Social Networks 47 (2016) 151–155

Contents lists available at ScienceDirect

Social Networks

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

Calling Dunbar's numbers

P. Mac Carron [a,*], K. Kaski [b], R. Dunbar [a,b]

[a] SENRG, Department of Experimental Psychology, University of Oxford, OX1 3UD United Kingdom
[b] Department of Computer Science, Aalto University School of Science, P.O. Box 15500, Espoo, Finland

ARTICLE INFO

Article history:

Keywords:
Social
Brain
Hypothesis
Communication
Ego
Networks

ABSTRACT

The social brain hypothesis predicts that humans have an average of about 150 relationships at any given time. Within this 150, there are layers of friends of an ego, where the number of friends in a layer increases as the emotional closeness decreases. Here we analyse a mobile phone dataset, firstly, to ascertain whether layers of friends can be identified based on call frequency. We then apply different clustering algorithms to break the call frequency of egos into clusters and compare the number of alters in each cluster with the layer size predicted by the social brain hypothesis. In this dataset we find strong evidence for the existence of a layered structure. The clustering yields results that match well with previous studies for the innermost and outermost layers, but for layers in between we observe large variability.

1. Introduction

In recent years the availability of communication data has allowed us to analyse the nature of human relationships and interactions on a much larger scale than previously available (see, for example, Onnela et al., 2007). Although modes of communication have changed however, our brain sizes have not, and it is suggested there is a cognitive constraint on the number of face-to-face social interactions one may have (Dunbar, 1993; Roberts et al., 2009). This constraint fits in a broad sense with the 'social brain hypothesis' which argues that the evolution of primate brains was driven by the need to maintain increasingly large social groups (Humphrey, 1976; Dunbar, 1992, 1998; Barton and Dunbar, 1997).

Individuals do not give equal weight to each relationship and evidence from the social brain hypothesis suggests that ego networks are structured into a sequence of layers with the size of each layer increasing as emotional closeness decreases (Dunbar, 1998; Hill and Dunbar, 2003). The mean number of friends in each has been found to be around 5, 15, 50 and 150 in the cumulative layers (i.e. on average 10 people in the second layer to make a total of 15) (Zhou et al., 2005; Hamilton et al., 2007). Beyond this there are even larger groupings suggested at 500 and 1500 (Dunbar, 1993; Zhou et al., 2005).

Recently these Dunbar layers have been observed in online social media, such as Facebook and Twitter (Dunbar et al., 2015) and an online computer game (Fuchs et al., 2014). These relationships are temporal, however, and the 150 in particular represents the amount of friends at a given time. If a new friend is made, an old one is most likely dropped, and the strength relationships changes quicker in the outer layers than the inner ones (Sutcliffe et al., 2012; Saramäki et al., 2014). However, other methods for estimating personal network sizes have found numbers larger than the outer Dunbar layer, these studies suggest an average personal network size of around 290 for Americans (Killworth et al., 1984; McCarty et al., 2001).

Here we use a mobile phone call dataset initially to ascertain whether layers of friends are detectable in an offline context. If we find evidence of these layers, we then test if they match the layer sizes previously identified using different clustering algorithms.

A European phone-call dataset over all 12 months of 2007 is used. This has 34.9 million users with almost 6 billion calls. About 6 million of these users are with the company (who provide coverage to approximately 20% of the country's population) for whom we have data on all calls they make.

The call frequency between two individuals represents the strength of a relationship and has been shown to correlate with emotional closeness (Roberts and Dunbar, 2011; Arnaboldi et al., 2013). Saramäki et al. (2014) have also shown that social signatures in cell phone data remain robust over time even with identity changes in the alters.

2. Methods

To eliminate casual calls and business calls, the data are filtered so that only calls are counted if there is at least one reciprocal call between the two users.

* Corresponding author.
E-mail address: padraig.maccarron@wolfson.ox.ac.uk (P. Mac Carron).

http://dx.doi.org/10.1016/j.socnet.2016.06.003

People vary in the extent to which they use their phones, with some using it as a regular means of communication with family and friends, and others using it only for social emergencies or to arrange meetings. While the former are likely to provide a full coverage of their social network, the latter won't. To avoid this kind of under-reporting, we censored the dataset so as to include only those individuals with a minimum number of alters. Since the average number of alters at a given time in personal, or ego-centric, networks is 150, with a natural range of approximately 100–250 (Hill and Dunbar, 2003; Zhou et al., 2005; Roberts et al., 2009), we set a value of 100 alters as the minimum cut-off. By doing so, we aimed to have a more complete distribution of actual ego networks, while not biasing against individuals who have naturally small networks. After this we lower the cut-off to 50 alters to observe the results for lower frequency users.

The degree k of an ego represents the number of alters called and the weighted degree w represents the total number of calls an ego makes. The degree distribution p_k and weighted distribution p_w are the fraction of vertices in a network with degree k and weighted degree w, respectively. Note that in empirical networks, the degree distributions are often found to have positive or right skew (Newman, 2003).

In order to estimate the functional forms of degree distribution, the method of Maximum Likelihood Estimators is used (Clauset et al., 2009; Edwards et al., 2007). Here we test different distributions; namely power law, exponential, stretched exponential, Gaussian (or normal) and log-normal distributions, and use the Akaike Information Criteria to select the best model (Akaike, 1974; Burnham and Anderson, 2002).

The data for each user is considered as a one dimensional array which we denote by W such that the minimum possible weight is $w_{\min} = 1$ when an ego calls an alter once. There is no real upper limit (beyond financial or time constraints) to the maximum number of calls a user can make to their preferred alter. In order to compare users, the data for user i is normalised by

$$\widehat{W} = \frac{W_i - W_{i\min}}{W_{i\max} - W_{i\min}}, \tag{1}$$

where W_i is the number of calls made to each alter and $W_{i\min}$ and $W_{i\max}$ are minimum and maximum number of calls they make to any of their alters. This ensures that, for each ego, the strongest interaction with an alter is 1 and the weakest is 0. A first estimate to identify the layers is to plot the probability density of all different weights for all users to ascertain if any pattern exists. A kernel density estimate is applied to the true probability density and the local minima are used to identify clusters (Rosenblatt et al., 1956; Parzen, 1962).

Many methods exist for data clustering, (see, for example, Jain et al., 1999; Gan et al., 2007). The vast majority of these algorithms, however, are for high-dimensional datasets (Jain, 2010). Here, although we are dealing with big data, we seek to break each individual's calls into clusters or layers. Thus we are dealing with one-dimensional clustering for each user, and from this we analyse the average layer sizes.

A common method for one-dimensional clustering is the Jenks natural breaks algorithm (Jenks, 1967). The Jenks algorithm is similar to k-means clustering in one dimension (Khan, 2012). It searches for the minimum distance between data points and the centres of the clusters they belong to as well as for maximum difference between cluster centres themselves. The goodness of fit can be calculated to optimise the number of clusters found (Coulson, 1987). A goodness of fit of 1.0 can only be attained when there is zero within-class variation (often when the number of clusters is the same size as the data). To choose the optimal number of clusters we take a threshold of 0.85 for the goodness of fit as suggested in Coulson (1987).

We also use a Gaussian Mixture Model which assumes that the data are generated from a number of Gaussian distributions (Day, 1969). Naively, we may assume that the layers are made up of Gaussian distributions with their means on the Dunbar numbers. The expectation maximisation algorithm is implemented for this (Dempster et al., 1977) and, again, the Akaike Information Criterion is used to assess the number of clusters in the data.

Another method for clustering the data, used here, is the head/tail breaks (Jiang, 2013). This method was developed for data with heavy-tailed distributions. It splits the data at the mean and taking the head (all values above the mean), it recursively splits each consecutive head at its mean. Our data is heavy tailed (Onnela et al., 2007), with most users calling many people a small number of times but calling their closer friends frequently. An advantage of the head/tail breaks is that the number of clusters is derived naturally from the distribution of the data.

3. Results

Although the mobile phone call dataset we study here contains almost complete data on over 6 million users, only a fraction of these have a degree $k \geq 100$. In order to test the hypothesis of the layers of different levels of emotional closeness, we analyse users

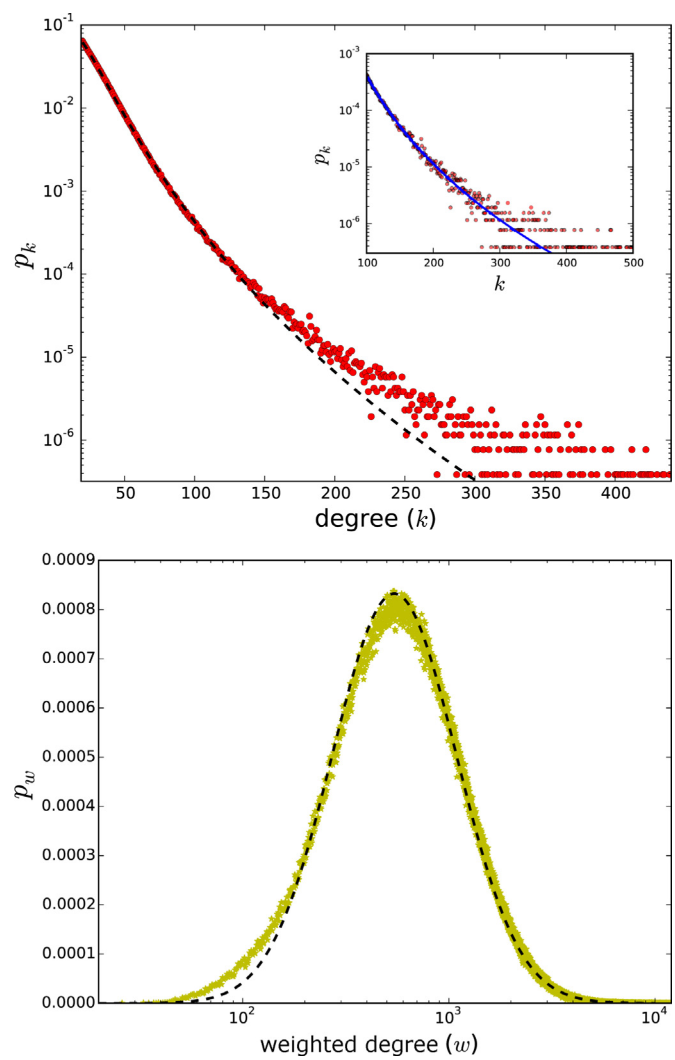

Fig. 1. On the left panel: The degree distribution and a log-normal fit. The inset shows users with degree $k \geq 100$ and a similar fit. On the right panel: The weighted degree distribution is shown, again with a fitted log-normal distribution.