

Раздел II. Анализ данных, моделирование и управление

УДК 004.942, 004.056

DOI 10.18522/2311-3103-2025-5-53-63

А.А. Магазёв, А.Ю. Никифорова

О ВЫЧИСЛЕНИИ СРЕДНЕГО ВРЕМЕНИ ИНФИЦИРОВАНИЯ В РАМКАХ ДИСКРЕТНОЙ МАРКОВСКОЙ ЭПИДЕМИОЛОГИЧЕСКОЙ МОДЕЛИ В ОТСУТСТВИИ ЛЕЧЕНИЯ

Моделирование распространения вирусов является актуальной областью исследований. Существует множество «непрерывных» эпидемических моделей, основанных на использовании систем дифференциальных уравнений. Недостатком таких моделей является то, что они имеют погрешность при описании начальной стадии распространения вируса и не учитывают особенности связей между индивидуумами. «Дискретные» модели, в которых время и количество инфицированных и восприимчивых узлов являются дискретными величинами, дают более точную картину эпидемического процесса. В этой работе мы изучаем некоторую дискретную марковскую модель в случае, когда лечение отсутствует. Это важный случай, поскольку его можно рассматривать либо как приближение к начальной фазе эпидемии, либо как модель эпидемий вирусов, которые трудно поддаются лечению. В первом разделе мы подробно описываем свойства исследуемой марковской модели. Во втором разделе, используя марковский подход, мы определяем среднее время заражения, то есть количество временных шагов, затраченных на заражение всех особей в популяции. Однако расчет среднего времени заражения в популяциях с большим количеством особей (или в сетях с большим количеством узлов) является сложной вычислительной задачей, поэтому в третьем разделе мы предлагаем соответствующую приближенную формулу для этого параметра при условии, что связность сети и вероятность распространения вируса малы. В четвертом разделе мы используем метод имитационного моделирования для расчета среднего времени заражения, а затем сравниваем результаты, полученные различными методами. Для проведения вычислительного эксперимента нами было разработано консольное приложение, написанное на языке программирования C++. Анализ значений среднего времени инфицирования, определенных тремя методами: методом точного вычисления фундаментальной матрицы M , вычислением с применением приближенной формулы и методом имитационного моделирования, показал, что методы хорошо согласуются между собой при заданных нами условиях. Полученная приближенная формула для среднего времени заражения является более простым в использовании вариантом расчета данного параметра.

Эпидемиологическая модель; компьютерный вирус; марковская цепь; случайный граф; среднее время инфицирования.

А.А. Magazev, A.Yu. Nikiforova

ON CALCULATING THE MEAN INFECTED TIME USING A DISCRETE MARKOV EPIDEMIOLOGICAL MODEL WITHOUT TREATMENT

Modeling of the spread of viruses is a relevant research field. There are a lot of «continuous» epidemic models based on the use of systems of differential equations. The disadvantage of such models lies in their error in describing the initial stage of virus propagation and in the fact that they ignore the specific features of inter-individual connections. «Discrete» models, in which the time and the number of infected and susceptible nodes are discrete values, provide a more accurate picture of the epidemic process. In this work, we study a discrete Markov model in the case when there is no treatment. This is an important case, since it can be viewed as either an approximation to the initial phase of an epidemic or as a model for epidemics of viruses that are difficult to treat. The first section provides a detailed description of the

properties of the Markov model used in this study. In the second section, using Markov approach, we define the mean infected time, i.e. the number of time steps taken to infect all individuals in the population. However, calculating the mean infected time in populations with a large number of individuals (or in networks with a large number of nodes) is computationally difficult problem, so in the third section we propose the corresponding approximate formula for this parameter. This approximation is designed for conditions of low network connectivity and a low probability of virus spread. In the fourth section, to validate our approximate formula, we compare its results against both exact calculations (using the fundamental matrix M) and data from simulation modeling. For the simulations, we developed a custom C++ console application. Our analysis demonstrates that all three methods yield consistent results under the specified conditions, confirming the practical utility of the simpler approximate formula.

Epidemic model; computer virus; Markov chain; random graph; mean infected time.

Введение. Различные типы вирусов, как компьютерных, так и биологических, продолжают оставаться серьезной угрозой для деятельности как больших корпораций, так и отдельных индивидуумов. Не смотря на усилия специалистов в разных областях, создающих новые вакцины для защиты от биологических вирусов и актуализирующих противовирусное программное обеспечение, ранее неизвестные модификации за время до создания «противоядия» часто успевают нанести большой ущерб. В связи с этим моделирование процессов распространения вирусов продолжает оставаться актуальным направлением исследований.

Математический подход к описанию эпидемий инфекционных заболеваний начал активно применяться в начале 20-ого века. Отметим, что в 1927 году А. Кермак и У. Маккендрик сформулировали качественно новый подход к описанию эпидемий с применением систем дифференциальных уравнений [1], который сразу стал широко использоваться и до сих пор является основой при построении «непрерывных» моделей эпидемий [2–4]. Примерно в то же время математиком Л. Ридом и врачом У.Х. Фростом из Университета Дж. Хопкинса была разработана модель биномиальной цепочки распространения болезней, которую они использовали на своих занятиях по биостатистике и эпидемиологии. Модель не была опубликована авторами, но другие специалисты подробно описывали её, указывая авторами Л. Рида и У. Фроста (см., например, статьи [5, 6] или монографию [7]).

В конце 80-х годов прошлого века, когда компьютеры и ИТ-технологии стали все шире и шире применяться во всех отраслях промышленности, а затем стали играть значительную роль и в личной жизни людей, возникла новая угроза – повсеместно стали распространяться вредоносные компьютерные программы, получившие название «компьютерных вирусов» из-за схожего с биологическими вирусами механизма распространения. Поэтому специалисты сразу же стали адаптировать модели распространения биологических вирусов и применять их для анализа эпидемий компьютерных вирусов. В 1987 году Ф. Коэном [8], а затем в 1990-х Дж. Кефартом и С. Уайтом [9, 10] были описаны первые математические модели распространения компьютерных вирусов. В данных моделях с успехом были применены ранее известные подходы и методы математической эпидемиологии.

Обширный перечень существующих математических моделей вирусных эпидемий делится на два класса: непрерывные и дискретные [11].

Непрерывные модели используют для описания эпидемий аппарат дифференциальных уравнений и допущение, что время и количество инфицированных и здоровых индивидов являются непрерывными переменными. С помощью таких моделей можно достаточно точно составить долгосрочный прогноз эпидемического процесса, а также оценить условия смены эпидемических режимов [12, 13]. Однако при описании начального этапа распространения вируса у таких моделей обнаруживается существенная погрешность по сравнению с экспериментальными данными. К тому же в них никак не учитывается особенность связей между индивидами в популяции.

Используя так называемые «дискретные» модели, можно получить более детальную картину, что особенно важно на начальной стадии эпидемии [14, 15]. Как следует из названия метода, время и количество инфицированных и здоровых индивидов являются в

данных моделях дискретными величинами, а популяция (компьютерная сеть) представляется в виде некоторого графа. Распространение вируса в таких моделях представляется как некоторый случайный процесс, ассоциированный с рассматриваемым графом.

В работе [14] представлена одна из дискретных марковских моделей, предназначенная для описания эпидемий компьютерных вирусов в связных сетях. Ранее мы уже проводили анализ некоторых аспектов данной модели [16–19]. Далее будем называть эту модель по первым буквам фамилий авторов – BSS-моделью.

Целью настоящей статьи является исследование BSS-модели в предельном режиме, когда лечение отсутствует, и получение приближенной формулы среднего времени инфицирования.

Структура настоящей статьи следующая. В первом разделе приводится подробное описание BSS-модели, после чего во втором разделе эта модель рассматривается при отсутствии лечения. Третий раздел будет посвящен вычислению среднего времени инфицирования сети. В четвертом – приведено описание проведенного вычислительного эксперимента. Заключение будет посвящено обобщению полученных результатов.

Описание BSS-модели. BSS-модель описывает процесс распространения вируса в компьютерных сетях, ассоциированных со связными графами. В её основу была положена модель Рида-Фроста, упомянутая нами выше [5–7], которая ранее применялась только для описания эпидемий биологических вирусов.

Далее приведем детальное описание BSS-модели.

Компьютерная сеть представляется в виде связного графа с N узлами. Узлами (вершинами) графа являются компьютеры, а возможные связи между ними – это ребра графа. Среди всех параметров, характеризующих данный граф, важным с точки зрения данной модели является связность сети, вычисляемая как $c = \langle k \rangle / N$, где $\langle k \rangle$ – средняя степень вершины:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N \frac{de g(i)}{N}.$$

Каждый из узлов сети может быть в одном из двух состояний: либо *восприимчивый* (**S**), либо *инфицированный* (**I**). Время предполагается дискретным, то есть изменения состояний узлов сети происходят в заданные моменты времени: $t = 0, 1, 2, \dots$. Таким образом, в каждый следующий момент времени восприимчивый узел может остаться восприимчивым или заразиться, а инфицированный – вылечиться или остаться инфицированным. Вероятность этих переходов задается следующими характеристиками:

δ – вероятность перехода **I** → **S** (вероятность лечения), которая является постоянной величиной, отражающей эффективность антивирусного программного обеспечения;

μ – вероятность перехода **S** → **I** (вероятность заражения), которая зависит от числа заражённых узлов.

Схема переходов между состояниями показана на рис. 1.

Количество инфицированных узлов I в каждый момент времени t объявляется состоянием сети. Поскольку число узлов S в восприимчивом состоянии в любой момент времени равно $N - I$, это число определяется состоянием сети однозначно. Таким образом, всего имеется $N + 1$ различных состояний сети: $I = 0, 1, \dots, N$.

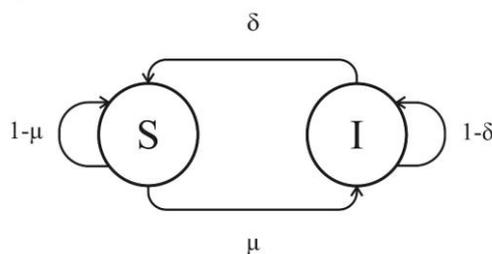


Рис. 1. Схема переходов между состояниями отдельного узла сети

В основу BSS-модели закладывается предположение о том, что каждое последующее состояние сети определяется только лишь предыдущим состоянием (свойство марковости). Тем самым динамика переходов между состояниями представляет собой некоторую *марковскую цепь* $\{X_t\}_{t \geq 0}$, для описания которой необходимо вычислить вероятности $\pi_{I,I'} = \Pr(X_{t+1} = I' | X_t = I)$ переходов между всевозможными парами состояний (I, I') . В работе [14] приводится явная формула для этой вероятности, которая имеет вид

$$\pi_{I,I'} = \sum_{x=\max\{0, I-I'\}}^{\min\{I, N-I'\}} C_I^x (\delta^x) (1 - \delta)^{I-x} C_{N-I}^{I'-I+x} \mu^{I'-I+x} (1 - \mu)^{N-I'-x}. \quad (1)$$

Здесь C_I^x – биномиальный коэффициент, $\mu(I)$ – вероятность восприимчивому узлу стать инфицированным, которая зависит от числа I инфицированных узлов и вычисляется как

$$\mu(I) = 1 - (1 - \beta c)^I, \quad (2)$$

где β – вероятность того, что инфицированный узел передаст инфекцию соседнему восприимчивому узлу. Как показывает данная формула, вероятность $\mu(I)$ зависит от β и c только через их произведение, поэтому вместо двух параметров β и c мы введём один параметр $a = \beta c$.

Вероятность $p_I(t)$ любого состояния I в произвольный момент времени t может быть найдена с помощью выражения

$$p_I(t) = \sum_{I'=0}^N p_{I'}(0) \pi_{I',I}^{(t)}, \quad (3)$$

где $\pi_{I',I}^{(t)}$ – вероятность перехода из состояния I' в состояние I ровно за t шагов, а $p_{I'}(0)$ – вероятность состояния I' в начальный момент времени. Как следствие, среднее число заражённых узлов в момент времени t будет равно

$$\langle I \rangle(t) = \sum_{I=0}^N p_{I'}(t) I.$$

В работе [14] авторы BSS-модели рассмотрели так же ее «непрерывную» версию, после чего они провели сравнение результатов этих двух моделей с результатами имитационного эксперимента. В итоге ими были получены условия, при которых данные всех трёх подходов хорошо согласуются друг с другом, а также условия, когда между ними имеется расхождение, в частности, вблизи так называемого *эпидемического порога*.

BSS-модель в отсутствии лечения. Описанная выше BSS-модель позволяет в принципе осуществить количественный анализ эпидемии, однако вычисление вероятностей состояний $p_I(t)$ с помощью равенства (3) вызывает значительные вычислительные трудности для графов с большим количеством узлов N . С другой стороны, исследование динамики эпидемий как раз и представляет особый интерес на больших графах с числом узлов $N \gtrsim 10^5$. В этой связи в данном разделе мы рассмотрим частный, но важный случай этой модели, когда в ней отсутствует переход $\mathbf{I} \rightarrow \mathbf{S}$, то есть $\delta = 0$. Данная ситуация может быть рассмотрена как приближение к начальной фазе эпидемии, либо как модель, описывающая эпидемии вирусов, плохо поддающихся лечению.

В случае, когда $\delta = 0$, формула (1) для переходных вероятностей состояний сети упрощается:

$$\pi_{I,I'} = \begin{cases} C_{N-I}^{I'-I} \mu(I)^{I'-I} (1 - \mu(I))^{N-I'}, & I' \geq I, \\ 0, & I' < I. \end{cases} \quad (4)$$

Здесь вероятность $\mu(I)$, как и ранее, определяется формулой (2). То, что $\pi_{I,I'} = 0$ при $I > I'$, согласуется с тем фактом, что в отсутствии лечения число инфицированных узлов в сети не может уменьшаться. Поскольку $\pi_{0,I'} = \delta_{0,I'}$, далее нет смысла рассматривать состояние $I = 0$, поэтому с этого момента мы будем считать, что возможные значения I и I' лежат в диапазоне от 1 до N .

Отметим, что формулу (4) можно также записать следующим образом:

$$\pi_{I,I+\Delta I} = \begin{cases} C_{N-I}^{\Delta I} \mu(I)^{\Delta I} (1 - \mu(I))^{N-I-\Delta I}, & \Delta I \geq 0, \\ 0, & \Delta I < 0, \end{cases}$$

где $\Delta I = I' - I$. Отсюда видно, что при фиксированном I величина $\Delta I = I' - I$, представляющая собой приращение заражённых узлов за единицу времени, распределена по биномиальному закону $\mathcal{B}(p, n)$ с параметрами $n = N - I$ и $p = \mu(I)$. Тем самым, матрица переходных вероятностей $\Pi = (\pi_{I,I'})$ в случае $\delta = 0$ является верхней треугольной матрицей, каждая строка которой представляет собой полный ряд вероятностей биномиального распределения в соответствии с формулой (4).

Заметим что, в отсутствии лечения, стартуя из некоторого состояния $X_0 = I_0$, где $0 < I_0 < N$, марковская цепь $\{X_t\}$ рано или поздно попадёт в состояние $I = N$, в котором все узлы сети будут заражены. Таким образом, данная цепь имеет поглощающее состояние $I = N$. Наличие такого состояния позволяет ввести важную количественную характеристику вирусной эпидемии – время инфицирования. По определению, временем инфицирования называется количество переходов T в соответствующей марковской цепи с момента $t = 0$ до момента $t = T$ попадания в поглощающее состояние $I = N$. Очевидно, что определённое таким образом время инфицирования является случайной величиной, имеющей некоторое распределение вероятностей. Для практических целей разумно рассматривать среднее время инфицирования τ , представляющее собой математическое ожидание случайной величины T . Ясно, что среднее время инфицирования является функцией параметров модели N , $\alpha = \beta c$ и I_0 :

$$\tau = \tau(\alpha, N, I_0). \quad (5)$$

Общий алгоритм вычисления величины τ хорошо известен [20]. Напомним его.

Состояние марковской цепи называется переходным, если с единичной вероятностью цепь посетит его только конечное число раз. Очевидно, что состояния $I = 1, \dots, N - 1$ в нашей марковской цепи $\{X_t\}$ являются переходными. Обозначим через Q подматрицу матрицы Π , включающую в себя только те строки и столбцы из Π , которые отвечают переходным состояниям. Ясно, что Q – это $(N - 1) \times (N - 1)$ -матрица, получающаяся из Π вычёркиванием N -ой строки и N -го столбца.

Так как все состояния, представляемые матрицей Q , переходные, $Q^n \rightarrow 0$ при $n \rightarrow \infty$. Это значит, что собственные значения матрицы Q по абсолютной величине строго меньше единицы, что влечёт обратимость матрицы $E - Q$. Рассмотрим матрицу

$$M = (E - Q)^{-1} = E + Q + Q^2 + Q^3 + \dots,$$

называемую фундаментальной матрицей марковской цепи. Смысл элементов этой матрицы следующий: (I, I') -элемент матрицы M равен среднему числу визитов состояния I , если стартовым является состояние I' . Таким образом, если мы стартуем из состояния I_0 , среднее время инфицирования будет равно сумме элементов строки I_0 фундаментальной матрицы M :

$$\tau = \sum_{I=1}^{N-1} M_{I_0, I}.$$

Данная формула позволяет нам вычислить среднее время инфицирования τ , однако при её применении могут возникать некоторые сложности.

Во-первых, как уже упоминалось выше, наиболее интересным с практической точки зрения является вычисление τ на графах с большим числом узлов $N \gtrsim 10^5$. Вычисление же обратных матриц в таких случаях затруднительно, в силу больших размеров соответствующих матриц. Во-вторых, элементы матрицы Q включают в себя биномиальные коэффициенты, расчет которых для больших N также имеет определенные вычислительные сложности, связанные с факториалами очень больших чисел. Кроме того, в выражениях для элементов матрицы Q также могут появляться очень большие степени очень маленьких величин, выходящие за пределы разрешенного диапазона представимых на ЭВМ чисел, что в итоге будет приводить к ошибкам вычисления.

Перечисленные трудности приводят к необходимости разработки более удобных алгоритмов вычисления среднего времени инфицирования. При этом вывод явного и точного выражения для функции $\tau(\alpha, N, I_0)$ вряд ли возможен, поэтому более перспективным является получение приближённой аналитической формулы для τ , при дополнительном предположении о малости параметра α , что обычно и имеет место на практике.

Среднее время инфицирования при малых α . В настоящем разделе мы рассмотрим BSS-модель в отсутствие лечения при $\alpha \ll 1$. Кроме того, мы будем считать, что эпидемия начинается с одного единственного заражённого узла: $I_0 = 1$.

Из формулы (4) нетрудно видеть, что при малых значениях α с точностью до слагаемых третьего и выше порядков мы получаем:

$$\begin{aligned} \pi_{I,I} &\approx 1 - I(N - I)\alpha + \frac{1}{2}I(N - I)(NI - I^2 - 1)\alpha^2, \\ \pi_{I,I+1} &\approx I(N - I)\alpha + \frac{1}{2}I(N - I)(I + 1 - 2I(N - I))\alpha^2, \\ \pi_{I,I+2} &\approx \frac{1}{2}I^2(N - I)(N - I - 1)\alpha^2. \end{aligned}$$

Остальные матричные элементы имеют третий и выше порядок малости по α , поэтому мы их игнорируем. Таким образом, с точностью до членов порядка α^2 для матричных элементов $\pi_{I,I'}$ при $I' \geq I$ мы можем записать:

$$\begin{aligned} \pi_{I,I'} &\approx \delta_{I,I'} + I(N - I)(\delta_{I+1,I'} - \delta_{I,I'})\alpha + \\ &+ \frac{1}{2}I(N - I)[(NI - I^2 - 1)\delta_{I,I'} + (I + 1 - 2NI + 2I^2)\delta_{I+1,I'} + I(N - I - 1)]\alpha^2, \end{aligned} \quad (7)$$

где $\delta_{I,I'}$ – символ Кронекера. Из полученной формулы (7) следует, что матрицу Q мы можем приближённо записать как

$$Q \approx E + \alpha Q_1 + \alpha^2 Q_2,$$

где E – единичная матрица порядка N_1 , а независящие от α матрицы Q_1 и Q_2 имеют ленточный вид. В частности, матрица Q_1 имеет следующий явный вид

$$Q_1 = \begin{pmatrix} -(N - 1) & N - 1 & 0 & 0 & \dots & 0 \\ 0 & -2(N - 2) & 2(N - 2) & 0 & \dots & 0 \\ 0 & 0 & -3(N - 3) & 3(N - 3) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -(N - 1) \end{pmatrix}.$$

В то же время матрицу Q_2 мы здесь не приводим из-за её достаточно громоздкого вида. Далее, будем искать фундаментальную матрицу M , обратную к матрице $E - Q$, в виде $M \approx \alpha^{-1}M_{-1} + M_0$, где M_{-1} и M_0 – не зависящие от α неизвестные пока матрицы порядка $N - 1$. Подобная приближённая форма матрицы M связана с особым характером матрицы $E - Q$ при $\alpha = 0$.

По определению обратной матрицы должно быть $(E - Q)M \approx E$, или в явном виде

$$(-\alpha Q_1 - \alpha^2 Q_2)(\alpha^{-1}M_{-1} + M_0) = E.$$

Раскрывая скобки и приравнивая подобные слагаемые при степенях α^0 и α^1 получаем:

$$-Q_1 M_{-1} = E, \quad Q_1 M_0 + Q_2 M_{-1} = 0.$$

Таким образом, для фундаментальной матрицы M мы получаем следующее приближённое равенство:

$$M \approx -\frac{1}{\alpha} Q_1^{-1} + Q_1^{-1} Q_2 Q_1^{-1}. \quad (8)$$

Обратная матрица Q_1^{-1} легко вычисляется и имеет вид:

$$Q_1^{-1} = - \begin{pmatrix} \frac{1}{N - 1} & \frac{1}{2(N - 2)} & \frac{1}{3(N - 3)} & \dots & \frac{1}{N - 1} \\ 0 & \frac{1}{2(N - 2)} & \frac{1}{3(N - 3)} & \dots & 0 \\ 0 & 0 & \frac{1}{3(N - 3)} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{N - 1} \end{pmatrix}.$$

Отметим теперь, что в соответствии с формулой (7):

$$(Q_2)_{I,I} = \frac{1}{2}I(N-I)(NI - I^2 - 1), \quad (Q_2)_{I,I+1} = \frac{1}{2}I(N-I)(I+1 - 2NI + 2I^2),$$

$$(Q_2)_{I,I+2} = \frac{1}{2}I(N-I)I(N-I-1),$$

а все остальные элементы матрицы Q_2 равны нулю. С другой стороны, нетрудно видеть, что первая строка матрицы $Q_1^{-1} Q_2 Q_1^{-1}$ будет иметь следующие элементы:

$$(Q_1^{-1} Q_2 Q_1^{-1})_{1,1} = \frac{(Q_2)_{1,1}}{(N-1)^2},$$

$$(Q_1^{-1} Q_2 Q_1^{-1})_{1,2} = \frac{(Q_2)_{1,1} + (Q_2)_{1,2}}{2(N-1)(N-2)} + \frac{(Q_2)_{2,2}}{4(N-2)^2},$$

$$(Q_1^{-1} Q_2 Q_1^{-1})_{1,3} = \frac{(Q_2)_{1,1} + (Q_2)_{1,2} + (Q_2)_{1,3}}{3(N-1)(N-3)} + \frac{(Q_2)_{2,2} + (Q_2)_{2,3}}{6(N-2)(N-3)} + \frac{(Q_2)_{3,3}}{9(N-3)^2},$$

...

Замечая теперь, что

$$(Q_2)_{I,I} + (Q_2)_{I,I+1} + (Q_2)_{I,I+2} = 0, \quad I = 1, 2, 3, \dots, N-3,$$

можно выписать общую формулу для элемента $(Q_1^{-1} Q_2 Q_1^{-1})_{1,I}$:

$$(Q_1^{-1} Q_2 Q_1^{-1})_{1,I} = \frac{(Q_2)_{I-1,I-1} + (Q_2)_{I-1,I}}{I(I-1)(N-I)(N-I+1)} + \frac{(Q_2)_{I,I}}{I^2(N-I)^2}, \quad I = 2, 3, \dots, N-1.$$

Подставляя сюда явные выражения для элементов матрицы Q_2 , после упрощения в итоге получаем:

$$(M_0)_{1,I} = (Q_1^{-1} Q_2 Q_1^{-1})_{1,I} = \frac{1}{2I} \left(1 - \frac{1}{N-I} \right).$$

Отметим, что эта формула верна для всех $I = 1, 2, \dots, N-1$.

Используя полученный результат для элементов первой строки фундаментальной матрицы M в соответствии с формулой (8) мы получаем следующее приближённое выражение:

$$M_{1,I} \approx \frac{1}{\alpha I(N-I)} + \frac{1}{2I} \left(1 - \frac{1}{N-I} \right), \quad I = 1, 2, 3, \dots, N-1.$$

Как мы уже отмечали в предыдущем разделе, среднее время инфицирования τ равно сумме элементов I_0 -ой строки фундаментальной матрицы M . Учитывая, что в нашем случае $I_0 = 1$ и применив полезную формулу

$$\sum_{k=1}^{n-1} \frac{1}{k(n-k)} = \frac{2H_{n-1}}{n},$$

где H_n – n -ое гармоническое число, для среднего времени инфицирования получаем следующую приближённую формулу:

$$\tau \approx \left(\frac{1}{\alpha N} + \frac{1}{2} - \frac{1}{N} \right) H_{N-1}. \quad (9)$$

В этой формуле мы игнорируем все слагаемые, имеющие второй и высшие порядки по α . Формула (9) – основной результат настоящего раздела.

Обсудим полученную формулу (9). Во-первых, из неё видно, что если $\alpha \rightarrow 0$, то $\tau \rightarrow \infty$, что отражает следующий очевидный факт: при малых α и/или β инфекция распространится на всю сеть через очень большой промежуток времени. Во-вторых, из этой формулы также следует, что с ростом α среднее время инфицирования τ уменьшается, что тоже понятно из общих соображений.

Гораздо интереснее поведение τ в зависимости от параметра N , т.е. от размера графа. С одной стороны, выражение в круглых скобках справа в формуле (9) уменьшается с ростом N , стремясь к значению $\frac{1}{2}$. С другой стороны, второй множитель в этой формуле

– гармоническое число H_{N-1} – медленно увеличивается с ростом N по логарифмическому закону. В результате этого их произведение, то есть среднее время инфицирования τ , ведет себя не монотонным образом: сначала возрастает, достигая некоторого максимума, потом убывает, а затем снова начинает медленно расти по логарифмическому закону.

Вычислительный эксперимент. Полученная нами формула (9) для среднего времени инфицирования является приближенной, причём это приближение тем лучше, чем меньше значение α . С другой стороны, неясно насколько данное приближение близко к истинному значению τ при различных значениях N . В данном разделе мы исследуем как согласуются между собой результаты, полученные тремя разными способами:

- 1) точным вычислением фундаментальной матрицы M и применением формулы (6);
- 2) с помощью приближенной формулы (9);
- 3) методом имитационного моделирования.

Предварительно, опишем подробнее метод имитационного моделирования, используемый нами для вычисления среднего времени инфицирования.

В начале мы генерируем случайный граф с заданным количеством узлов N и заданной связностью c . В нашем исследовании мы используем случайные графы класса Эрдёша-Реньи, поскольку методы генерации этих графов известны, а их свойства хорошо изучены [21]. На первом шаге мы заражаем один случайный узел (напомним, что $I_0 = 1$). На втором шаге этот узел заражает соседние узлы с заданной вероятностью передачи вируса β . В свою очередь, на третьем шаге уже зараженные узлы инфицируют связанные с ними узлы также с вероятностью β и т.д. Описанные шаги повторяются до тех пор, пока инфекция не распространится на все узлы графа. После этого подсчитывается количество временных шагов от момента $t = 0$ до момента $t = T$ заражения последних восприимчивых узлов. Таким образом мы получили время инфицирования T , то есть время попадания в поглощающее состояние $I = N$. Далее мы генерируем новый граф с теми же параметрами N и c , и повторяем описанные выше шаги до момента заражения всех узлов. В результате получаем новое значение времени инфицирования T . Эксперимент повторяется заданное количество раз (в нашем случае – 1000), после чего полученный массив T усредняется, то есть вычисляется среднее время инфицирования τ при заданных параметрах N и $\alpha = \beta c$. Затем процесс повторяется для других значений N и α . Данный алгоритм был реализован в виде консольного приложения, написанного на языке программирования C++ (на эту программу нами получено свидетельство о регистрации [17]).

С использованием разработанной нами программы проведено две серии вычислительных экспериментов при различных значениях параметра α : $\alpha = 0,001$ и $\alpha = 0,00325$, в каждой из которых количество узлов графа изменялось от $N_1 = 100$ до $N_{35} = 3500$ с шагом $\Delta N = 100$.

В каждом эксперименте оценивалось среднее время инфицирования τ . Эта же величина при тех же входных значениях N и α вычислялась нами и двумя другими методами: с помощью применения прямого марковского подхода (формула (6)) и в соответствии с приближенным результатом (9). Результаты сравнения всех трех методов представлены на рис. 2 и 3. На этих рисунках приведены графики зависимости среднего времени инфицирования τ от числа узлов в графе N при фиксированном значении α . На графиках синяя сплошная линия получена по формуле (6), пунктирная серая — по приближенной формуле (9), а треугольниками обозначены значения τ , полученные имитационным моделированием.

Как видно из обоих рисунков, среднее время инфицирования τ , вычисленное прямым способом с помощью формулы (6), довольно неплохо согласуется с результатами имитационного эксперимента, откуда можно сделать вывод, что случайные графы Эрдёша-Реньи хорошо воспроизводят результаты BSS-модели, построенной на основе марковских цепей. С другой стороны, приближенная формула (9) достаточно близко воспроизводит результаты прямого подхода в случае $\alpha = 0,001$, в то время как при $\alpha = 0,00325$ приближенные значения τ начинают немного расходиться с соответствующими точными значениями, особенно, при больших N . Этого и следовало ожидать, так как приближен-

ная формула (9) получена нами при условии, что α достаточно близко к нулю. Тем не менее надо отметить, что величина соответствующей погрешности $\Delta\tau$ стабильна и практически не увеличивается с ростом N .

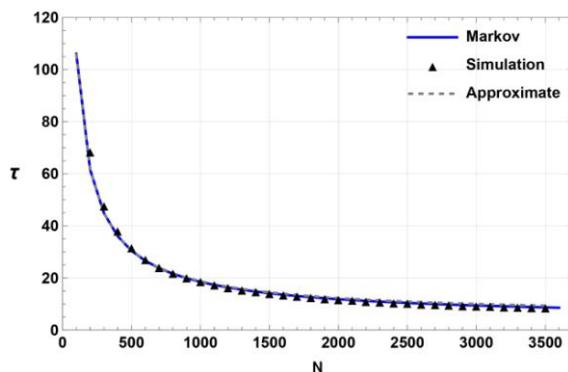


Рис. 2: Зависимость среднего времени инфицирования τ от N при $\alpha = 0,001$

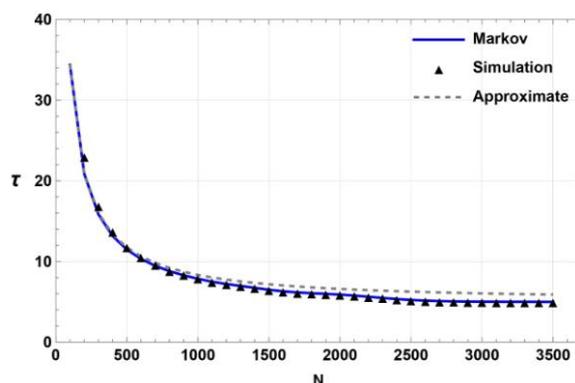


Рис. 3: Зависимость среднего времени инфицирования τ от N при $\alpha = 0,00325$

Заключение. В данной статье нами была рассмотрена одна из дискретных марковских моделей [14] в предельном случае, когда процесс распространения вируса происходит в отсутствие лечения. Основным результатом является полученная приближенная формула для расчета среднего времени инфицирования τ , то есть длительности эпидемии. Данная формула является приближенной аналитической оценкой длительности эпидемии и представляет собой удобную альтернативу прямому способу получения данного параметра, связанного с вычислением фундаментальной матрицы M соответствующей марковской цепи (6). Отметим, что вычисление данной матрицы сопряжено с большими вычислительными трудностями особенно при больших размерах сети (популяции). В последнем разделе статьи мы сравнили три подхода к определению среднего времени инфицирования: метод точного вычисления фундаментальной матрицы M , вычисление с применением приближенной формулы и метод имитационного моделирования. Результаты сравнения показали, что все три метода хорошо согласуются при достаточно малых значениях α .

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Kermack W. O., McKendrick A. G.* A contribution to the mathematical theory of epidemics // Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character. – 1927. – Vol. 115, No. 772. – P. 700-721.
2. *Persoons R., Van Mieghem P.* Finding patient zero in susceptible-infectious-susceptible epidemic processes // Physical Review E. – 2024. – Vol. 110, No. 4. – P. 044308.

3. *Achterberg M. A., Van Mieghem P.* Moment closure approximations of susceptible-infected-susceptible epidemics on adaptive networks // *Physical Review E*. – 2022. – Vol. 106, No. 1. – P. 014308.
4. *Notarmuzi D. et al.* Critical avalanches of susceptible-infected-susceptible dynamics in finite networks // *Physical Review E*. – 2023. – Vol. 107, No. 2. – P. 024310.
5. *Abbey H.* An examination of the Reed-Frost theory of epidemics // *Human biology*. – 1952. – Vol. 24, No. 3. – P. 201.
6. *Fine P.E.M.* A commentary on the mechanical analogue to the Reed-Frost epidemic model // *American journal of epidemiology*. – 1977. – Vol. 106, No. 2. – P. 87-100.
7. *Hoppensteadt F.C.* *Mathematical methods of population biology*. – Cambridge University Press, 1982. – No. 4.
8. *Cohen F.* Computer viruses: theory and experiments // *Computers & security*. – 1987. – Vol. 6, No. 1. – P. 22-35. – DOI: 10.1016/0167-4048(87)90122-2.
9. *Kephart J.O., White S.R.* Directed-graph epidemiological models of computer viruses // *Computation: the micro and the macro view*. – 1992. – P. 71-102.
10. *Kephart J.O., White S.R.* Measuring and modeling computer virus prevalence // *Proceedings 1993 IEEE Computer Society Symposium on Research in Security and Privacy*. – IEEE, 1993. – P. 2-15. – DOI: 10.1109/RISP.1993.287647.
11. *Pastor-Satorras R. et al.* Epidemic processes in complex networks // *Reviews of modern physics*. – 2015. – Vol. 87, No. 3. – P. 925-979.
12. *Granger T. et al.* Stochastic compartment model with mortality and its application to epidemic spreading in complex networks // *Entropy*. – 2024. – Vol. 26, No. 5. – P. 362. – DOI: 10.3390/e26050362.
13. *Singh P., Gupta A.* Generalized SIR (GSIR) epidemic model: An improved framework for the predictive monitoring of COVID-19 pandemic // *ISA transactions*. – 2022. – Vol. 124. – P. 31-40.
14. *Billings L., Spears W.M., Schwartz I.B.* A unified prediction of computer virus spread in connected networks // *Physics Letters A*. – 2002. – Vol. 297, No. 3-4. – P. 261-266.
15. *Далингер Я.М., Бабанин Д.В., Бурков С.М.* Математические модели распространения вирусов в компьютерных сетях различной структуры // *Информатика и системы управления*. – 2011. – № 4. – С. 3-11.
16. *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. – № 1 (61). – С. 92-104.
17. *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.
18. *Никифорова А.Ю.* Приближённая оценка условий прекращения эпидемии компьютерного вируса в связных сетях, ассоциированных со случайными графами // *Моделирование, оптимизация и информационные технологии*. – 2023. – Т. 11, № 4 (43). – DOI: 10.26102/2310-6018/2023.43.4.034.
19. *Magazev A.A., Nikiforova A.Y.* On the Applicability of a Markov Virus Spread Model to E-mail Graphs // *2023 Dynamics of Systems, Mechanisms and Machines (Dynamics)*. – IEEE, 2023. – P. 1-4. – DOI: 10.26102/2310-6018/2023.43.4.034.
20. *Lawler G. F.* *Introduction to stochastic processes*. – Chapman and Hall/CRC, 2018. – 234 p.
21. *Erdos P., Renyi A.* On Random Graphs // *Publicationes Mathematicae (Debrecen)*. – 1959. – Vol. 6. – P. 290-297.

REFERENCES

1. *Kermack W. O., McKendrick A. G.* A contribution to the mathematical theory of epidemics, *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 1927, Vol. 115, No. 772, pp. 700-721.
2. *Persoons R., Van Mieghem P.* Finding patient zero in susceptible-infectious-susceptible epidemic processes, *Physical Review E*, 2024, Vol. 110, No. 4, pp. 044308.
3. *Achterberg M. A., Van Mieghem P.* Moment closure approximations of susceptible-infected-susceptible epidemics on adaptive networks, *Physical Review E*, 2022, Vol. 106, No. 1, pp. 014308.
4. *Notarmuzi D. et al.* Critical avalanches of susceptible-infected-susceptible dynamics in finite networks, *Physical Review E*, 2023, Vol. 107, No. 2, pp. 024310.
5. *Abbey H.* An examination of the Reed-Frost theory of epidemics, *Human biology*, 1952, Vol. 24, No. 3, pp. 201.
6. *Fine P.E.M.* A commentary on the mechanical analogue to the Reed-Frost epidemic model, *American journal of epidemiology*, 1977, Vol. 106, No. 2, pp. 87-100.
7. *Hoppensteadt F.C.* *Mathematical methods of population biology*. Cambridge University Press, 1982. – No. 4.

8. Cohen F. Computer viruses: theory and experiments, *Computers & security*, 1987, Vol. 6, No. 1, pp. 22-35. DOI: 10.1016/0167-4048(87)90122-2.
9. Kephart J.O., White S.R. Directed-graph epidemiological models of computer viruses, *Computation: the micro and the macro view*, 1992, pp. 71-102.
10. Kephart J.O., White S.R. Measuring and modeling computer virus prevalence, *Proceedings 1993 IEEE Computer Society Symposium on Research in Security and Privacy*. IEEE, 1993, pp. 2-15. DOI: 10.1109/RISP.1993.287647.
11. Pastor-Satorras R. et al. Epidemic processes in complex networks, *Reviews of modern physics*, 2015, Vol. 87, No. 3, pp. 925-979.
12. Granger T. et al. Stochastic compartment model with mortality and its application to epidemic spreading in complex networks, *Entropy*, 2024, Vol. 26, No. 5, pp. 362. DOI: 10.3390/e26050362.
13. Singh P., Gupta A. Generalized SIR (GSIR) epidemic model: An improved framework for the predictive monitoring of COVID-19 pandemic, *ISA transactions*, 2022, Vol. 124, pp. 31-40.
14. Billings L., Spears W.M., Schwartz I.B. A unified prediction of computer virus spread in connected networks, *Physics Letters A*, 2002, Vol. 297, No. 3-4, pp. 261-266.
15. Dalingер YA.M., Babanin D.V., Burkov S.M. Matematicheskie modeli rasprostraneniya virusov v komp'yuternykh setyakh razlichnoy struktury [The mathematical models of the spreading of viruses in computer networks with the diferent structures], *Informatika i sistemy upravleniya* [Computer science and control systems], 2011, No. 4, pp. 3-11.
16. Bel'chenko A.O., Magazev A.A., Nikiforova A.Yu. Priblizhennaya otsenka srednego chisla zarazhennykh uzlov v markovskoy modeli rasprostraneniya komp'yuternykh virusov [An approximate evaluation of the infected nodes number for a Markov model of viruses spreading], *Matematicheskie struktury i modelirovanie* [Mathematical Structures and Modeling], 2022, No. 1 (61), pp. 92-104.
17. Magazev A.A., Nikiforova A.Yu. Programma dlya otsenki srednego vremeni rasprostraneniya komp'yuternogo virusa v setyakh, assotsirovannykh so sluchaynymi grafami: svidetel'stvo o registratsii elektronnoho resursa [A program for estimating the average spread time of a computer virus in networks associated with random graphs]. Moscow: FIPS, 2023. Patent RF № 2023614819 from 06.03.2023.
18. Nikiforova A.Yu. Priblizhennaya otsenka usloviy prekrashcheniya epidemii komp'yuternogo virusa v svyaznykh setyakh, assotsirovannykh so sluchaynymi grafami [An approximate evaluation of the conditions for the termination of a computer virus epidemic in connected networks associated with random graphs], *Modelirovanie, optimizatsiya i informatsionnye tekhnologii* [The scientific journal Modeling, Optimization and Information Technology], 2023, Vol. 11, No. 4 (43). DOI: 10.26102/2310-6018/2023.43.4.034.
19. Magazev A.A., Nikiforova A.Y. On the Applicability of a Markov Virus Spread Model to E-mail Graphs, *2023 Dynamics of Systems, Mechanisms and Machines (Dynamics)*. IEEE, 2023, pp. 1-4. DOI: 10.26102/2310-6018/2023.43.4.034.
20. Lawler G. F. Introduction to stochastic processes. Chapman and Hall/CRC, 2018, 234 p.
21. Erdos P., Renyi A. On Random Graphs, *Publicationes Mathematicae (Debrecen)*, 1959, Vol. 6, pp. 290-297.

Магазев Алексей Анатольевич – Омский государственный технический университет; e-mail: magazev@mail.ru; г. Омск, Россия; кафедра «Комплексная защита информации»; профессор; ORCID 0000-0002-8725-9183; Scopus Author ID 6507004666.

Никифорова Ангелина Юрьевна – Омский государственный технический университет; e-mail: skt-omgtu@mail.ru; г. Омск, Россия; кафедра «Комплексная защита информации»; старший преподаватель; ORCID: 0000-0002-0981-7127.

Magazev Alexey Anatolyevich – Omsk State Technical University, e-mail: magazev@mail.ru; Omsk, Russia; the Department of Comprehensive Information Security; professor, ORCID 0000-0002-8725-9183; Scopus Author ID 6507004666.

Nikiforova Angelina Yurevna – Omsk State Technical University; e-mail: skt-omgtu@mail.ru; Omsk, Russia; the Department of Comprehensive Information Security; Senior lecturer; ORCID 0000-0002-0981-7127.