

Semi-empirical Theory of Distributions of Users of Social Online Networks by the Number of Contacts

Ibragim Suleimenov¹ and Akhat Bakirov¹

¹Almaty University of Power Engineering and Telecommunications, Almaty, Republic of Kazakhstan

Abstract. Based on the empirical analysis of contacts in the social network VKontakte, quantitative patterns are established that describe the distribution of users of social networks by the number of contacts (the number of friends). The parameter entering into the differential equation, whose solution is the proposed distribution, makes sense, similar to the Dunbar number and approximately coincides with it in magnitude. The fact that parameters characterizing a given distribution, depending on the size of the communication space in different cities, has been shown. However, the basic parameter that characterizes the fractal dimension of the communication space remains constant for all cities studied.

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 the campaigns for the promotion of goods on the market are currently conducted in social online networks [1-5], and such networks are already being used not as an auxiliary tool for campaigning, but as the main one [6].

Social online networks linking a large number of people are an ideal platform for carrying out various information impacts of a mass nature, including in the context of information wars [7-9].

At the same time, modern telecommunication technologies make it possible to receive online education, which creates a number of advantages (availability of university education courses around the world, saving time for moving, convenience for people with high demands) [10-12]. Mass open online courses (MOOK) are beginning to gain popularity [13-15].

The consistent use of social online networks for various purposes requires their adequate theoretical description at a quantitative level.

In this paper we analyze the distribution of users of social online networks by the number of friends (mutual contacts) and a model describing this distribution is proposed.

2 Experimental

Data used in the work were collected on the basis of direct collection of information on the number of

friends of users of the social online network "VKontakte".

A sample of users residing in the following cities was randomly created: Almaty, Moscow, Kiev, Pavlodar, Novosibirsk. Counted the number of friends each user has. Based on these data, the dependencies shown in Fig. 1 - 5 (points). The abscissa is the number n , which are at the center of the intervals on which the axis is divided. The scale of the partition was chosen unevenly. The ordinate is the ratio of the total number of users, whose number of friends fit into this interval, to the extent of this interval. This construction approximately corresponds to the density of the 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.

3 Discussion

3.1 Theoretical model

The distribution of the number of users by the number of friends in the general case can be found on the basis of the following system of equations

$$\frac{dn_j}{dt} = n_{j-1} \sum_{i=0} w_{j-1,i} n_i - n_j \sum_{i=0} w_{j,i} n_i - \frac{n_j}{\tau}, \quad (1)$$

where n_j is the number of users whose number of friends is j , $w_{j,i}$ is the frequency of the formation of a new connection between persons having friends j and i ,

τ is the lifetime of the network user (the actual time of use of the resource).

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

$$\frac{d}{dt}n(x) = n(x - \Delta x) \int w(x - \Delta x, y)n(y)dy - n(x) \int w(x, y)n(y)dy - \frac{n(x)}{\tau} \quad (2)$$

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

$$\frac{d}{dt} \frac{n(x)}{\Delta x} = -n(x) \int \frac{\partial}{\partial x} w(x, y)n(y)dy - \frac{dn}{dx} \int w(x, y)n(y)dy - \frac{1}{\tau} \frac{n(x)}{\Delta x} \quad (3)$$

From this we obtain that for the stationary case

$$\frac{df}{dx} = f(x)K(x), \quad (4)$$

where

$$K(x) = \frac{\int \frac{\partial}{\partial x} w(x, y)f(y)dy - \frac{1}{\tau}}{\int w(x, y)f(y)dy} \quad (5)$$

It is essential that the functions entering the multiplier for the function $f(x)$ in the 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 according to the number of friends through a stream directed towards their increase. This, in particular, allows us to use heuristic considerations when finding the function $K(x)$.

These considerations are as follows. The literature currently uses the Dunbar number [16-19], which represents the maximum number of permanent social contacts (links) that one person can support. Initially, this number was obtained when studying the influence of the size and development of the neocortex on the size of the monkeys flock [16-17], but now it is increasingly used to describe the communication space [20-21].

It can be assumed that the Dunbar number describes the potential causing the flow in the 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 one can expect that the derivative in formula (4) will have a positive sign (and vice versa).

The simplest entry satisfying this requirement has the form

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

where the factor α takes into account the fractal nature of the communication space.

This character of the communication space follows from obvious considerations. The formation of communication between two persons is facilitated if they have a common friend, i.e. there is a certain chain connecting them. Inclusion of such chains leads to a power-law dependence in formula (6).

Equation (6) is of first order and is easily integrable, we have

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

where C is the integration constant.

Further

$$f = \frac{A}{((a-x)^2 + C)^\beta} \quad (8)$$

where the normalization factor A is introduced into the notation,

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

3.2 Comparison with experimental data

A comparison of the experimental results with the theoretical calculations based on the proposed model is presented in Fig. 1-5.

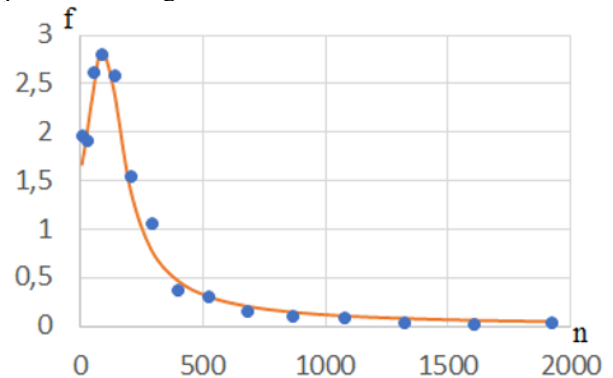


Figure 1. Comparison of experimental data with the theoretical calculation for Almaty, the sample is 3000 users.

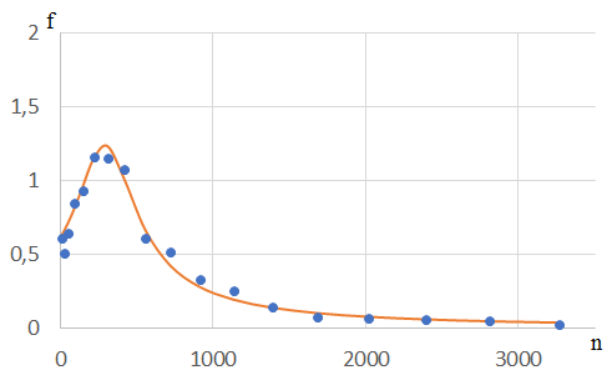


Figure 2. Comparison of experimental data with the theoretical calculation for Moscow, the sample is 3000 users.

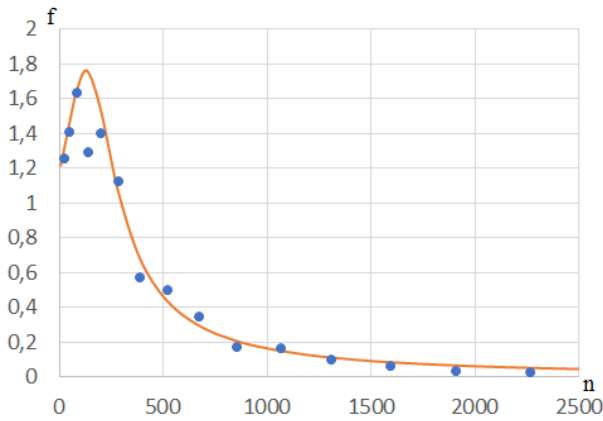


Figure 3. Comparison of experimental data with the theoretical calculation for Kiev, the sample is 3000 users.

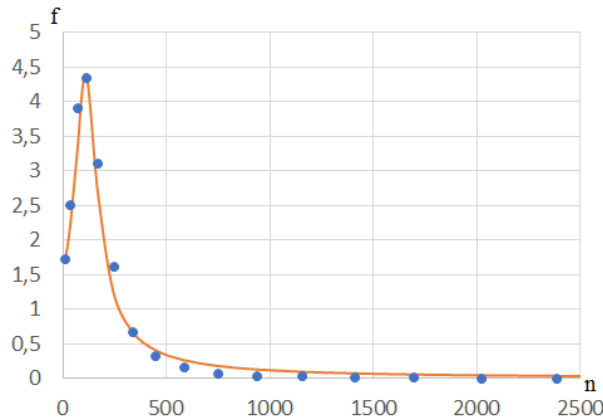


Figure 4. Comparison of experimental data with the theoretical calculation for Pavlodar, the sample is 3000 users.

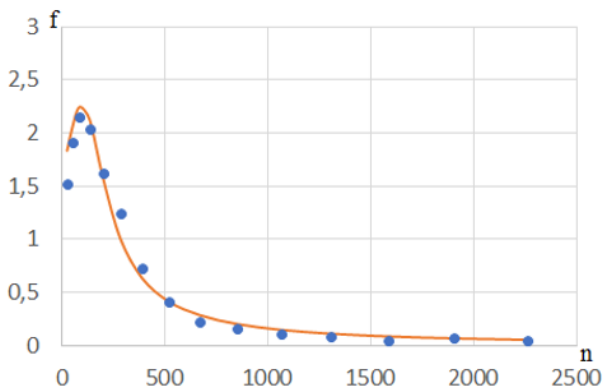


Figure 5. Comparison of experimental data with the theoretical calculation for Novosibirsk, the sample is 3000 users.

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

The control parameters that define the theoretical curve were determined on the basis of experimental data using the method of least squares. The numerical values obtained, which are used to construct the theoretical curves presented in Fig. 1 - 5, 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 dependences, fit within the range (100-230) indicated in the literature data [16-19]. It is also seen that the analog of the Dunbar number for Moscow proves to be much higher than for Almaty and Novosibirsk, which could be expected from general considerations.

The most significant result is that the fractal dimension of the communication space is the same for all three cities, with a value of $2/3$ 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 constants of formula (8) for different cities.

City	A	C	a	β
Almaty	820	6000	95	0,654
Moscow	1290	44000	290	0,657
Kiev	1100	20000	130	0,656
Pavlodar	800	3000	110	0.648
Novosibirsk	1120	14000	95	0,651

4 Conclusions

Thus, on the basis of heuristic considerations and literary data on the nature of the Dunbar number, it is possible to 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 justified also because the fractal dimension obtained for various cities in which Runet is actively used remains with a high accuracy constant. The next step in the study is to expand the sample to more cities, and also to test for another segment of the Internet (not Runet)

References

1. W. Chen, C. Wang, Y. Wang. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. *Scalable influence maximization for prevalent viral marketing in large-scale social networks*. p. 1029-1038 (2010).
2. S. Doyle. Journal of Database Marketing & Customer Strategy Management. *The role of social networks in marketing*. **1**, p. 60-64 (2007).
3. M. Stelzner. Social media examiner. *2014 Social Media Marketing Industry Report*. p. 1-52 (2014).
4. D. Chaffey. Smart Insights: Social Media Marketing. *Global social media research summary 2016* (2016).
5. F. Eggers et al. Journal of small business management. *Technologies that support marketing and market development in SMEs—evidence from social networks*. **2**, p. 270-302 (2017).

6. E. Turban et al. Springer. *Electronic Commerce 2018: A Managerial and Social Networks Perspective* (2017).
7. E. Hunter, P. Pernik International Centre for Defence and Security. *The challenges of hybrid warfare* (2015).
8. J. Klausen Studies in Conflict & Terrorism. *Tweeting the Jihad: Social media networks of Western foreign fighters in Syria and Iraq*. **1**, p. 1-22 (2015).
9. Y. R. Kamalipour, M. Friedrichsen. Introduction: Transformation in Journalism and News Media. *Digital Transformation in a Global World*. p. 1-4 (2017).
10. L. de-Marcos et al. Computers & Education. *An empirical study comparing gamification and social networking on e-learning*. **75**, p. 82-91 (2014).
11. L. de-Marcos et al. Computers in Human Behavior. *Social network analysis of a gamified e-learning course: Small-world phenomenon and network metrics as predictors of academic performance*. **60**, p. 312-321 (2016).
12. P. De Meo et al. Information Sciences. *Combining trust and skills evaluation to form e-Learning classes in online social networks*. **405**, p. 107-122 (2017).
13. C. Bremer, D. Univ. Bibliothek Frankfurt am Main. *Weiß How to analyze participation in a (C) MOOC?* p. 992-1002 (2013).
14. M. Khalil, H. Brunner, M. Ebner. International Journal of Emerging Technologies in Learning (iJET). *Evaluation grid for xMOOCs*. **4**, p. 40-45 (2015).
15. J. F. Colas, P. B. Sloep, M. Garreta-Domingo. The International Review of Research in Open and Distributed Learning. *The effect of multilingual facilitation on active participation in MOOCs*. **4** (2016).
16. R. I. M. Dunbar. Behavioral and brain sciences. *Coevolution of neocortical size, group size and language in humans*. **4**, p. 681-694 (1993).
17. R. Dunbar. Harvard University Press. *Grooming, gossip, and the evolution of language* (1998).
18. R. I. M. Dunbar. Annals of human biology. *The social brain hypothesis and its implications for social evolution*. **5**, p. 562-572 (2009).
19. S. Shultz, R. I. M. Dunbar. I Know What You Are Thinking: Brain Imaging and Mental Privacy. *The social brain hypothesis: an evolutionary perspective on the neurobiology of social behaviour*. p. 13-28 (2012).
20. W. He. Networked Public. *Social Media: Tools and Space for Networked Public Communication*. p. 123-167. (Springer, Berlin, Heidelberg, 2017).
21. S. Banerjee. Interdisciplinary Description of Complex Systems: INDECS. *A biologically inspired model of distributed online communication supporting efficient search and diffusion of innovation*. **1**, p. 10-22 (2016).