УДК 004.942, 004.056
DOI: 10.26102/2310-6018/2023.43.4.034

Приближенная оценка условий прекращения эпидемии компьютерного вируса в связных сетях, ассоциированных со случайными графами

А.Ю. Никифорова✉

Омский государственный технический университет, Омск, Российская Федерация

Резюме. Математическое моделирование эпидемий компьютерных вирусов является важнейшим направлением теоретических исследований в области информационной безопасности. В настоящей работе исследуется марковская модель распространения компьютерных вирусов, основанная на модели Рида-Фроста. Основная цель статьи – анализ применимости модифицированной модели Рида-Фроста к классу сетей, ассоциированных со случайными графами Эрдёша-Реньи. В частности, проверялось влияние отношения вероятности лечения к вероятности заражения на прекращение распространения компьютерного вируса. Результаты, полученные с помощью данной модели, сравнивались с результатами имитационного моделирования при различных значениях параметров эпидемии и характеристик сети. В проведенных расчетах и экспериментах изменялись следующие параметры: вероятность заражения, вероятность лечения, а также связность сети. Для расчетов применялась система символьных вычислений Wolfram Mathematica. Для проведения вычислительного эксперимента была использована программа на языке C++, написанная ранее автором и его научным руководителем. Проведенные исследования показали, что при определенных параметрах условие прекращения эпидемии подтверждается как теоретическими расчетами, так и результатами экспериментов. Однако эпидемия прекращается раньше, чем достигается расчетное пороговое значение. В дальнейших исследованиях автор планирует дать более точную теоретическую оценку условий прекращения эпидемии.

Ключевые слова: компьютерный вирус, вероятность заражения, вероятность излечения, случайный граф, модель Рида-Фроста, восприимчивый узел.

Для цитирования: Никифорова А.Ю. Приближенная оценка условий прекращения эпидемии компьютерного вируса в связных сетях, ассоциированных со случайными графами. *Моделирование, оптимизация и информационные технологии.* 2023;11(4). URL: https://moitvivt.ru/ru/journal/pdf?id=1483 DOI: 10.26102/2310-6018/2023.43.4.034

An approximate evaluation of the conditions for the termination of a computer virus epidemic in connected networks associated with random graphs

A.Y. Nikiforova✉

Omsk State Technical University, Omsk, the Russian Federation

Abstract. Mathematical modeling of computer virus epidemics is the most important area of theoretical research in the field of information security. This paper examines a Markov model of the computer virus spread based on the Reed–Frost model. The main aim of the article is to analyze the applicability of the modified Reed-Frost model to the class of networks associated with random Erdos-Renyi graphs. In particular, the effect of the ratio of the probability of cure to the probability of infection on stopping the spread of a computer virus was tested. The results of this model are compared with ones obtained via the simulation modeling for different values of epidemic parameters and network characteristics. In the calculations and experiments carried out, the following parameters changed: the probability of infection, the probability of cure, as well as the connectivity of the network. The Wolfram Mathematica symbolic computing system was used for calculations. A C++ program written earlier by the author and their supervisor was used to conduct the computational experiment. The studies show that, under certain parameters, the condition for ending the epidemic is confirmed by both theoretical calculations and experimental results. However, the epidemic vanishes before the threshold value calculated is reached. In the future, the author plans to give a more accurate theoretical assessment of the conditions for ending the epidemic.

Keywords: computer virus, probability of infection, probability of cure, random graph, Reed-Frost model, susceptible node.

For citation: Nikiforova A.Y. An approximate evaluation of the conditions for the termination of a computer virus epidemic in connected networks associated with random graphs. *Modeling, Optimization and Information Technology*. 2023;11(4). URL: https://moitvivt.ru/ru/journal/pdf?id=1483 DOI: 10.26102/2310-6018/2023.43.4.034 (In Russ.).

Введение

Эпидемии различных вирусов с древнейших времен до наших дней представляют собой серьезную угрозу для человека. Медики проводили наблюдения и писали статьи на эту тему со времен Гиппократа [1]. В XVIII веке в борьбу с эпидемиями включились математики. Даниилом Бернулли в 1760 году была предложена первая математическая модель, доказавшая эффективность вакцинации против оспы для уменьшения количества заболевших [2]. Далее в 1927 году Кермак и Маккендрик представили свою математическую модель эпидемий [3]. Практически в это же время математиком Лоуэллом Ридом и врачом Уэйдом Хэмптоном Фростом из Университета Джонса Хопкинса была разработана модель биномиальной цепочки распространения болезней, которую они использовали на своих занятиях по биостатистике и эпидемиологии, но опубликовали лишь в 1950 году [4].

Во второй половине XX века, когда компьютеры и интернет-технологии стали применяться во всех отраслях экономики, а также в личной жизни людей, появились новые опасности – вредоносные компьютерные программы (далее – ВКП), которые приносили значительный ущерб. За ВКП прочно закрепилось название «компьютерный вирус» из-за похожего механизма распространения. Обнаружив это сходство, уже в 1980-е годы ученые стали исследовать возможность применения моделей распространения биологических вирусов для борьбы с ВКП. В 1987 году Коэном [5], а затем в 1990-х Кефартом и Уайтом [6, 7] были описаны первые математические модели эпидемий компьютерных вирусов. В настоящее время такие исследования активно продолжаются [8-10]. В журнале «Physics Letters A» в 2002 году авторами L. Billings, W. M. Spears и I. B. Schwartz была опубликована статья «A unified prediction of computer virus spread in connected networks» [8], в которой была представлена модель Рида-Фроста, модифицированная для моделирования эпидемий в сложных связных компьютерных сетях. Особенностью данной модели является то, что она может быть полностью описана на языке марковских цепей и время в ней предполагается дискретным. В работах [11-13] автор с коллегами показали, что модифицированная модель Рида-Фроста хорошо согласуется с имитационными моделями компьютерных эпидемий на классе случайных графов Эрдёша-Реньи [14]. В статье [8] L. Billings и соавторы также описали примерную зависимость прекращения эпидемии компьютерного вируса от свойств компьютерной сети и вероятностей заражения и лечения. Целью данного исследования является анализ применимости условий прекращения эпидемии вируса, предложенных в [8], для компьютерных сетей, ассоциированных со случайными графами Эрдёша-Реньи. Результаты, полученные с помощью данной модели, будут сопоставлены с результатами имитационного моделирования при различных значениях параметров эпидемии и характеристик сети.

Материалы и методы

Для достижения поставленной цели проведем дальнейшее исследование модели, базирующейся на модели Рида-Фроста, которая была разработана в 1920-е годы для изучения распространения биологических вирусов. В 2002 году L. Billings с соавторами в [8] адаптировали ее для моделирования эпидемий компьютерных вирусов.

В этой модели компьютерная сеть описывается посредством случайного графа. Вершины графа соответствуют компьютерам, а возможные связи между ними – ребрам графа. Граф состоит из конечного числа вершин (или узлов) N. Время предполагается дискретным. Каждый узел сети может находиться в одном из двух состояний: быть либо инфицированным I, либо восприимчивым S. В каждый следующий момент времени инфицированный узел может остаться инфицированным либо вылечиться с вероятностью, которую обозначим δ, значение которой является постоянной величиной, отражающей эффективность антивирусного программного обеспечения. В этот же момент восприимчивый узел может остаться восприимчивым либо стать инфицированным с вероятностью, которую обозначим μ. Она, в отличие от δ, является переменной величиной, зависящей от β – вероятности передачи вируса от зараженного узла к восприимчивому, и от связности – с, которая вычисляется как среднее относительное число соседей у узлов сети. Число I инфицированных узлов будет определять марковское состояние сети. Поскольку S в любой момент времени равно N - I, число восприимчивых узлов определяется состоянием сети однозначно. Исходя из вышеперечисленного, представим компьютерную сеть в виде марковской цепи, состояния которой определяются числом зараженных узлов I. Полное число возможных состояний этой модели равно $N + 1$.

Динамика соответствующей марковской цепи описывается вероятностями переходов из состояний I в состояния I'.

$$p_{I,I'} = \Sigma_{x=\max\{0,I-I'\}}^{\min\{I,S-I'+I\}} C_I^x(\delta^x)(1-\delta)^{I-x} C_S^{x+I'-I} \mu^{x+I'-I}(1-\mu)^{S-(x+I'-I)}, \qquad (1)$$

где вероятность μ восприимчивому узлу стать заражённым в состоянии I задается равенством

$$\mu(I) = \Sigma_{k=0}^{I}(1-(1-\beta)^k) C_I^k c^k (1-c)^{I-k} = 1-(1-\beta c)^I. \qquad (2)$$

В формулах (1 и 2) C_I^x обозначает биномиальный коэффициент $C_I^x = \frac{I\,!}{x\,!\,(I-x)\,!}$.

С помощью вероятностей вида (1) можно составить так называемую переходную матрицу $\Pi = (p_{I,I'})$. С помощью этой матрицы в соответствии со стандартными методами теории марковских цепей можно вычислить вероятности различных состояний сети. Для этого также требуется задать начальные вероятности, которые обычно определяются так:

$$p_I(0) = \begin{cases} 1, & I = I_0 \\ 0, & I \neq I_0 \end{cases}, \qquad (3)$$

где I_0 – начальное число зараженных узлов. Зная вероятности состояний сети в произвольный момент времени, можно вычислить среднее число зараженных узлов в данный момент времени. По определению – это сумма каждого состояния, умноженная на соответствующую ему вероятность.

Для проведения вычислений в соответствии с описанной моделью было разработано несколько подпрограмм в системе символьных вычислений Wolfram

Mathematica. При работе с подпрограммами имеется возможность задавать следующие начальные условия: число узлов в графе N, его связность c, вероятность передачи вируса от зараженного узла к восприимчивому β, вероятность выздоровления δ, количество зараженных узлов в начальный момент времени. При выборе значения связности c нужно принять во внимание следующие ограничения: с одной стороны, связность реальных компьютерных сетей мала [13], с другой стороны, согласно [14] становится несвязным, если выполняется условие

$$c \le \frac{\ln N}{N}. \tag{4}$$

В работе [8] L. Billings и соавторы приводят следующее условие прекращения эпидемии компьютерного вируса

$$\delta \ge N\beta c. \tag{5}$$

С учетом граничного условия связности, получаем следующее выражение

$$\frac{\ln N}{N} \le c \le \frac{\delta}{N\beta}. \tag{6}$$

Упростим полученное выражение, умножив на N, принимая во внимание, что $Nc = \langle k \rangle$, обозначающее среднюю степень узла графа, получаем:

$$\ln N \le \langle k \rangle \le \frac{\delta}{\beta}. \tag{7}$$

Обозначим $k_{\text{порог}} = Nc = \langle k \rangle$. Таким образом, когда $\frac{\delta}{\beta} \ge k_{\text{порог}}$, эпидемия вируса прекращается, в обратном случае, даже при наличии лечения некоторое количество компьютеров сети будет заражено.

В данном исследовании примем N = 100 (число компьютеров в небольшом предприятии). Граф с N = 100 в соответствии с выражением (4) станет несвязным при $c \approx 0{,}046$, поэтому для первой серии вычислений примем c = 0,05, тогда для заданных условий $k_{\text{порог}} = 5$, следовательно, согласно неравенству (7) эпидемия вируса должна прекращаться при $\frac{\delta}{\beta} \ge 5$. Для второй серии вычислений примем c = 0,07, тогда для заданных условий $k_{\text{порог}} = 7$, следовательно, согласно неравенству (7) эпидемия вируса должна прекращаться при $\frac{\delta}{\beta} \ge 7$.

Были проведены расчеты при разных значениях и соотношениях между δ и β, по результатам которых были построены графики зависимости количества зараженных узлов от времени. Далее эти графики будут сравниваться с графиками, полученными по результатам экспериментов.

Для проведения вычислительного эксперимента нами был разработан алгоритм и написана программа на языке С++, с помощью которой мы имеем возможность создавать случайные графы и имитировать распространение вируса [15].

В программе можно задавать и изменять следующие входные параметры: количество узлов в графе N, связность графа c, вероятность передачи вируса β, вероятность излечения зараженного узла δ, начальное количество зараженных узлов I_0, количество повторений эксперимента для заданных условий, время наблюдения за эпидемией.

На первом шаге программа генерирует случайный граф по заданным параметрам N и c. Для первой серии экспериментов, как и для первой серии вычислений, задаем N = 100, c = 0,05 и соответствующие теоретическим значения δ и β. Во второй серии экспериментов примем c = 0,07.

В данном исследовании была выбрана тактика изначально заражать только 1 узел. Далее этот узел с заданной вероятностью β пытается заразить своих соседей, после чего зараженные узлы могут вылечиться с вероятностью δ. В следующий момент времени описанные шаги повторяются до тех пор, пока значение времени не достигнет заранее выбранного значения. После этого весь описанный эксперимент повторяется заданное количество раз. В итоге все полученные данные усредняются и выгружаются в файл с расширением txt, который в дальнейшем загружается в систему символьных вычислений Wolfram Mathematica для построения графика распространения вируса и сравнения с графиком, вычисленным теоретически.

Результаты

Автором было проведено более 30 экспериментов с разными значениями δ, β и c. Далее на Рисунках 1-3 приведены типичные графики, иллюстрирующие полученные результаты.

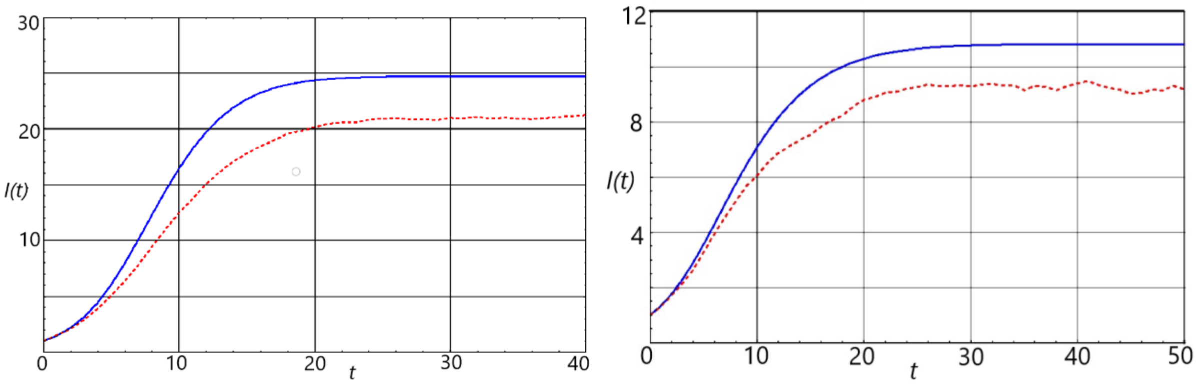

Рисунок 1 – Зависимость среднего числа зараженных узлов *I* от времени:
слева - при c = 0,05, δ = 0,5, β = 0,2, δ/β = 2,5; справа - при c = 0,07, δ = 0,7, β = 0,15, $\delta/\beta \approx 4{,}67$.
Сплошная линия – теоретический график, пунктирная линия – экспериментальный
Figure 1 – Dependence of the mean number of infected nodes I on time: on the left – at c = 0,05, δ = 0,5, β = 0,2, δ/β = 2,5; on the right - at c = 0,07, δ = 0,7, β = 0,15, $\delta/\beta \approx 4{,}67$.
The solid line is a theoretical graph, the dotted line is an experimental one

Обсуждение

На Рисунке 1 представлены 2 графика, на которых $\frac{\delta}{\beta}$ значительно меньше $k_{\text{порог}}$, и несмотря на лечение эпидемия не прекращается, что подтверждает справедливость выражения (5).

На Рисунке 2 представлены 2 графика, на которых $\frac{\delta}{\beta}$ равно $k_{\text{порог}}$, и, как предсказывало выражение (5), наблюдается вымирание вируса и прекращение эпидемии.

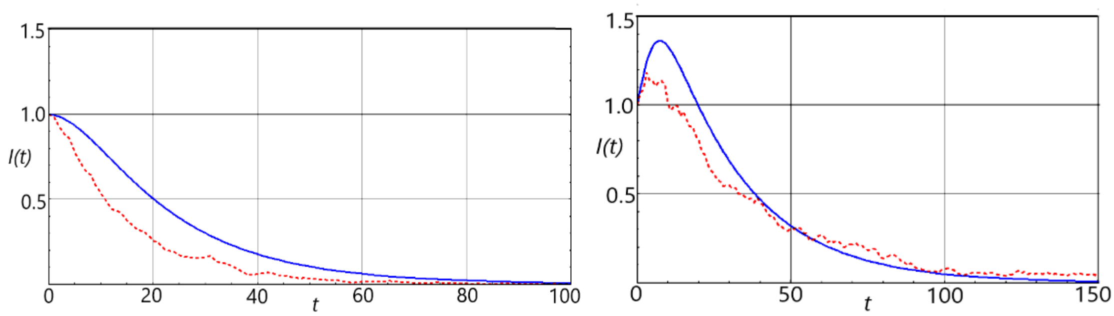

Рисунок 2 – Зависимости количества зараженных узлов от времени:
слева - при c = 0,05, δ = 0,5, β = 0,1, δ/β = 5; справа - при c = 0,05, δ = 0,9, β = 0,2, δ/β=4,5.
Сплошная линия – теоретический график, пунктирная линия – экспериментальный
Figure 2 – Dependence of the mean number of infected nodes I on time: on the left - at c = 0,05, δ = 0,5, β = 0,1, δ/β = 5; on the right - at c = 0,05, δ = 0,9, β = 0,2, δ/β=4,5.
The solid line is a theoretical graph, the dotted line is an experimental one

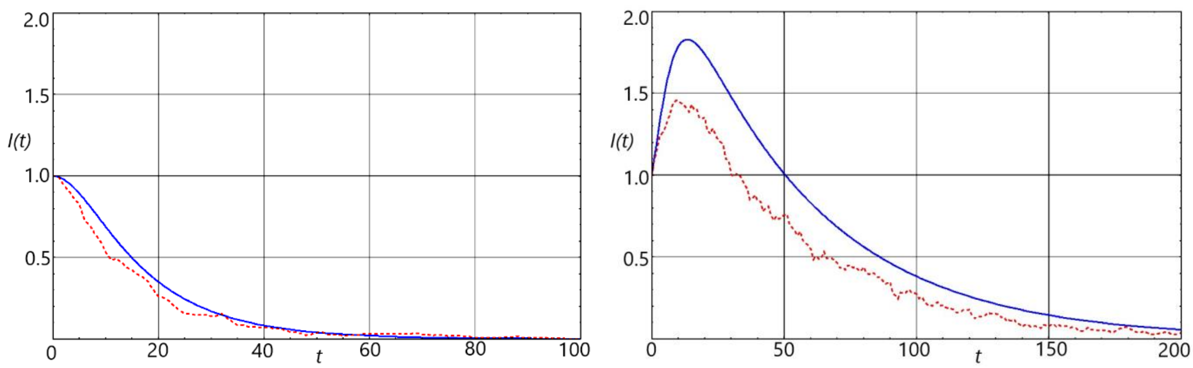

Рисунок 3 – Зависимости количества зараженных узлов от времени:
слева – при c = 0,07, δ = 0,7, β = 0,1, δ/β = 7; справа – при c = 0,07, δ = 0,7, β = 0,117, δ/β≈6.
Сплошная линия – теоретический график, пунктирная линия – экспериментальный
Figure 3 – Dependence of the mean number of infected nodes I on time: on the left - at c = 0,07, δ = 0,7, β = 0,1, δ/β = 7; on the right - at c = 0,07, δ = 0,7, β = 0,117, δ/β≈6.
The solid line is a theoretical graph, the dotted line is an experimental one

На Рисунке 3 представлены 2 графика, на которых $\frac{\delta}{\beta} = (k_{\text{порог}} - 1)$, но, вопреки выражению (5), наблюдается вымирание вируса и прекращение эпидемии.

При разных значениях δ, β и c было проведено более 30 экспериментов, расчеты и эксперименты показывают, что эпидемия прекращается при $\frac{\delta}{\beta} \cong (k_{\text{порог}} - 1)$.

Заключение

Когда отношение вероятности лечения δ к вероятности заражения β больше или значительно меньше $k_{\text{порог}}$, результаты вычислений и экспериментов хорошо согласуются с условием $\delta \geq N\beta c$, предложенным в [8]. Однако проведенные исследования показали, что эпидемия прекращается раньше, чем достигается расчетное пороговое значение. Автор планирует в дальнейших исследованиях дать более точную теоретическую оценку условий прекращения эпидемии.

СПИСОК ИСТОЧНИКОВ

1. Гиппократ. *Избранные книги.* Перевод с греческого профессора В.И. Руднева. М.: Гос. изд-во биол. и мед. лит-ры; 1936. 736 с.
2. Григорьян А.Т., Ковалёв Б.Д. *Даниил Бернулли, 1700—1782*. М.: Наука; 1981. 320 с.
3. Kermack W.O., McKendrick A.G. Contributions to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 1927;771(115):700–721.
4. Abbey H. An examination of the Reed-Frost theory of epidemics. *Human Biology*. 1952;24(3):201–233.
5. Cohen F. Computer viruses: Theory and experiments. *Computers & Security*. 1987;6(1):22–35. DOI: 10.1016/0167-4048(87)90122-2.
6. Kephart J., White S. Directed-graph epidemiological models of computer viruses. *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*. 1991;(May 20-22):343–359. DOI: 10.1142/9789812812438_0004.
7. Kephart J., White S. Measuring and modeling computer virus prevalence. *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*. 1993;(May 24-26):2–15. DOI: 10.1109/RISP.1993.287647.
8. Billings L., Spears W.M., Schwartz I.B. A unified prediction of computer virus spread in connected networks. *Physics Letters A*. 2002;297(3–4):261–266. DOI: 10.1016/S0375-9601(02)00152-4.
9. Van de Bovenkamp R., Van Mieghem P. Survival time of the susceptible-infected-susceptible infection process on a graph. *Physical Review E*. 2015;92(3):1–16. DOI: 10.1103/PhysRevE.92.032806.
10. Pastor-Satorras R., Castellano C., Van Mieghem P., Vespignani A. Epidemic processes in complex networks. *Reviews of Modern Physics*. 2015;87(3):925–978. DOI: 10.1103/RevModPhys.87.925.
11. Никифорова А.Ю., Магазев А.А. О вероятности заражения восприимчивого узла в модели Рида-Фроста. *Прикладная математика и фундаментальная информатика*. 2020;7(4):34–41. DOI: 10.25206/2311-4908-2020-7-4-34-41.
12. Бельченко А.О., Магазев А.А., Никифорова А.Ю. Приближённая оценка среднего числа заражённых узлов в марковской модели распространения компьютерных вирусов. *Математические структуры и моделирование*. 2022;61(1):92–104. DOI: 10.24147/2222-8772.2022.1.92-104.
13. Networks Repository. An Interactive Scientific Network Data Repository. URL: https://networkrepository.com (дата обращения: 20.11.2023).
14. Erdős P., Renyi A. On Random Graphs. *Publicationes Mathematicae (Debrecen)*. 1959;6:290–297.
15. Магазев А.А., Никифорова А.Ю. *Программа для оценки среднего времени распространения компьютерного вируса в сетях, ассоциированных со случайными графами*: свидетельство о регистрации электронного ресурса. М.: ФИПС, 2023. № 2023614819 от 06.03.2023.

REFERENCES

1. Gippokrat. *Selected books.* Translated from the Greek professor V.I. Rudnev. Moscow, Gos. izd-vo biol. i med. lit-ry; 1936. 736 p. (In Russ.).
2. Grigor'jan A.T., Kovaljov B.D. *Daniil Bernulli, 1700—1782*. Moscow, Nauka=Nauka Publishers; 1981. 320 p. (In Russ.).

3. Kermack W.O., McKendrick A. G. Contributions to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 1927;772(115):700–721.
4. Abbey H. An examination of the Reed-Frost theory of epidemics. *Human Biology*. 1952;24(3):201–233.
5. Cohen F. Computer viruses: Theory and experiments. *Computers & Security*. 1987;6(1):22–35. DOI: 10.1016/0167-4048(87)90122-2.
6. Kephart J., White S. Directed-graph epidemiological models of computer viruses. *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*. 1991;(May 20-22):343–359. DOI: 10.1142/9789812812438_0004.
7. Kephart J., White S. Measuring and modeling computer virus prevalence. *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*. 1993;(May 24-26):2–15. DOI: 10.1109/RISP.1993.287647.
8. Billings L., Spears W. M., Schwartz I. B. A unified prediction of computer virus spread in connected networks. *Physics Letters A*. 2002;297(3–4):261–266. DOI: 10.1016/S0375-9601(02)00152-4.
9. Van de Bovenkamp R., Van Mieghem P. Survival time of the susceptible-infected-susceptible infection process on a graph. *Physical Review E*. 2015;92(3):1–16. DOI: 10.1103/PhysRevE.92.032806.
10. Pastor-Satorras R., Castellano C., Van Mieghem P., Vespignani A. Epidemic processes in complex networks. *Reviews of Modern Physics*. 2015;87(3):925–978. DOI: 10.1103/RevModPhys.87.925.
11. Nikiforova A.Yu., Magazev A. A. On the probability of infection of a susceptible node for the reed-frost model. *Prikladnaja matematika i fundamental'naja informatika=Applied Mathematics and fundamental Computer Science*. 2020;7(4):34–41. DOI: 10.25206/2311-4908-2020-7-4-34-41. (In Russ.).
12. Bel'chenko A.O., Magazev A.A., Nikiforova A.Yu. An approximate evaluation of the infected nodes number fora markov model of viruses spreading. *Matematicheskie struktury i modelirovanie=Mathematical Structures and Modeling*. 2022;1(61):92–104. DOI: 10.24147/2222-8772.2022.1.92-104. (In Russ.).
13. Networks Repository. An Interactive Scientific Network Data Repository. URL: https://networkrepository.com (дата обращения: 20.11.2023).
14. Erdős P., Renyi A. On Random Graphs. *Publicationes Mathematicae (Debrecen)*. 1959;6:290–297.
15. Magazev A.A., Nikiforova A.Yu. *A program for estimating the average time of the spread of a computer virus in networks associated with random graphs*: certificate of registration of an electronic resource. Moscow, FIPS= Federal Institute of Industrial Property, 2023. No. 2023614819 от 06.03.2023. (In Russ.).

ИНФОРМАЦИЯ ОБ АВТОРЕ / INFORMATION ABOUT THE AUTHOR

Никифорова Ангелина Юрьевна, аспирант, Омский государственный технический университет, Омск, Российская Федерация.
e-mail: skt-omgtu@mail.ru
ORCID: 0000-0002-0981-7127

Angelina Y. Nikiforova, Postgraduate Student, Omsk State Technical University, Omsk, the Russian Federation.

Статья поступила в редакцию 05.12.2023; одобрена после рецензирования 18.12.2023; принята к публикации 28.12.2023.

The article was submitted 05.12.2023; approved after reviewing 18.12.2023; accepted for publication 28.12.2023.