

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теории функций и стохастического анализа

**МОДЕЛИРОВАНИЕ РАСПРОСТРАНЕНИЯ ЭПИДЕМИЙ В  
ГРАФАХ**

**АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ**

Студентки 2 курса 248 группы  
направления 09.04.03 — Прикладная информатика

механико-математического факультета

Арюковой Татьяны Сергеевны

Научный руководитель

доцент, к. э. н.

\_\_\_\_\_

А. Р. Файзлиев

Заведующий кафедрой

д. ф.-м. н., доцент

\_\_\_\_\_

С. П. Сидоров

Саратов 2021

## ВВЕДЕНИЕ

**Актуальность темы.** В повседневной жизни нас окружает огромное количество вирусов, которые могут нанести огромный вред всем живым существам. Но этого можно избежать, если контролировать ситуацию и вовремя принимать необходимые меры. Пандемия коронавируса не стала исключением в данной ситуации, и что бы избежать еще больших потрясений, необходимо понимать, какие способы в большей мере подходят для моделирования данной эпидемии, что бы избежать отрицательных последствий.

Данная работа представляет интерес поскольку в настоящее время эпидемия коронавируса не утихла и требует пристального внимания.

**Целью** магистерской работы является моделирование и анализ распространение эпидемии коронавируса в графах.

**Объект исследования** - вспышки пандемии и ее последствия.

**Предмет исследования** - методы моделирования эпидемий и анализ полученных результатов.

Для достижения поставленных целей в работе необходимо решить следующие **задачи**:

- рассмотреть некоторые эпидемиологические подходы моделирования;
- изучить стратегии многошагового прогнозирования временных рядов;
- изучить необходимую информацию о пандемии коронавируса;
- определить инструменты для исследования;
- построить и проанализировать модель распространения пандемии в Москве;
- провести численные эксперимент с моделью SEIR.

**Практическая значимость** производимого исследования заключается в создании математической модели распространения эпидемии и ее анализа, для получения результатов на существующих данных. По результатам работы можно сделать вывод о предстоящих мерах борьбы с пандемией.

**Структура и содержание магистерской работы.** Работа состоит из введения, 8 разделов, заключения, списка использованных источников, состоящий из 25 наименований и одного приложения. Общий объем работы составляет 55 страниц.

## Основное содержание работы

Во **введении** обосновывается актуальность темы работы, формулируется цель работы и решаемые задачи, отмечается практическая значимость полученных результатов.

В **первом** разделе ставится задача предстоящей работы.

Во **втором** разделе производится обзор эпидемиологических подходов. Классическая модель SI основывается на утверждении: для любого промежутка времени верно, что количество человек, присоединившихся к больным, равно количеству человек, переставших быть здоровыми.

Модель SI для графа "Мир тесен" применяется в случаях, когда в каждый момент времени здоровый имеет примерно  $(k)$  контактов с больными, а значит  $(k)$  возможностей заразиться.

Граф «Мир тесен» («Маленький мир») может быть смоделирован следующим способом: изначально берется кольцо из  $N$  вершин, где каждая вершина симметрично смежна с  $2K$  ближайшими соседями; далее для каждой вершины каждое её ребро с вероятностью  $p$  обрывают на другом конце и соединяют с случайной вершиной. Процедура выдает случайный граф, в котором типичное расстояние  $L$  между двумя произвольно выбранными вершинами растёт пропорционально логарифму от числа вершин  $N$ :  $L \propto \log N$ , а также в котором среднее количество соседей  $k = 2K$ .

Модель SI для безмасштабной сети — граф, в котором степени вершин распределены по степенному закону:

$$P(k) \sim k^{-\gamma},$$

где степень вершины — количество непосредственных соседей,  $P(k)$  — доля вершин степени  $k$ ,  $2 < \gamma < 3$  — параметр. Примеры данного вида графов из жизни — граф авиаперелетов, граф финансовых операций, граф соавторов статей.

Симуляция эпидемии на произвольном графе.

Пусть есть социальный граф, каждая вершина которого имеет одну из меток:  $s$  — человек здоров;  $i$  — человек болен;  $\beta$  — вероятность заражения здорового при контакте с больным. Тогда шаг алгоритма распространения болезни реализуется таким образом:

1. Выбираем вершину с меткой  $i$  и генерируем случайное число  $x \in (0,1)$ .
2. Если  $x < \beta$ , то выбираем случайную соседнюю соседнюю вершину. Если ее метка  $- s$ , то меняем ее на  $i$ .

В **третьем** разделе рассматриваются стратегии многошагового прогнозирования временных рядов. Задача многошагового прогнозирования временного ряда (долговременного прогнозирования) состоит в получении прогнозов на  $H$  шагов вперед  $[y_{T+1}, y_{T+2}, \dots, y_{T+H}]$  временного ряда  $[y_1, \dots, y_T]$ , состоящего из  $T$  наблюдений, где  $H > 1$  определяет горизонт прогнозирования (количество шагов вперед для предсказания). Ниже воспользуемся стандартным обозначением, где  $f$  и  $F$  — функции, моделирующие зависимость между прошедшими и будущими наблюдениями,  $d$  соответствует количеству наблюдений прошлого, на которые опирается предсказание, а  $w$  представляет собой слагаемое, включающее ошибку модели и внешние возмущения/шум. Так же произвели сравнение стратегий долговременного прогнозирования временных рядов: итерационный, прямой, DirRec, MIMO и DIRMO. Как было сказано, DirRec - стратегия является комбинацией прямого и итерационного подходов, а DIRMO — комбинацией прямого и MIMO-подхода.

В **четвертом** разделе были рассмотрены более подробно предлагаемые подходы, а именно: основные обозначения, итерационный подход, прямой подход, DirRec, MIMO и DIRMO.

В **пятом** разделе был изучен объект исследования - пандемия коронавируса. А именно, что коронавирусы — это семейство РНК-содержащих вирусов, которые чаще всего вызывают заболевания у животных, а некоторые из них — у людей. У людей инфекция обычно протекает в легкой форме с симптомами ОРВИ, не вызывая тяжелых осложнений. Так же были изучены такие вопросы, как распространение вируса, симптомы, продолжительность болезни и инкубационного периода и кто находится в группе риска.

В **шестом** разделе были подробно изучены такие инструменты для исследования, как Модель SIR, которая получила заслуженную популярность в силу простоты построения и использования. Ее применение позволяет точно моделировать эпидемии гриппа и других заболеваний в больших городах, вводить новые параметры и анализировать разные сценарии.

Система уравнений SIR:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta IS}{N} \\ \frac{dI}{dt} &= \frac{\beta IS}{N} - \tau I \\ \frac{dR}{dt} &= \tau I\end{aligned}$$

Модель SEIR, именно по которой и развиваются по-настоящему опасные эпидемии, поскольку длительный инкубационный период может препятствовать своевременному обнаружению заболевания. В этом случае есть риск, что заболевание охватит значительное число индивидуумов в популяции. Инфекция развивается по схеме «восприимчивые» — «контактные» — «инфицированные» — «выздоровевшие» и описывается системой уравнений:

$$\begin{aligned}\frac{dS}{dt} &= \mu N - \mu S - \beta \frac{I}{N} S \\ \frac{dE}{dt} &= \frac{I}{N} S - (\mu + a) E \\ \frac{dI}{dt} &= a E - (\tau + \mu) I \\ \frac{dR}{dt} &= \tau I - \mu R\end{aligned}$$

Модель SIS - «восприимчивые — инфицированные — восприимчивые» применима при анализе распространения заболеваний, к которым не вырабатывается иммунитет, например гриппа и ОРВИ. Она описывается следующей системой уравнений:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta SI}{N} + \tau I \\ \frac{dI}{dt} &= \frac{\beta SI}{N} - \tau I\end{aligned}$$

Модель MSEIR - построенная для заболевания с инкубационным периодом и учитывающая иммунитет детей, приобретенный внутриутробно, — одна из самых сложных для анализа в силу наличия большого числа независимых параметров. Система уравнений для нее выглядит так:

$$\begin{aligned}\frac{dM}{dT} &= B - \delta M - \mu M \\ \frac{dS}{dT} &= \delta M - \beta SI - \mu S \\ \frac{dE}{dT} &= \beta SI - (\epsilon + \mu) E \\ \frac{dI}{dT} &= \epsilon E - (\tau + \mu) I \\ \frac{dR}{dT} &= \tau I - \mu R\end{aligned}$$

А так же были рассмотрены суперинфекции - процесс, при котором одна (первая) инфекция заражает организм, а после происходит заражение второй инфекцией, устойчивой к лекарствам для первой. Возникающие в клетках процессы могут приводить к мутациям и появлению новых видов инфекций и штаммов вирусов.

**Седьмой** раздел посвящен математическому моделированию распространения эпидемии коронавируса COVID-19 в Москве. В работе используется дискретное логистическое уравнение, которое также имеет два параметра: показатель роста численности заболевших, максимальное значение численности жителей, которые потенциально могут быть инфицированы коронавирусом COVID-19.

Логистическое уравнение для описания распространения коронавируса COVID-19 имеет вид:

$$y_{n+1} = \lambda y_n \left(1 - \frac{y_n}{N}\right),$$

где  $y_n = y(t_n)$  – численность популяции заболевших в  $n$ -й момент времени – коэффициент (показатель) роста численности популяции;  $N$  – максимальное значение численности жителей, которые потенциально могут заразиться коронавирусом COVID-19. Этот параметр зависит от целого ряда факторов, таких как численность населения, его скученность или плотность, устойчивость к заболеванию, дисциплинированность населения во время карантинных мероприятий и др.

Эта модель показывает, что численность популяции заболевших быстро растет, пока она мала ( $y_n \ll N$ ) и начинает убывать, когда заболевших становится много.

Далее были рассмотрены результаты расчета по математической модели. Для проверки адекватности использования логистического уравнения распространения коронавируса COVID-19 в Москве использовались данные по распространению вируса в Китае. Начальное значение нормированного множителя  $N$  для Китая подбирали, используя соотношение. Для Китая количество заболевших на конец эпидемии составляло  $\bar{y} = 82500$ . Начальное приближение среднего значения параметра показателя роста популяции брали равным  $\langle \alpha \rangle \sim 1,12$ .

После проведения анализа соотношения нормировочного множителя к числу жителей Москвы рассмотрены 4 сценария развития распространения эпидемии коронавируса COVID-19.

- 1 сценарий:  $N = 1\,000\,000$  чел. ( $N/A \cdot 100\% = 7,1\%$ );
- 2 сценарий:  $N = 760\,000$  чел. (5,4%);
- 3 сценарий:  $N = 500\,000$  чел. (3,6%);
- 4 сценарий:  $N = 300\,000$  чел. (2,1%).

Значения параметров показателей роста численности популяции заболевших представлены ниже:

Сценарий 1:  $N = 1\,000\,000$  чел.

- $\alpha = 1,14$  с 5 по 21 марта;
- $\alpha = 1,111$  с 22 марта по 1 апреля;
- $\alpha = 1,081$  с 1 апреля.

Сценарий 2:  $N = 760\,000$  чел.

- $\alpha = 1,14$  с 5 по 21 марта;
- $\alpha = 1,111$  с 22 марта по 1 апреля;
- $\alpha = 1,083$  с 1 апреля.

Сценарий 3:  $N = 500\,000$  чел.

- $\alpha = 1,14$  с 5 по 21 марта;
- $\alpha = 1,113$  с 22 марта по 1 апреля;
- $\alpha = 1,084$  с 1 апреля.

Сценарий 4:  $N = 300\,000$  чел.

- $\alpha = 1,14$  с 5 по 21 марта;
- $\alpha = 1,108$  с 22 марта по 1 апреля;
- $\alpha = 1,094$  с 1 апреля.

В завершении данного раздела были проведены обсуждение результатов. Первый сценарий – самый тяжелый, назовем его НьюЙоркским, второй – Уханьским (нормировочный множитель совпадает с нормировочным множителем, взятым для Уханя). Последний четвертый сценарий – самый легкий, назовем Израильским (нормировочный множитель как у Израиля).

Так же был сделан вывод, что значения показателя численности заболевших падают с введением мер по изоляции и уменьшению контактов.

В **восьмом** разделе был проведен численный эксперимент с моделью

seir. Произведено моделирование, для которого были взяты следующие параметры, ориентируясь на открытые данные. Предполагая, что болезнь в среднем длится 14 дней (по крайней мере, сколько длится легкая форма, на которую приходится до 80% случаев), найдем значение  $\gamma=1/14=0,0714$ . Примем  $\beta=3/14=0,2143$ . Величину  $\theta=0,6$  заимствуем из работы Pengpeng et al. С учетом средней длительности инкубационного периода в 3 дня, возьмем  $k=1/3=0,33$ . Население России примем равным  $N=144,5 \cdot 10^6$  человек. В качестве начальных условий используем данные по России на 2 апреля, когда введенные в конце марта меры по ограничению распространения инфекции должны были возыметь свое действие, а именно:

$$\begin{aligned} S_0 &= 3548, \\ I_0 &= 3283, \\ E_0 &= 0,5I_0. \end{aligned}$$

Оценку  $E_0$  – произвольная оценка.

Далее была проведена оптимизация, в ходе которой было использовано среднеквадратичное отклонение как мера погрешности модели. Целевая функция примет вид:

$$f(\mathbf{x}) = \frac{1}{M} \sqrt{\sum_{i=1}^M (F_i - \mathcal{F}_i)^2 + \sum_{i=1}^M (G_i - \mathcal{G}_i)^2},$$

где  $M$  — число точек,  $F$  — общее число случаев заражения, которое дает модель,  $\mathcal{F}$  — реальное общее число случаев,  $G$  — число больных в текущий момент, которое дает модель,  $\mathcal{G}$  — реальное общее текущих активных случаев.

На следующем этапе было произведено сравнение трех европейских стран, находящихся в топ-5 по количеству заболевших, и Россию. Видно, что по темпам развития эпидемии Россия отстает примерно на месяц, и что вот уже как месяц рост общего числа случаев во всех трех странах практически линейный (и даже медленнее линейного), в отличие от результатов, полученных в SEIR модели.

Далее была произведена модификация Модели Seir. В качестве примера была построена модель эпидемии в виде клеточного автомата, где каждая



клетка взаимодействует только с 4 соседними. Это эквивалентно тому, что у каждого индивида популяции 4 социальных контакта — это очень маленькое число для человеческой популяции, но тем быстрее проявится эффект ограничения социальных связей. На каждой итерации с вероятностью 0,1 каждый из 4-х соседей зараженной клетки может быть заражен. Болезнь длится в среднем 14 дней. Результаты моделирования для пула из 200x200 клеток представлены на рисунке ниже, где  $k$  — номер итерации.

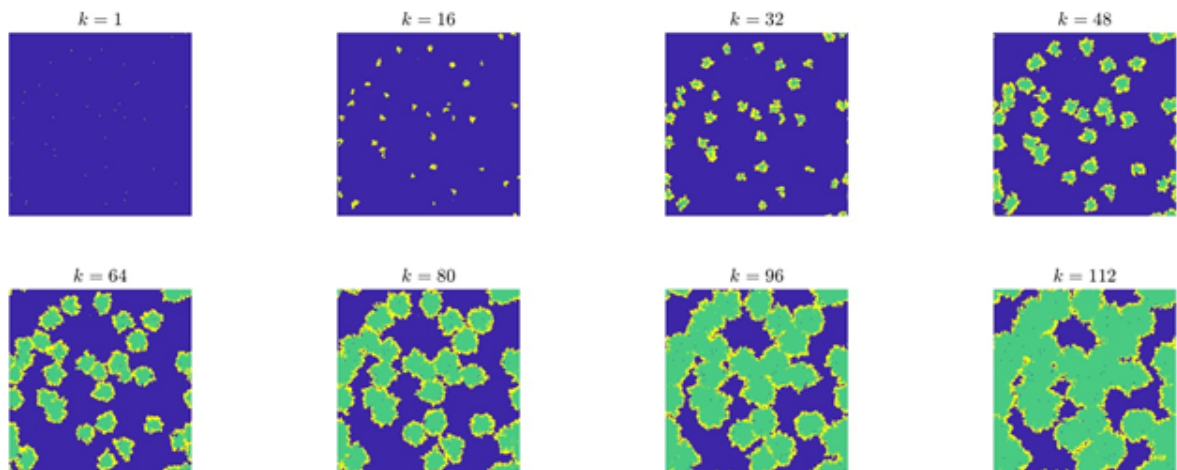


Рисунок 1 – Результаты моделирования

Синий цвет обозначает восприимчивых, желтый — заболевших, зеленый — выздоровевших. Графики численности заболевших так: после недолгой фазы субэкспоненциального роста, как в модели SEIR, наблюдается затяжная фаза линейного роста - как в реальности.

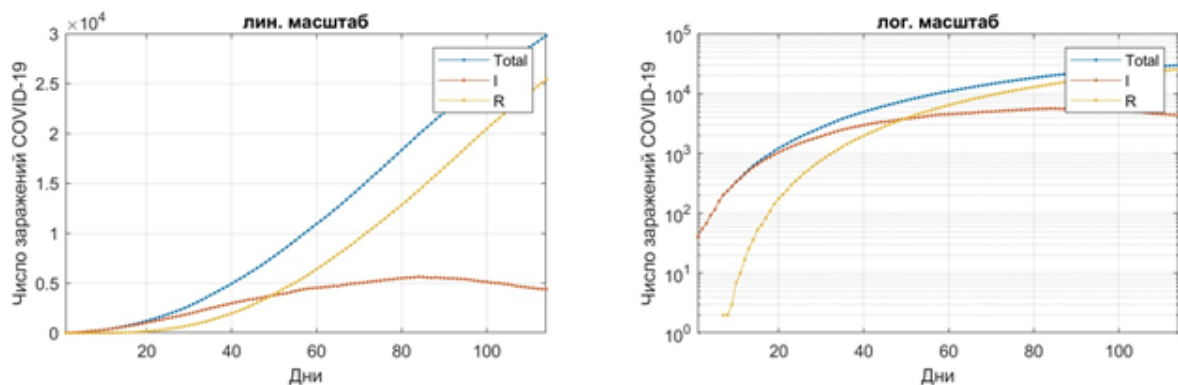


Рисунок 2 – Результаты моделирования

Затем были рассмотрены два сценария, где нельзя точно сказать, по какому из них будут развиваться события дальше.

Первый - сценарий США. Запоздав с принятием ключевых решений, которые позволили бы замедлить рост эпидемии в начале, страна и на момент рассмотрения не в силах справиться с естественным приростом новых случаев.

Второй - сценарий Германии. Благодаря дисциплине в стране, рост происходит медленнее линейного. Более того, чтобы сделать модель релевантной, мне пришлось вручную добавить 6 апреля увеличение коэффициента выздоровления  $\gamma$  в 1,7 раз, иначе такое резкое падение числа случаев в терминах модели SEIR не объяснить.

**В заключении** приведены результаты магистерской работы.

## Основные результаты

1. Рассмотрены некоторые эпидемиологические подходы моделирования. Из них к данному датасету применима только стохастическая SI-модель. SI-модель проигрывает предлагаемому методу в качестве предсказания меток пользователей.;
2. Рассмотрены несколько стратегий прогнозирования временных рядов для модели машинного обучения в рамках задачи предсказания распространения эпидемий, лучшее качество показала рекурсивная одношаговая модель, опирающаяся на историю двух последних наблюдений.
3. Изучена необходимая информация о пандемии коронавируса, а именно распространение вируса, симптомы, продолжительность болезни и инкубационного периода и кто находится в группе риска.;
4. Построена и проанализирована модель распространения пандемии в Москве. В проведенной практической работе показано использование дискретного логистического уравнения для моделирования распространения коронавируса COVID-19 в Москве. Определены показатели роста численности популяции инфицированных коронавирусом COVID-19 для 4-х сценариев развития эпидемии в Москве. Для более простого варианта – рассчитанное время пика приходится на 15.04 с приростом числа заболевших на пике 1210 чел. и общей численностью 13 636 чел. Для более сложного варианта время пика – 24.04 с приростом числа заболевших 3033 чел. и общей численностью на пике 38 130 человек.;
5. Проведен численный эксперимент с моделью SEIR.