Regularities in the Social Network’s User Distribution by the Number of Mutual Contacts

Ibragim E Suleimenov, Akhat S Bakirov

1 Almaty University of Power Engineering and Telecommunications, Almaty, Kazakhstan;
2 National Academy of Engineering of the Republic of Kazakhstan, Almaty, Kazakhstan;
esenych@yandex.ru; axatmr@mail.ru

ABSTRACT

This paper describes a study in which the pattern of distribution of the number of friends among users of the Vkontakte social network among residents of eight large cities of the CIS (Moscow, St. Petersburg, Almaty, Novosibirsk, Tashkent, Kiev, Yekaterinburg, Pavlodar) was studied. Experimental data show that this distribution is of a similar nature for all selected cities. A semi-empirical model was built, on the basis of which an explicit form of theoretical dependence was obtained, describing the nature of the distribution of the number of mutual contacts (“friends”) of users of social online networks. It is shown that this theoretical dependence agrees with satisfactory accuracy with experimental data for a sufficiently large sample of cities. It is established that the Dunbar number, which is included in the dependencies considered as a control parameter, is a characteristic of the communication environment of each specific city and correlates with the population of the city.

Keywords: Online social networks, Dunbar number, City communication environment.

1 Introduction

The study of the communication space formed by the rapid development of telecommunication technologies is of considerable interest in several aspects.

Thus, a significant part of campaigns to promote products on the market is currently conducted in online social networks [1-5], and such networks are already used not as an auxiliary means of campaigning, but as the main one.

Social online networks connecting a large number of people are an ideal platform for conducting various informational influences of a mass nature, including in the framework of information wars [6-11].

At the same time, modern telecommunication technologies make it possible to receive education online, which creates a number of advantages (access to educational courses of universities around the world, saving travel time, convenience for people with advanced requirements) [12-16]. Massive open online courses (MOOCs) are beginning to gain great popularity [17-21].

The consistent use of social online networks for various purposes requires their adequate theoretical description on a quantitative level.
In this paper, we analyze the distribution of users of social online networks by the number of friends (mutual contacts) and propose a model that describes this distribution.

2 Research method

The data used in the work was collected on the basis of direct collection of information regarding the number of friends of users of social online networks VKontakte.

A random sample of users living in the following cities was created: Almaty, Moscow, Novosibirsk, Kiev, Pavlodar, Yekaterinburg, St. Petersburg, Tashkent. The number of friends each user has was counted. Based on these data, the dependencies presented in Fig. 1 - 8 (dots). The abscissa axis shows the numbers n falling at the center of the intervals on which the axis is divided. The partition scale is selected uneven. The ordinates show the ratio of the total number of users, whose number of friends falls within this interval, to the length of this interval. Such a construction approximately corresponds to the density of distribution of users by the number of friends.

It can be seen that the curves obtained for various cities in which the Runet is actively exploited have a similar character.

In general, the distribution of the number of users by the number of friends can be found on the basis of the following equation:

\[ \frac{dn_j}{dt} = n_{j-1} \sum_{i=0}^{n_j} w_{j-1,i} n_i - n_j \sum_{i=0}^{n_j} w_{j,i} n_i - \frac{n_j}{\tau} \]  

(1)

nj – the number of users whose friends are j, w_{j,i} – the frequency of the formation of a new relationship between people with j and i friends, \( \tau \) – network user lifetime (actual time of using the resource).

Assuming that the functions under consideration with a change in the number j change slowly, we can proceed to the continuous form of writing equation (1). Replacing the discrete numbers j and i with continuous variables x and y, we can write

\[ \frac{d}{dx} n(x) = n(x - \Delta x) \int w(x - \Delta x, y)n(y)dy - n(x) \int w(x, y)n(y)dy - \frac{n(x)}{\tau} \]  

(2)

Applying the Taylor series expansion up to quadratic terms, we have

\[ \frac{d}{dx} n(x) = -n(x) \int \frac{\partial}{\partial x} w(x, y)n(y)dy - \frac{dn}{dx} \int w(x, y)n(y)dy - \frac{1}{\tau} \frac{n(x)}{\Delta x} \]  

(3)

Whence we get that for the stationary case

\[ \frac{df}{dx} = f(x)K(x) \]  

(4)

where

\[ K(x) = \left[ \frac{\frac{\partial}{\partial x} w(x,y)f(y)dy - \frac{1}{\tau}}{\int w(x,y)f(y)dy} \right] \]  

(5)

It is essential that the functions included in the factor for the function f(x) in expression (4) are determined by the entire profile of the distribution under consideration.

Expression (4) allows us to interpret the establishment of an equilibrium distribution of users by the number of friends through a stream directed towards their increase. This, in particular, allows us to use heuristic considerations in finding the function K(x).
These considerations are as follows. The literature currently uses the Dunbar number [22-25], which represents the maximum number of constant social contacts (ties) that one person can support. Initially, this number was obtained by studying the influence of the size and development of the neocortex on the size of a flock of monkeys [22-23], but now it is increasingly used to describe the communication space [26-29].

It can be assumed that the Dunbar number describes the potential causing the flow in a positive direction along the x axis in the model under consideration. If the number of friends of this user is less than the Dunbar number, then we can expect that the derivative in formula (4) will have a positive sign (and vice versa).

The simplest record satisfying this requirement has the form

$$\frac{df}{dx} = (a - x) f^{\alpha} \quad (6)$$

where the factor $\alpha$ takes into account the fractal nature of the communication space.

This nature of the communication space follows from visual considerations. The formation of a relationship between two persons is facilitated if they have a common friend, i.e. there is a certain chain linking them. The inclusion of such chains leads to a power-law dependence in formula (6) and the fractal nature of the communication space as a whole.

Equation (6) is of the first order and is easily integrated; we have

$$-(1 - \alpha) f^{1-\alpha} = (a - x)^2 + C \quad (7)$$

where $C$ is the integration constant.

From here

$$f = \frac{A}{(a-x)^{\alpha} + C} \quad (8)$$

where the normalization factor $A$ is entered in the record,

$$\beta = \frac{1}{\alpha - 1}; \quad \alpha = 1 + \frac{1}{\beta} \quad (9)$$

A comparison of the experimental results with theoretical calculations based on the proposed model is presented in Fig. 1-8.

The theoretical curves (solid lines) in these figures were obtained based on the model described above. It is seen that there is good agreement between the experimental and theoretical data.

The control parameters defining the theoretical curve were determined on the basis of experimental data using the least squares method. The obtained numerical values, using which theoretical curves are plotted, shown in Fig. 1 to 8 are summarized in Table 1.

It can be seen that analogues of the Dunbar number obtained on the basis of the analysis of the presented dependencies fit into the range (100-230) indicated in the literature [22-25]. It is also seen that the analogue of the Dunbar number for Moscow is noticeably higher than for Almaty and Novosibirsk, which could be expected from general considerations.
This means that the analogue of the Dunbar number included in the dependencies used as a parameter can be considered as a characteristic of the communication space for each specific city.

The most significant result is that the fractal dimension of the communication space is the same for all eight cities, and the value is 2/3 with high accuracy. It can be assumed that this value reflects some fundamental features of the formation of the communication space.

Table 1. Empirically determined constant formulas (8) for various settlements and the population of cities.

<table>
<thead>
<tr>
<th>City</th>
<th>A</th>
<th>C</th>
<th>a</th>
<th>$\beta$</th>
<th>Population, thousand people</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almaty</td>
<td>820</td>
<td>5500</td>
<td>95</td>
<td>0.654</td>
<td>1806.6</td>
</tr>
<tr>
<td>Moscow</td>
<td>1290</td>
<td>43000</td>
<td>300</td>
<td>0.657</td>
<td>12506.5</td>
</tr>
<tr>
<td>Novosibirsk</td>
<td>1120</td>
<td>14000</td>
<td>95</td>
<td>0.651</td>
<td>1612.8</td>
</tr>
<tr>
<td>Kiev</td>
<td>1100</td>
<td>20000</td>
<td>130</td>
<td>0.656</td>
<td>2934.5</td>
</tr>
<tr>
<td>Pavlodar</td>
<td>800</td>
<td>3000</td>
<td>110</td>
<td>0.648</td>
<td>334.9</td>
</tr>
<tr>
<td>Yekaterinburg</td>
<td>800</td>
<td>9800</td>
<td>120</td>
<td>0.649</td>
<td>1468.8</td>
</tr>
<tr>
<td>St. Petersburg</td>
<td>716</td>
<td>9350</td>
<td>205</td>
<td>0.65</td>
<td>5351.9</td>
</tr>
<tr>
<td>Tashkent</td>
<td>200</td>
<td>1500</td>
<td>140</td>
<td>0.654</td>
<td>2424.1</td>
</tr>
</tbody>
</table>

It should be noted that the obtained values of the Dunbar number (considered as parameters of the theoretical dependence) correlate with the population of the city (Fig. 9).

This figure shows that the Dunbar number is not an absolute indicator, but significantly depends on the characteristics of the communication environment, mainly the population of the city.

This circumstance cannot be surprising - the larger the city, the greater the number of communications a particular person is involved. He, on average, is better informed, more sociable, etc. Of course, these factors are not the only ones, but we can speak with all certainty about the influence of the nature of the communication environment on the value of the Dunbar number.

3 Conclusion

In conclusion, we note that based on heuristic considerations and literature data on the nature of the Dunbar number, we can propose a simple semi-empirical model that adequately describes the experimental distributions of users of social online networks by the number of friends.

This model is based on the assumption of the fractal nature of the communication space. This assumption seems to be justified also because the fractal dimension obtained for various cities in which the Runet is actively used remains constant with high accuracy.
REFERENCES


