Physica A 393 (2014) 418–426

Contents lists available at ScienceDirect

Physica A

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

The gravitational law of social interaction

Moshe Levy [a,*], Jacob Goldenberg [b]

[a] *The Hebrew University, Jerusalem 91905, Israel*
[b] *IDC, Herzliya, Israel*

HIGHLIGHTS

- We empirically investigate the relation between distance and link probability.
- Four very different social networks are examined.
- We find that the probability decreases as the inverse of the distance squared.
- This is the exact unique distance dependence that ensures network searchability.

ARTICLE INFO

Article history:
Received 7 May 2013
Received in revised form 4 August 2013
Available online 31 August 2013

Keywords:
Social networks
Searchability
Small world
Link probability
Distance dependence
Gravitational law

ABSTRACT

While a great deal is known about the topology of social networks, there is much less agreement about the geographical structure of these networks. The fundamental question in this context is: how does the probability of a social link between two individuals depend on the physical distance between them? While it is clear that the probability decreases with the distance, various studies have found different functional forms for this dependence. The exact form of the distance dependence has crucial implications for network searchability and dynamics: Kleinberg (2000) [15] shows that the small-world property holds if the probability of a social link is a power-law function of the distance with power −2, but not with any other power. We investigate the distance dependence of link probability empirically by analyzing four very different sets of data: Facebook links, data from the electronic version of the Small-World experiment, email messages, and data from detailed personal interviews. All four datasets reveal the same empirical regularity: the probability of a social link is proportional to the inverse of the square of the distance between the two individuals, analogously to the distance dependence of the gravitational force. Thus, it seems that social networks spontaneously converge to the exact unique distance dependence that ensures the Small-World property.

1. Introduction

A great deal has been learned about the topology of networks, and social networks in particular [1–4]. Yet, the geographical structure of these networks is still not fully understood. Different studies report different functional forms of the dependence of the probability of a social link on the distance. The exact distance dependence has important implications for network searchability and for understanding, and possibly influencing, social dynamics, and it is the focus of the present study.

* Corresponding author. Tel.: +972 25883219; fax: +972 25881341.
E-mail addresses: mslm@mscc.huji.ac.il, mslm@huji.ac.il (M. Levy), jacob.goldenberg@mail.huji.ac.il (J. Goldenberg).

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.08.067

Studies that have investigated the geographical structure of social networks typically focus on one particular network, and investigate it in depth. These studies report different results regarding the role of distance in social interaction. For example, in mobile phone communications the probability of a social link depends on the distance r between individuals as $1/r^2$ [5]. Similar results are obtained when analyzing mobile phone communications data at the city level [6]. In contrast, in the LiveJournal network the probability of a link is proportional to $1/r^{1.2} + \varepsilon$, where ε is a background probability that is independent of distance [7]. In email communications between employees within an HP lab the link probability was found to follow $1/r$ [8]. Similar findings are obtained in Ref. [9], where an algorithm to determine a person's location by information about her friends' locations is developed. Refs. [10,11] study location-based social networks, such as Brightkite, Foursquare, and Gowalla, and find that the probability of a social link in these networks generally depends on the distance as a power-law, with power in the range of −0.5 to −1.0. For an excellent review of these and other studies, see Ref. [12]. Do these different results stem from the different social networks analyzed? Are the differences due to the different methodologies employed? (For example, some studies use proxies, such as the average check-in location, to determine a person's "home location", while others use exact addresses or zip-codes.) Perhaps some of the differences are due to the very different distance scales involved? (In the HP email study the maximal distance is 1000 feet, or about 0.3 km, while the maximal distance in the other studies is in the order of hundreds, or even thousands, of km.) The purpose of this study is to analyze a variety of social networks within a unified framework, and to examine whether a general regularity emerges. We investigate the geographical structure of social networks by analyzing four independent and very different sets of data: Facebook links, the social links of participants in the electronic-version Small-World experiment [13], email communication, and data from in-depth personal interviews reported in Ref. [14].

2. Data and results

2.1. Facebook links

"MyPersonality" is a Facebook application that allows users to examine their personality profile. The application also allows users to share their personal data, including their home address zip-code.[1,2]

By examining all pairs of Facebook friends (i.e. pairs of linked users) in the MyPersonality database, we constructed the distribution of link distances. The dataset contains 289,432 users who provided their zip-codes. Of these, we found 531,223 linked pairs with reported zip-codes. The distance between any two linked users was calculated by taking each user's location as the "center of mass" of his zip-code area, and calculating the distance between the two locations.

The probability of a social link is estimated by dividing the *actual* number of existing links at a certain distance by the *potential* number of links at this distance (i.e. the total number of pairs, linked and unlinked, located at this distance one from the other). What is the potential number of social links at a given distance r? For a given person, the potential number of links at distance r is the total number of people located at distance r from her (i.e. in a ring of radius r around her). Under the simplistic assumption of a plane with uniform population density ρ, the number of individuals populating a thin ring of distance r and width Δr around the individual is $2\pi r \Delta r \rho$, i.e. the potential number is proportional to r. At high resolution, the population density is, of course, not uniform – cities are much denser than rural areas. However, when looking at the inter-city resolution, as we do here, the empirical potential number of links is indeed almost perfectly linear in r. Fig. 1A shows the potential number of links as a function of the distance for the MyPersonality database. For every distance range $(r, r + \Delta r)$ we calculate the total number of pairs (linked and unlinked) located at this distance one from the other. The linear fit is very good with $R = 0.996$ ($R^2 = 0.99$).

Panel 1B reports the distribution of *actual* Facebook links as a function of physical distance, on a double-logarithmic scale. The number of links is fitted very closely by a power law. The power is −1.08 with a standard error of 0.03, i.e. the observed distribution of the number of links is very closely approximated by the inverse of the distance, $1/r$, shown by the dashed line with slope −1.

Given that the actual number of links at a distance r is closely approximated by $1/r$, and that the potential number of links is proportional to r, this implies that the probability of a social link is proportional to $\frac{(1/r)}{r} = \frac{1}{r^2}$. This is directly confirmed in Panel C of Fig. 1, which shows the ratio between the actual number of links and the potential number of links directly calculated for each distance interval. The scale is double-logarithmic, and thus the linear fit implies a power-law distribution. The best-fit power is -1.98 ± 0.11. Thus, the probability of a link is fitted very well by a power-law of the distance with power −2 (shown by the dashed line):

$$p(r) \propto \frac{1}{r^2}. \tag{1}$$

[1] We thank David Stidwell, the developer of the "My Personality" application, for this anonymized data.

[2] About 5% of the users reported their zip-codes. The sample of Facebook users who voluntarily report their location is not systematically biased relative to the entire user population [9].