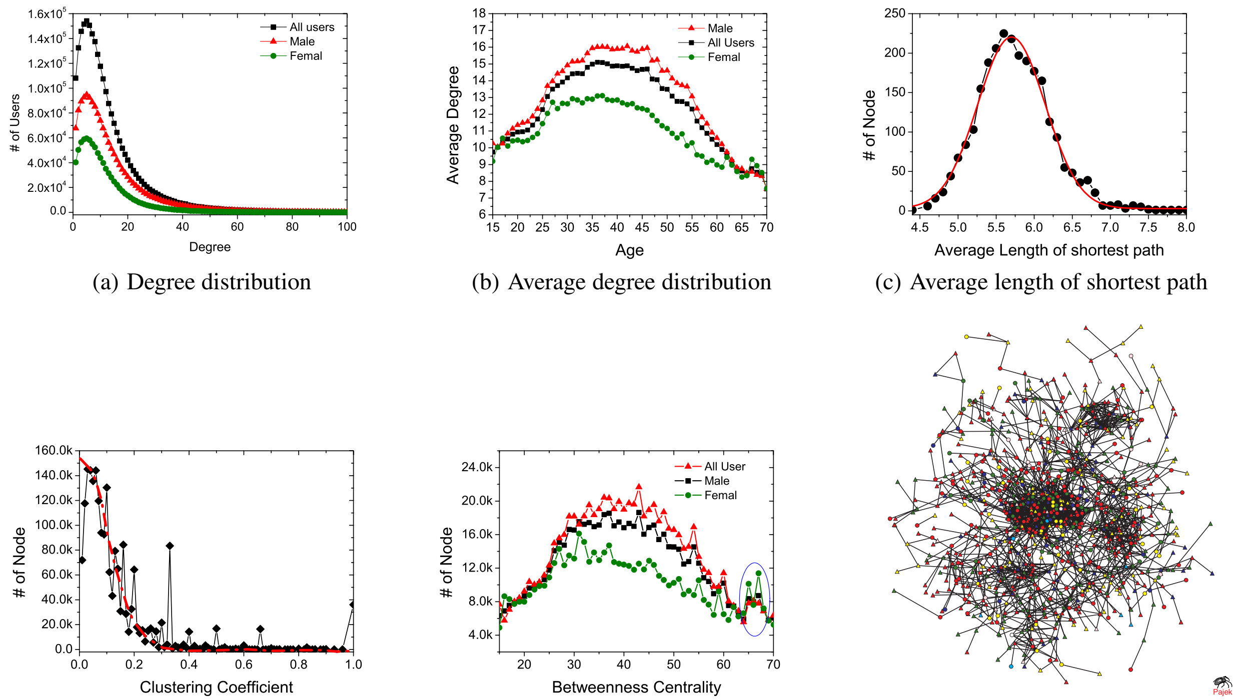
3.1 Degree Distribution

The degree of the node in mobile social network represents the number of friends the user has. So in certain aspect the degree indicates how important or active the user is. The degree distribution of the mobile social network is illustrated in Figure 1(a) and the average degree of the different age and gender is illustrated in Figure 1(b).

From Figure 1(a), we find that the degree distribution almost fits the power law distribution like other social networks, e.g. email network, co-authorship network, but the main difference is that the curve has a peak at the degree five. We believe that this is not a noise but a true difference, because every people have some certain friends to contact in social lives, so the peak is not at degree one like other social networks.

From Figure 1(b), we find some very interesting results about human behavior: 1) the middle-age people are more active than the young and old people. The distribution of the average degree reflects a person's behavior in the whole life: begin entering the society from about 20 years old; become very active in middle age (about 40 years old); and decay in the old age (after 60 years old); 2) the male is always more active than the female in the social network except in the old age, which is very interesting and

WWW 2009, April 20–24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.



(a) Degree distribution (b) Average degree distribution (c) Average length of shortest path

(d) Clustering Coefficient (e) Betweenness Centrality (f) The mobile social network

Figure 1: Our experimental studies of mobile social networks on real mobile phone call logs

conflicts with our traditional thought. We think this unusual result may be explained by Fischer and Oliker' theory [2], they suggest that age and lifestyle stage account for network differences, with young married men having larger networks than their wives, and the reverse being true for older married couples.

3.2 Shortest Paths and Diameter

A shortest path between two nodes is a minimal length path between them. The diameter is length of the longest shortest path. It is impossible in practice to obtain the exact value of diameter in a large-scale network because of the too high computational cost. So in this paper we approximate the bounds for the diameter by using the algorithms proposed in [4]. We obtain that the lower bound is 15 and the upper bound is 17.

We compute the average shortest path by sampling 2537 nodes in the network in Figure 1(c) and the overall average shortest path is 5.75, which proves the famous Six Degree of Separation Theory proposed by Stanley Milgram from another point of view.

3.3 Clustering Coefficient

The clustering coefficient measures the extent of the inter-connectivity between the neighbors of a node. If the neighborhood is fully connected, the value is 1 and a value close to 0 means that there are hardly any connections in the neighborhood.

The distribution of clustering coefficient versus nodes number is plotted in Figure 1(d), from which we can conclude that the distribution fits power law distribution and the mobile social network is sparse as a whole but very dense in some local region.

3.4 Betweenness Centrality

Betweenness is the most widely used metric to measure the important of a node in a network. Vertices that occur on many shortest paths between other vertices have higher betweenness.

It is very expensive to obtain the exact value of betweenness in large-scale network, so we use the sampling methods proposed in [3]. The distribution with different age and gender is plotted in Figure 1(e). Because of the sampling the curve is fluctuated very badly, but we can conclude the similar results about human behavior of the different age and gender as Figure 1(b).

4. CONCULUSION

In this paper, we analyze the statistical properties and structure of a real large-scale mobile social network by using some important network metrics from the view of users' age and gender. We find that mobile social network is a typical scale-free network and small-world network and also find some very interesting results about the human behavior of different age and gender in the social network.

This work is supported by the National Natural Science Foundation of China under Grant No.60703066 and No.60874082.

5. REFERENCES

[1] Batagelj, V., and Mrvar, A.: Pajek. Program for large Network Analysis. Connections, 1998.

[2] Fischer, C.S., and Oliker, S.J.: A research note on friendship, gender and the life cycle. Social Forces, 62,124-133, 1983.

[3] Geisberg, R., Sanders, P., and Scultes, D.: Better approximation of betweenness centrality. In Proceedings of the 10th workshop on Algorithm Engineering and Experiments (ALENEX'08).SIAM, 2008.

[4] Magnien, C., Latapy, M., and Habib, M.: Fast computation of empirically tight bounds for the diameter of massive graphs, http://www-rp.lip6.fr/ latapy/Diameter/, 2007.