DOI: https://doi.org/10.26565/2304-6201-2023-59-01

УДК (UDC) 519.216

Danilevskyi Mykhailo *PhD student; V.N. Karazin Kharkiv National University, Svobody Square, 4, Kharkiv-22, Ukraine, 61022.*
e-mail: m.danilevskiy@gmail.com
https://orcid.org/0009-0000-0030-2218

Yanovsky Volodymyr *Doctor of Physical and Mathematical Sciences, professor; V. N. Karazin Kharkiv National University, sq. Svobody 4, Kharkiv, Ukraine, 61000; Institute of Single Crystals, National Academy of Sciences of Ukraine, Nauki Ave. 60, Kharkiv, Ukraine, 61001*
e-mail: yanovsky@isc.kharkov.ua
https://orcid.org/0000-0003-0461-749X

Modeling and Analyzing the Simplest Network of Telephone Subscribers

Abstract. Dynamic networks such as social, transport and biological networks are widely represented in the modern world. Modeling complex networks as time-varying structures opens up additional opportunities for studying their properties.

Purpose. The goal of the work is to model the simplest dynamic network of telephone subscribers. The main focus is on experiments with the resulting model and studying how the number of subscribers influences the network properties.

Research methods. The work uses the Monte Carlo method of stochastic dynamics of discrete states using time steps of the same length, as well as the methods for constructing computer models, the methods for analyzing the properties of networks, the least squares method and others. The computer model has been developed in Python using the Pandas, Numpy and NetworkX libraries.

Results. The simplest model of a network of telephone subscribers has been designed, where subscribers are connected randomly and disconnected after the phone conversation. In the model, the average daily number of outgoing calls from subscribers is distributed according to the lognormal law. The experiments have been carried out with different numbers of subscribers, but for the same time period. Based on the data obtained from the experiments, we analyzed such network properties as number of connections, density, degree distribution, average clustering coefficient, and average shortest path length.

Conclusions. The developed computer model of the simplest dynamic network of telephone subscribers forms a model similar to a random Erdes-Rényi graph, but the degrees of the vertices or the number of connections between subscribers are distributed according to a lognormal law. The developed computer model can serve as the basis for the development of more complex models and the study of the dynamic properties of such networks.

Keywords: *dynamic complex network, mobile call graph, telephone network, lognormal distribution, degree distribution, network density, clustering coefficient, average shortest path length.*

How to quote: Danilevskyi M., Yanovsky V., “Modeling and Analyzing the Simplest Network of Telephone Subscribers” *Bulletin of V.N. Karazin Kharkiv National University, series Mathematical modeling. Information technology. Automated control systems*, vol. 59, pp. 6-15, 2023. https://doi.org/10.26565/2304-6201-2023-59-01

Як цитувати: Danilevskyi M., Yanovsky V. Modeling and Analyzing the Simplest Network of Telephone Subscribers. *Вісник Харківського національного університету імені В.Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2023. вип. 59. С.6-15. https://doi.org/10.26565/2304-6201-2023-59-01

1. Introduction

Research of the social networks has been developing for many decades. The psychiatrist Jacob Moreno, who became interested in the dynamics of social interactions within groups in the 1930s [1], is usually considered as the founder of the field. “A social network is any network in which nodes represent people and edges represent some form of connection between them, such as friendship” [2]. In this paper, we create a model and study a dynamic network of telephone subscribers or a mobile call graph. This network can be seen as a social network in which contacts are made via a telephone. The peculiarity of the network connected by the telephone calls is that connections are severed after the end of telephone conversation between subscribers. At any moment in time, only pairwise connections exist, but a fully connected network can be observed taking into account the connections that existed over a certain period of time. Thus, in each period there are different configurations of connections between subscribers, and accordingly, the network of telephone subscribers can be considered a dynamic one.

Most networks, including a telephone subscriber network, change over time. However, researchers often view networks as static entities. In some cases, this may be a reasonable approximation, but we can learn a lot by observing, analyzing and modeling network changes over time. For example, in some networks only connections may change and nodes remain permanent, in others, nodes may appear or disappear. One of the approaches to modeling social networks, including networks of telephone subscribers, is random graphs. The random graph model has been introduced in 1950-1960 by Paul Erdős and Alfred Rényi [3]. A random graph is a network model in which the values of certain properties are fixed, but other properties are random. For instance, a graph with a fixed number of nodes and edges is one of the simplest models in which edges between nodes are randomly assigned. This approach is quite primitive and has a number of disadvantages, for example, the lack of correlation between the degrees of neighboring nodes, and the degrees of nodes are distributed according to Poisson's law, which is rarely the case in real networks. Additionally, in real networks structures called communities are formed, in random networks they are absent. The next stage in the development of models of complex networks is the Watts-Strogatz model [4]. This model creates a network with the properties of a "small world" – a high clustering coefficient and a short length of the average shortest path between the vertices of the network. As a result of the analysis of the Internet network topology by Albert and Barabasi [5], it has been discovered that the distribution of vertex degrees follows a power-law. Such networks are called scale-free networks. It has been empirically established that social, communication, biological, citation graphs, Internet links, and other systems could be sufficiently modeled by scale-free graphs. It is believed that the most important characteristics of social networks, in addition to high clustering, short average paths and the presence of communities, include assortative mixing and a wide degree distribution. In [6], the authors have presented a model of such a network in which new vertices are added both to random vertices and to neighboring ones, which leads to implicit preferential joining.

Real world networks can have billions of users, and sometimes it is even impossible to determine their size. In such conditions, network modeling is an effective way to study complex networks. It helps to get an idea about the network structure, understand how the network changes when its parameters change, and also study the processes occurring in networks, for example, the dissemination of information. Dynamic network models can be divided into two groups: deterministic and stochastic. Stochastic models take into account the random nature of network parameters, and therefore are better suited for modeling social networks. For example, in [7], the authors have noted the limitations of the analytical approach and created a stochastic model of the telephone network using the GPSS/H language. Using the created model, the authors have determined the operational characteristics of the existing corporate telephone network and assessed the performance of the designed networks before their installation. In [8], a discrete event simulator for communication networks OSSIm is presented. A brief but comprehensive overview of graph modeling of complex communication networks and their application to social network analysis is provided in [9]. Based on various models of physical networks, the functions for generating artificial social networks have been created in Matlab [10], thus adding a social component to the models of physical networks. Currently, complex networks can be analyzed and modeled by describing models in programming languages, including C++, Python, R, as well as using specialized software packages such as NetworkX [11] and NetworkKit [12]. In works [13-16], the authors analyze the structure and dynamics of social networks in which connections between people are carried out by using a mobile communication network.

In this work, we present a simple model of social network in which connections between people are established via telephone calls. Based on the data obtained after experiments, such network properties as the number of edges, density, degree distribution, average clustering coefficient, and average shortest path length have been analyzed.

2. Modeling dynamic network of telephone subscribers

In this work, we present a model of a social network in which connections between people are established via telephone calls. A network of telephone subscribers is characterized by a certain number of nodes (subscribers) and connections between them. Connections in such network occur when one subscriber contacts another, and the duration of their contact is greater than 0. The number of connections or call attempts is varied for different types of subscribers. For instance, usually people call 5-7 times a day, while sales managers may call 200-300 times a day. As a result, at each moment of time a certain number of pairwise connections exist. These connections cannot be considered a network. Only if we increase the time scale, it is possible to observe a network formed during a given period of time.

As a result of computer modeling, a simple model of a telephone network has been developed. The main simplification is that subscribers choose to contact each other randomly. The number of calls over a period of time is assigned to each subscriber according to the lognormal distribution. While subscribers are in contact, they cannot make or receive calls. The call duration is modeled by the probability of ending a call at a given time.

The input parameters of the model are the following: number of subscribers, parameters of the lognormal distribution for assigning the number of calls per period of time, duration of the experiment.

3. Description of the experiment

Conducting experiments involves simulating the network of telephone subscribers over a certain period of time. The number of subscribers is fixed and set beforehand. Time in this model is discrete, one unit of time is considered to be one minute. If a subscriber calls 5 times a day, then the probability of a call at any minute is defined as 5 divided by 1440 and therefore equals to 0.00347. The experimenter sets the number of subscribers, duration of the experiment, as well as the parameters of the lognormal distribution for modeling number calls per day for each subscriber. The subscriber selects a subscriber randomly according to a uniform distribution law. The probability of a call ending at any given time is 0.99, which determines its duration. If subscribers have established a connection, they are blocked and can neither make nor receive calls. The progress of the experiment is recorded in the table:

Table 1. Experiment data sample

idx	caller_id	callee_id	start_time	finish_time
0	67	22	0	1
1	49	64	2	3
2	44	54	3	4

idx is the record number, caller_id is the caller identifier, callee_id is the callee identifier, start_time is the call start time, finish_time is the call finish, duration is the call duration.

The input data for the experiment are: the number of subscribers is 1000, the duration is 10080 time units (minutes) or 7 days, the parameters of the lognormal distribution correspond to normal subscribers ($\mu = 1.1, \sigma = 1.0$) [17].

4. Study of the obtained data

As a result of running a model, subscribers made 34834 contacts between each other. Based on this data, we examine the distribution of the number of subscribers by the number of calls per day. The chart (Fig.1) shows a histogram where the circles represent fraction of subscribers who called a certain number of times in one day. This distribution can be compared with the distribution obtained in the work based on the results of experiments, where it has been found that such data are distributed according to the lognormal law [17]. Let us examine how well the model confirms with the experimental data.

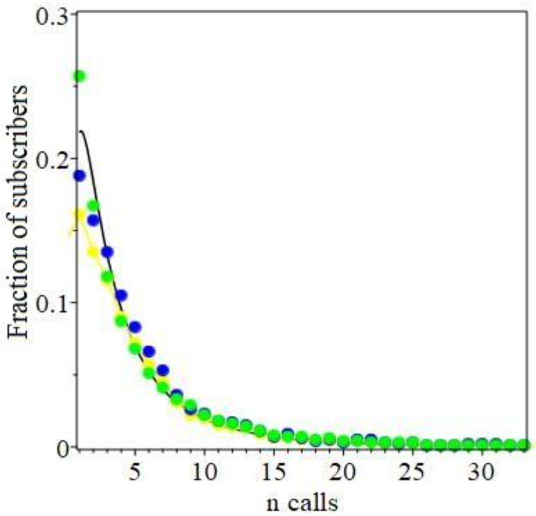

Fig. 1. Distribution of the number of subscribers by the number of calls per day. Green - experimental data from the real network. Simulation data: yellow - all subscribers are included, blue - only those subscribers who made at least one call. The lognormal law is shown as a black curve.

The green line corresponds to experimental data from the telephone network of subscribers [17]. The data obtained after experiments with the model is depicted by the blue and yellow circles. The distribution of subscribers corresponding to the blue line has been calculated without taking into account subscribers who made at least one call during the experiment. The distribution of subscribers corresponding to the yellow line has been calculated taking into account subscribers who did not make a single call during the calls simulations. The lognormal distribution (black curve) demonstrates the agreement between the simulation data and the experimental observations.

The experimental data include subscribers who made at least one call during the experiment, i.e., there are no subscribers without calls. There are such subscribers in the simulated network, which corresponds to the distribution shown by the yellow line. As a result, differences in distributions arise.

As can be seen from Fig. 1, the blue curve – the distribution of subscribers who made at least 1 call – is in good agreement with the experimental data (the green curve). Such subscribers form a network of connections over a certain period. For a more detailed study of the properties of the telephone subscriber network, we will use the data obtained as a result of additional experiments with the developed computer model.

5. Analysis of the experiments with the model

To study the properties of the modeled dynamic network, we have conducted additional experiments with a different number of subscribers, and measured the dependencies of such network properties as well as determined the type of the resulting network. We have conducted 50 experiments in which the initial number of subscribers was 10, and in subsequent experiments the number of subscribers increased by 10 up to 500 subscribers. The graphs of 6 networks out of 50 are shown in Fig. 2. The topology of the resulting networks reflects the random nature of the choice of subscribers.

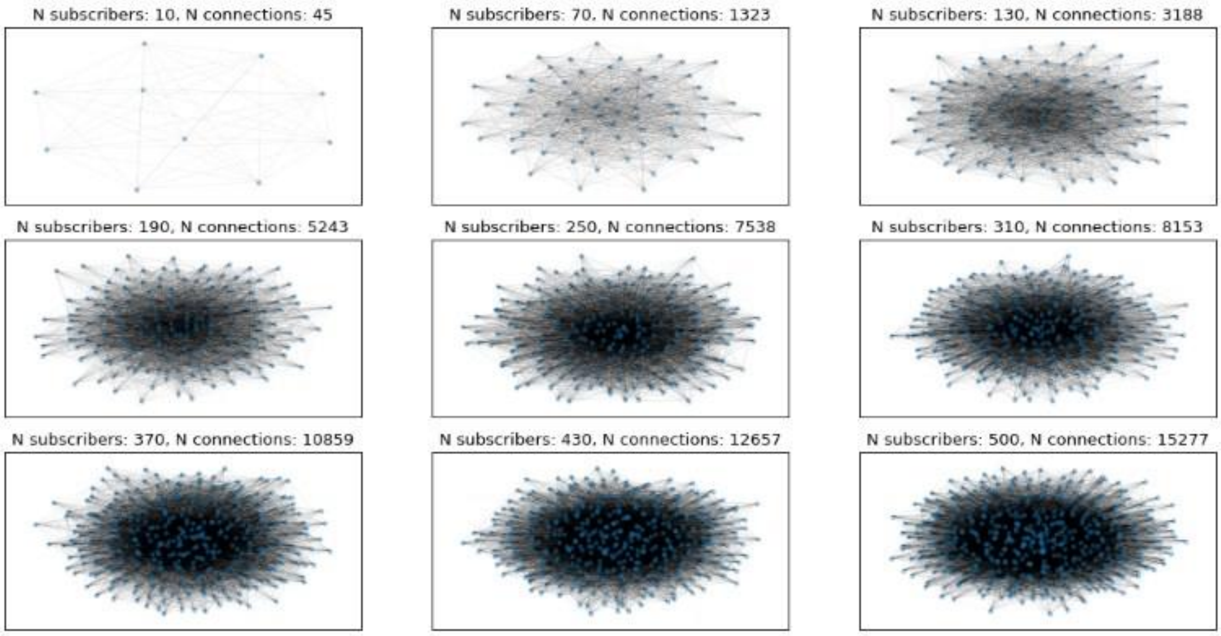

Fig. 2. Graphs of the resulting networks with selected number of subscribers

The number of connections in such networks increases linearly from 45 to 15277 depending on the number of subscribers. As a result of the simulation, we obtained data on the relationship between the number of connections and the number of subscribers. Linear approximation of these data by using the least squares method gives the dependence shown in Fig. 3 with a black line. It is easy to notice good agreement of the data with the given dependence.

$$n = 32 \times (N - 32) \tag{1}$$

where n is the number of connections, N is the number of subscribers.

From this dependence it follows that each new subscriber creates 32 new connections. The deviation is observed in case of a small number of subscribers. There are not yet enough subscribers in this case to establish an asymptotic number of connections.

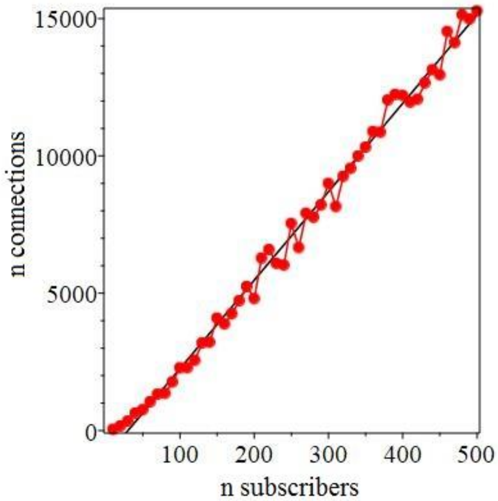

Fig. 3. Relationship between the number of connections and the number of subscribers in simulated networks. Red circles are the simulation data, black ones are the direct approximation.

Network density is another property that we have analyzed, particularly, the changes in network density with an increase in the number of subscribers. Density is the fraction of actually present edges to the maximum possible number of edges. "It can be thought of as the probability that a pair of nodes, picked uniformly at random from the whole network, is connected by an edge" [2]. As a result of the network simulation, the maximum density was 1.00 in a network with 10 subscribers, the minimum 0.12 was observed in a network with 500 subscribers. In Fig. 4 the simulated data is shown in red, and the approximating dependency in black. The data analysis has shown that the network density follows a hyperbolic dependence:

$$\rho = \frac{1.2}{1 + \frac{N}{58}} \quad (2)$$

where ρ is the network density, N is the number of subscribers.

Good agreement with the simulation data is evident in Fig. 4. Thus, as the number of subscribers increases, the network density decreases and tends to 0.

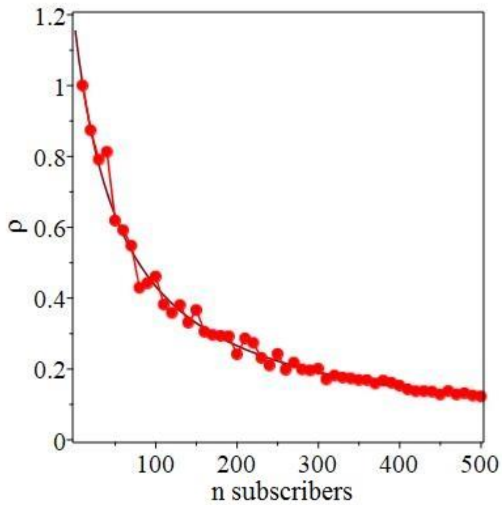

Fig. 4. Relationship between network density and the number of subscribers. Simulation data are in red, hyperbolic dependence curve are in black.

During the experiment, the network subscribers managed to connect with 50-60 unique subscribers in networks with 200 or more subscribers (Fig. 5). This subscriber or graph node property is known as the degree.

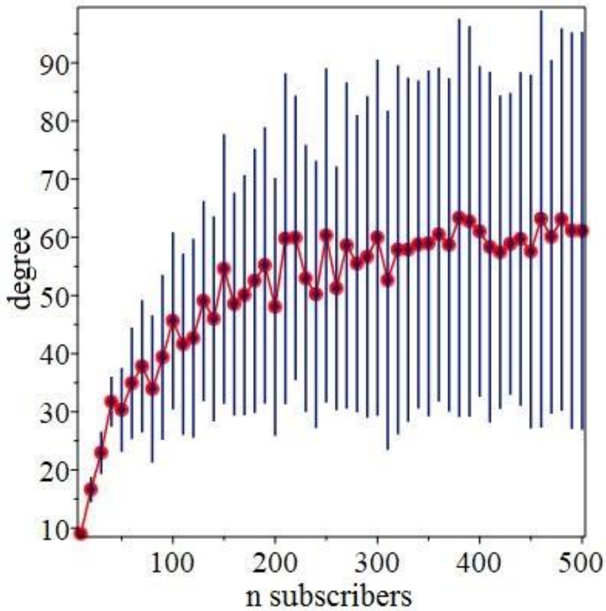

Figure 5. Relationship between the average node degree (number of created connections) and the number of subscribers in the network

Let us consider one of the fundamental properties of the network – the degree distribution. The degree distribution tells us the frequency with which nodes of different degrees appear in the network [2]. As a result of simulating 50 networks, we have obtained the average degree distribution presented in Fig. 6.

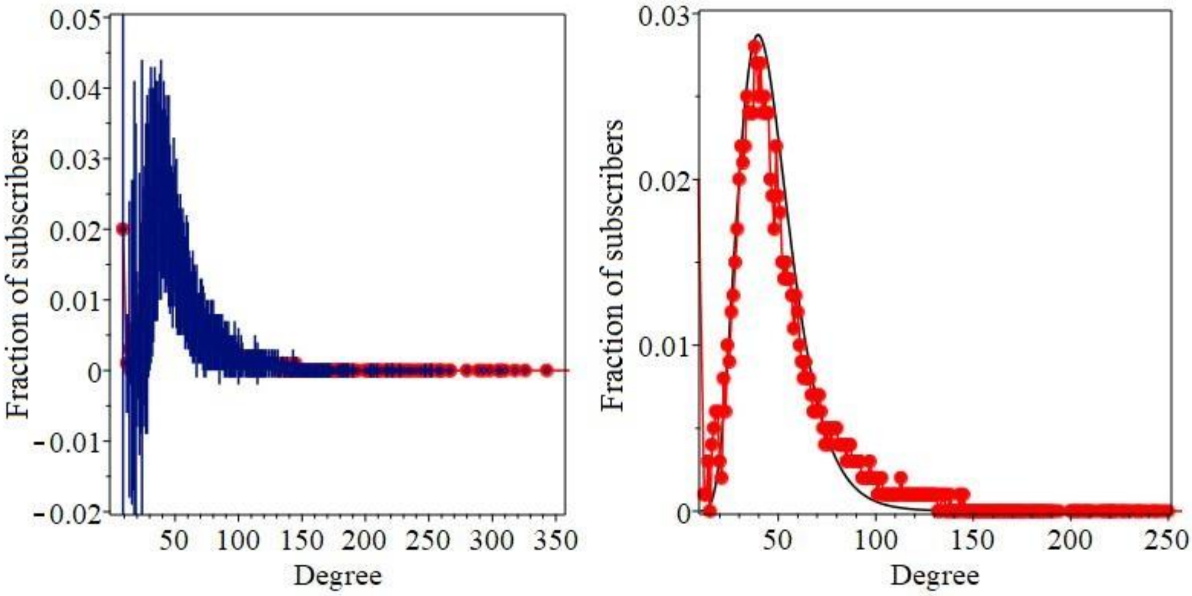

Fig. 6. Distribution of connections established by each subscriber, or node degree distribution. On the left is the distribution of node degrees with standard deviations. On the right is the simulation data in red, the black curve is the lognormal distribution (3) with $\mu = 3.79$ and $\sigma = 0.332$. Good agreement with the simulation data is observed.

The degree distribution corresponds well with the lognormal distribution with parameters $\mu = 3.79$ and $\sigma = 0.332$. Therefore, we observe that with a random selection of subscribers, the distribution of degrees in the network repeats the distribution of the number of subscribers by the number of calls.

$$f(x) = \frac{e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}}{x\sigma\sqrt{2\pi}} \qquad (3)$$

where σ and μ are parameters that determine the lognormal distribution.

The next important network property is the clustering coefficient. The clustering coefficient is the average probability that two neighbors of the same node are themselves neighbors or friends [2]. The higher the value of clustering coefficient, the better information is transmitted over the network. As we can see from the chart in Fig. 7, the clustering coefficient decreases from 1.00 in the fully connected network with 10 subscribers to 0.21 in the network with 500 subscribers.

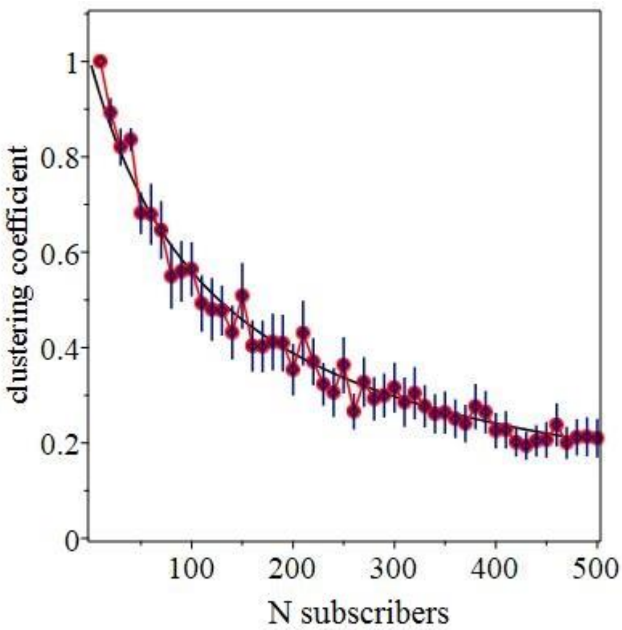

Fig. 7. Relation between the average clustering coefficient and number of subscribers in simulated networks. The red dots are the simulation data, the black curve is the dependence (4).

The simulation data is also well approximated by the hyperbolic dependency, which is shown in Fig. 7 with the black curve. It is clear that as the number of network subscribers increases, the clustering coefficient decreases:

$$C = \frac{1}{1+\frac{N}{127}} \qquad (4)$$

where C is the network clustering coefficient, N is the number of subscribers.

Finally, we consider the dependency between the average shortest path and the number of subscribers in the simulated networks, which is presented in the Fig. 8. The lengths of all links between adjacent nodes are considered equal to one. The shortest path in a network, also sometimes called a geodesic path, is the shortest walk between a given pair of nodes, i.e., the walk that traverses the smallest number of edges [2]. In the case of the telephone network and connections between people, this indicator characterizes the average number of contacts required to connect any two subscribers.

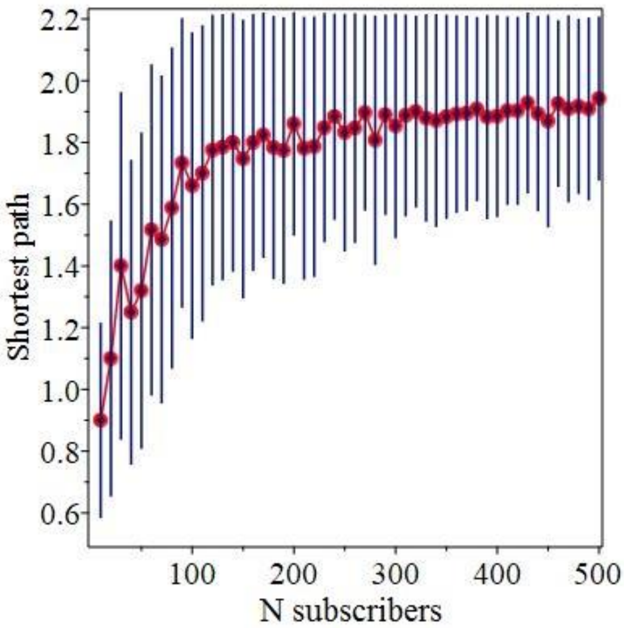

Figure 8. Relation between the average shortest path length and the number of subscribers in networks

In this telephone network model, in which subscribers communicate randomly, the average shortest path is quite low (0.90 – 1.94).

6. Conclusions

In this paper, we propose a simple model of a dynamic network of telephone subscribers. The main simplification of the model is that subscribers choose each other randomly. That leads to some restrictions which determine the network topology and the dynamics of its properties. The basic properties of the network such as the number of connections, density, clustering coefficient, degree distribution and the average shortest path length in the network and their dependence on changes in the number of network nodes or subscribers have been obtained. The modeling results have shown that the number of connections increases linearly with the number of subscribers, which results in 32 connections when a new subscriber appears on the network. Network density decreases with increasing number of subscribers according to the hyperbolic law (2). The network clustering coefficient also decreases with increasing number of subscribers according to a similar hyperbolic law (4). Another important network characteristic that affects the propagation of information is the length of average shortest path between two subscribers. It could be noted that a low average path length of 2 is possible even in a low-density network with the density of 0.12. It may indicate that the network does not need to be fully connected so that subscribers communicate with each other through a small number of contacts. The degree distribution or the number of subscriber connections obtained by the simulation is in good agreement with the lognormal distribution. It is interesting to compare the results obtained with data on real subscriber networks. Due to the random formation of connections, the resulting model creates a network that resembles the Erdes-Rényi random graph model, but the degrees of the vertices are distributed according to the lognormal law.

REFERENCES

1. J. L. Moreno, The first book on group psychotherapy, 3rd ed. Oxford, England: Beacon, House, 1957, pp. xxiv, 138.
2. M. Newman, Networks, vol. 1. Oxford University Press, 2018. DOI: https://doi.org/10.1093/oso/9780198805090.001.0001 (Last accessed: 20.08.2022).
3. P. Erdos and A. Renyi, “On the evolution of random graphs,” Publ. Math. Inst. Hungary. Acad. Sci., vol. 5, pp. 17–61, 1960. URL: https://snap.stanford.edu/class/cs224w-readings/erdos60random.pdf (Last accessed: 20.08.2022).

4. D. Watts and S. Strogatz, “Collective dynamics of ‘small-world’ networks,” Nature, vol. 393, pp. 440–442, 1998, DOI: https://doi.org/10.1038/30918 (Last accessed: 20.08.2022).
5. A.-L. Barabási and R. Albert, “Emergence of Scaling in Random Networks,” Science, vol. 286, no. 5439, pp. 509–512, 1999, DOI: https://doi.org/10.1126/science.286.5439.509 (Last accessed: 20.08.2022).
6. R. Toivonen, J.-P. Onnela, J. Saramäki, J. Hyvönen, and K. Kaski, “A model for social networks,” Physica A: Statistical Mechanics and its Applications, vol. 371, no. 2, pp. 851–860, Nov. 2006, DOI: https://doi.org/10.1016/j.physa.2006.03.050 (Last accessed: 20.08.2022).
7. S. Niaki and Z. B. Rad, “Designing a communication network using simulation,” Scientia Iranica, vol. 11, pp. 165–180, Aug. 2004. URL: https://www.researchgate.net/publication/236107639_Designing_a_communication_network_using_simulation (Last accessed: 20.08.2022).
8. E. Uleia, “Discrete Event Simulator for Communication Networks,” 2007. URL: http://2007.telfor.rs/files/radovi/09_13.pdf (Last accessed: 20.08.2022).
9. B. S. Khan and M. A. Niazi, “Modeling and Analysis of Network Dynamics in Complex Communication Networks Using Social Network Methods,” ArXiv, vol. abs/1708.00186, 2017, URL: https://arxiv.org/ftp/arxiv/papers/1708/1708.00186.pdf (Last accessed: 20.08.2022).
10. T. Johansson, “Generating artificial social networks,” The Quantitative Methods for Psychology, vol. 15, no. 2, pp. 56–74, 2019, DOI: https://doi.org/10.20982/tqmp.15.2.p056 (Last accessed: 20.08.2022).
11. “NetworkX — NetworkX documentation.” URL: https://networkx.org/ (Last accessed: 20.08.2022).
12. C. L. STAUDT, A. SAZONOVS, and H. MEYERHENKE, “NetworKit: A tool suite for large-scale complex network analysis,” Network Science, vol. 4, no. 4, pp. 508–530, 2016, DOI: https://doi.org/10.1017/nws.2016.20 (Last accessed: 20.08.2022).
13. G. Miritello, E. Moro, R. Lara, R. Martínez-López, S. G. B. Roberts, and R. I. M. Dunbar, “Time as a limited resource: Communication Strategy in Mobile Phone Networks.” arXiv, Jan. 11, 2013. URL: https://arxiv.org/abs/1301.2464 (Last accessed: 20.08.2022).
14. A. A. Nanavati et al., “On the Structural Properties of Massive Telecom Call Graphs: Findings and Implications,” in Proceedings of the 15th ACM International Conference on Information and Knowledge Management, in CIKM ’06. New York, NY, USA: Association for Computing Machinery, 2006, pp. 435–444. DOI: https://doi.org/10.1145/1183614.1183678 (Last accessed: 20.08.2022).
15. J.-P. Onnela et al., “Structure and tie strengths in mobile communication networks,” Proc. Natl. Acad. Sci. U.S.A., vol. 104, no. 18, pp. 7332–7336, May 2007, DOI: https://doi.org10.1073/pnas.0610245104 (Last accessed: 20.08.2022).
16. M. Zignani, C. Quadri, S. Gaitto, and G. P. Rossi, “Exploiting all phone media? A multidimensional network analysis of phone users’ sociality.” Jan. 14, 2014. Accessed: Oct. 19, 2023. URL: https://arxiv.org/abs/1401.3126 (Last accessed: 20.08.2022).
17. V. Danilevskiy and V. Yanovsky, “Statistical properties of telephone communication network,” arXiv preprint arXiv:2004.03172, 2020. URL: https://arxiv.org/pdf/2004.03172.pdf (Last accessed: 20.08.2022).

Данілевський Михайло Вікторович — *аспірант; Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків-22, Україна, 61022*
e-mail: m.danilevskiy@gmail.com
https://orcid.org/0009-0000-0030-2218

Яновський Володимир Володимирович — *доктор фізико-математичних наук, професор, професор кафедри штучного інтелекту та програмного забезпечення Харківський національний університет імені В. Н. Каразіна, майдан Свободи 4, Харків-22, Україна, 61022 Завідувач теоретичним відділом, інститут монокристалів НАН України, пр. Науки 60, Харків, Україна, 61001*
e-mail: yanovsky@isc.kharkov.ua
https://orcid.org/0000-0003-0461-749X

Моделювання та аналіз найпростішої мережі телефонних абонентів

Актуальність. Динамічні мережі представлені у широкому спектрі областей сучасного світу, включаючи соціальні, транспортні та біологічні мережі. Моделювання складних мереж як структур, що змінюються в часі, відкриває додаткові можливості для вивчення їх властивостей.

Мета. Метою роботи є моделювання найпростішої динамічної мережі телефонних абонентів. Основна увага зосереджена на експериментах з отриманою моделлю та дослідження впливу кількості абонентів на властивості мережі.

Методи дослідження. У роботі використовуються метод Монте-Карло стохастичної динаміки дискретних станів із використанням часових кроків однакової довжини, а також методи побудови комп'ютерних моделей, методи аналізу властивостей мереж, метод найменших квадратів та інші. Комп'ютерна модель розроблена мовою Python із використанням бібліотек Pandas, Numpy та NetworkX.

Результати. Розроблено найпростішу модель мережі телефонних абонентів, у якій абоненти обирають інших абонентів випадковим чином, а зв'язки існують тільки під час телефонної розмови. В моделі середньоденна кількість вихідних дзвінків абонентів розподілена за логнормальним законом. Проведено експерименти з моделлю з різною кількістю абонентів, але за однаковий часовий відрізок. На підставі отриманих даних про дзвінки, розглянуті такі властивості мереж як кількість зв’язків, щільність, розподіл вершин, середній коефіцієнт кластеризації та середня довжина найкоротшого шляху.

Висновки. Розроблена комп'ютерна модель найпростішої динамічної мережі телефонних абонентів формує модель схожу до випадковий граф Ердеша-Реньї, але при цьому ступені вершин або кількість зв'язків абонентів розподілено за логнормальним законом. Розроблена комп'ютерна модель може бути основою розробки складніших моделей та вивчення динамічних властивостей подібних мереж.

Ключові слова: *складна динамічна мережа, граф мобільних викликів, телефонна мережа, логнормальний розподіл, розподіл ступенів, щільність мережі, коефіцієнт кластеризації, середня довжина найкоротшого шляху*.

Надійшла у першій редакції 18.06.2023, в останній - 21.08.2023.
The first version has been received on 18.06.2023, the final version - on 21.08.2023.