

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теории функций и стохастического анализа

**Разработка программного продукта для моделирования
распространения эпидемий на графах**

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студентки 2 курса 248 группы

направления 09.04.03 – Прикладная информатика

механико-математического факультета

Задворной Ольги Александровны

Научный руководитель
доцент, к. э. н.

подпись, дата

А. Р. Файзлиев

Зав. кафедрой
д. ф.-м. н., доцент

подпись, дата

С. П. Сидоров

Саратов 2023

Введение.

Актуальность темы. Выпускная квалификационная работа раскрывает тему моделирования эпидемий. На инфекционные заболевания сейчас приходится 43% от общего числа болезней. Лекарств и вакцин редко бывает достаточно, чтобы остановить инфекционное заболевание — не менее важно понимать, как патоген, ответственный за заболевание, распространяется среди населения. Так как речь зачастую идет о человеческих жизнях, невозможно предпринимать какие-либо меры вслепую или только на основании уже сложившейся ситуации, не располагая любой дополнительной информацией. Поэтому всё более возрастающее значение приобретает развитие способов моделирования эпидемий и их дальнейшего прогнозирования.

Общим методом, используемым для математического моделирования инфекционных заболеваний, являются *компартиментные модели*, в которых популяция разделяется по группам (например, восприимчивые, инфицированные, выздоровевшие). Люди при этом могут перемещаться между группами, а порядок расположения букв в аббревиатуре названий моделей обычно показывает структуру потока между группами.

Происхождение таких моделей относится к началу XX века, а важными работами являются работы Росса в 1916 году, Росса и Хадсона в 1917 году, Кермака и Маккендрика в 1927 году и Кендалла в 1956 году. Модель Рида-Фроста также является значимым предком современных подходов к эпидемиологическому моделированию.

Модели чаще всего выполняются с использованием обыкновенных дифференциальных уравнений (которые являются *детерминированными*), но также могут использоваться со *стохастической* (случайной) структурой, которая более реалистична, но гораздо сложнее в анализе. Здесь распространенным подходом являются *агентные модели*.

Модели пытаются предсказать такие вещи, как распространение болезни, общее число инфицированных, продолжительность эпидемии, а также оценить различные эпидемиологические параметры, такие как базовое репродуктивное число. Такие модели могут показать, как различные мероприятия здравоохранения могут повлиять на исход эпидемии, например, какой метод

введения ограниченного количества вакцин является наиболее эффективным для данной популяции.

Целью магистерской работы является написание программы для моделирования распространения эпидемий на различных видах случайных и реальных графов.

В ходе работы рассмотрены следующие **задачи**, способствующие осуществлению обозначенной цели:

- Исследовать классические модели эпидемий;
- Изучить некоторые из рассмотренных моделей на контактной сети;
- Рассмотреть важные характеристики сетей, влияющие на распространение эпидемий;
- Выявить перспективные стратегии сдерживания эпидемий и их виды;
- Описать агентный подход моделирования эпидемий;
- Определить требования к создаваемой программе;
- Реализовать программу для моделирования распространения эпидемий, включая различные модели и их усложнения;
- Визуализировать результаты моделирования и сравнить их между собой.

Практическая значимость проводимого исследования заключается в том, что на основании построенных компьютерных моделей распространения эпидемий можно сложить представление о подобном распространении в реальных сетях, сравнивая между собой различные модификации и усложнения и определяя наиболее реалистичные из них. По результатам симуляций можно делать выводы о состоятельности и эффективности применяемых схем распространения и сдерживающих мер.

Работа прошла апробацию на различных конференциях. В частности, на XI Международной молодежной научно-практической конференции «Математическое и компьютерное моделирование в экономике, страховании и управлении рисками» в ноябре 2022 года и на ежегодной студенческой конференции «Актуальные проблемы математики и механики», которую проводил механико-математический факультет СГУ в апреле 2023 года, в секции «Анализ данных» и на итоговой студенческой научной конференции СГУ в мае 2023 года.

Структура и содержание магистерской работы. Работа состоит из введения, трёх разделов, заключения, списка использованных источников из 25 наименований и четырёх приложений. В первом разделе «Детерминированное моделирование эпидемий на графах» даются основные сведения о базовых моделях распространения эпидемий и их сравнения, рассматриваются детерминированный подход, описывающий распространение через систему дифференциальных уравнений, и его применимость к сетям. Второй раздел «Модификации моделей распространения эпидемий» содержит возможные модификации разрабатываемой системы, то есть введение дополнительных факторов, прибавляющих системе большую реалистичность. Третий раздел «Реализация стохастического подхода к моделированию эпидемий» посвящен непосредственно написанию программы и анализу результатов. В нем приводится описание используемого подхода, реализованных моделей и графов, на которых они запускаются, а также делаются выводы по полученным данным.

Основное содержание работы.

В первом разделе «Детерминированное моделирование эпидемий на графах» даются основные сведения о базовых моделях распространения эпидемий и их сравнения, рассматриваются детерминированный подход, описывающий распространение через систему дифференциальных уравнений, и его применимость к сетям.

Эпидемиология разработала надежную аналитическую и численную основу для моделирования распространения патогенов. Этот подход основан на *двух фундаментальных гипотезах*:

Гипотеза 1. Компартиментализация. *В моделях эпидемии каждого человека классифицируют в зависимости от стадии заболевания.* Самая простая классификация предполагает, что человек может находиться в одном из трех состояний (компартиментов):

- **Восприимчивые** (Susceptible, S);
- **Инфицированные** (Infectious, I);
- **Выздоровевшие** (Recovered, R).

Моделирование некоторых заболеваний требует дополнительных состояний. Люди при этом могут перемещаться между группами.

Гипотеза 2. Однородное смешивание. *Каждый человек имеет одинаковые шансы вступить в контакт с инфицированным.*

Эта гипотеза устраняет необходимость знать точную сеть контактов, по которой распространяется болезнь, заменяя ее предположением о том, что любой может заразить кого угодно.

В этом разделе была представлена структура моделирования эпидемии, основанная на этих двух гипотезах. Более конкретно, исследована динамика трех часто используемых моделей эпидемий, так называемых моделей **SI**, **SIS** и **SIR**, которые помогают понять основные моменты моделирования эпидемий.

В зависимости от характеристик патогена нужны разные модели, чтобы фиксировать динамику эпидемической вспышки. Прогнозы моделей **SI**, **SIS** и **SIR** согласуются друг с другом на ранних стадиях эпидемии: когда количество инфицированных людей невелико, болезнь распространяется свободно, а количество инфицированных увеличивается экспоненциально. Результаты для большого промежутка времени разные: в модели **SI** заражаются все; модель **SIS** либо достигает эндемического состояния, при котором всегда инфицирована конечная часть людей, либо инфекция исчезает; в модели **SIR** все выздоравливают. Репродуктивное число R_0 предсказывает долгосрочную судьбу эпидемии: при $R_0 > 1$ патоген сохраняется в популяции, а при $R_0 < 1$ он вымирает естественным путем.

Стоит отметить, что рассмотренные детерминированные модели игнорировали тот факт, что человек вступает в контакт только со своими сетевыми соседями в соответствующей контактной сети. Это также означает, что инфицированный обычно заражает только $\langle k \rangle$ других индивидуумов (где $\langle k \rangle$ — средняя степень вершины в графе, среднее число соседей), игнорируя вариации в степени узлов. Для более точного предсказания динамики эпидемий была рассмотрена роль, которую контактная сеть играет в эпидемических явлениях.

Для сетевых эпидемий справедливо: человек может передавать патоген только тем, с кем он вступает в контакт, следовательно, патогены распространяются по сложной сети контактов. Более того, эти контактные сети часто

не масштабируются, поэтому $\langle k \rangle$ недостаточно для характеристики их топологии.

Если патоген распространяется по сети, люди с большим количеством связей с большей вероятностью будут контактировать с инфицированным человеком, следовательно, они с большей вероятностью будут инфицированы. Математический формализм должен рассматривать степень каждого узла как неявную переменную. Это достигается с помощью *блочной-степенной аппроксимации*, которая различает узлы на основе их степени и предполагает, что узлы с одинаковой степенью статистически эквивалентны.

Модели **SI**, **SIS** и **SIR** были дополнительно рассмотрены в случайных и безмасштабных сетях. Дальнейший анализ показал, что учет топологии сети значительно изменяет предсказательную силу моделей эпидемии. Было получено два фундаментальных результата:

- В большой безмасштабной сети $\tau = 0$ (где τ — характерное время распространения патогена), что означает, что вирус может мгновенно достичь большинства узлов;
- В большой безмасштабной сети $\lambda_c = 0$ (где λ_c — эпидемический порог), что означает, что даже вирусы с небольшой скоростью распространения могут сохраняться в популяции.

Оба результата являются следствием способности концентраторов транслировать патоген на большое количество других узлов.

Второй раздел «Модификации моделей распространения эпидемий» содержит возможные модификации разрабатываемой системы, то есть описание вводимых дополнительных факторов, прибавляющих системе большую реалистичность.

Сетевые эпидемии предсказывают, что скорость, с которой патоген распространяется, зависит от распределения степеней соответствующей контактной сети. Действительно, было обнаружено, что $\langle k^2 \rangle$ влияет как на характерное время τ , так и на эпидемический порог λ_c . Однако ни один из этих выводов не имеет значения, если сеть, по которой распространяется патоген, является случайной — в этом случае прогнозы сетевых эпидемий неотличимы от прогнозов традиционных моделей эпидемий. Поэтому была исследована структура нескольких контактных сетей, встречающихся при эпидемических

явлениях, и предложены прямые эмпирические доказательства значимости лежащих в основе различий в степени.

Были рассмотрены такие важные особенности реальных сетей, как географическая мобильность, паттерны локальных контактов, учет времени взаимодействия, прерывистый характер контактов, корреляция степеней, образование сообществ, возможность как простого, так и сложного заражения.

Некоторые сетевые характеристики замедляют патоген, другие способствуют его распространению. Поэтому эти эффекты должны быть учтены, если стоит цель предсказать распространение реального патогена. Хотя эти закономерности имеют очевидное отношение к инфекционным заболеваниям, они также влияют на распространение идей, продуктов, компьютерных вирусов и даже некоторых неинфекционных заболеваний.

Во второй половине раздела подробно рассматриваются *стратегии сдерживания эпидемий*, а также описываются принципы случайной и выборочной видов иммунизации.

Некоторые из наиболее распространенных стратегий сдерживания эпидемий включают следующие:

- *Мероприятия по снижению передачи патогена.*

Например, для снижения скорости передачи воздушно-капельных или контактных патогенов можно использовать маски для лица, перчатки и обработку рук и рабочих поверхностей дезинфицирующими средствами.

- *Мероприятия по сокращению числа контактов.*

При заболеваниях с серьезными последствиями для здоровья вводятся такие меры, как изоляция пациентов, закрытие учреждений и ограничение доступа к часто посещаемым общественным местам, таким как кинотеатры и торговые центры. Это делает сеть более разреженной за счет уменьшения количества контактов между отдельными лицами, что соответственно снижает скорость передачи.

- *Вакцинация.*

Вакцинация навсегда удаляет привитые узлы из сети, поскольку они не могут быть инфицированы и не могут распространять болезнь. Вакцинация также снижает скорость распространения, повышая вероятность вымирания патогена.

Стратегии иммунизации определяют, как вакцины, методы лечения или лекарства распределяются среди населения. В идеале, если существует лечение или вакцина, ее следует вводить каждому инфицированному человеку или тем, кто подвержен риску заражения патогеном. Тем не менее, часто соображения стоимости, трудности охвата всех лиц, подверженных риску, и реальные или предполагаемые побочные эффекты лечения препятствуют полному охвату. Учитывая эти ограничения, стратегии иммунизации направлены на минимизацию угрозы пандемии путем наиболее эффективного распределения имеющихся вакцин или методов лечения.

Известные способы сдерживания эпидемий важно учитывать при их моделировании и дальнейшем прогнозе их развития при применении различных мер.

Третий раздел «Реализация стохастического подхода к моделированию эпидемий» посвящен непосредственному написанию программы и анализу результатов. В нем приводится описание используемого подхода, реализованных моделей и графов, на которых они запускаются, а также делаются выводы по полученным данным.

Избранным для реализации программы стал один из методов стохастического моделирования эпидемии — *агентный подход*. Преимущество данного подхода заключается в том, что реально включить все виды деталей в модель на индивидуальном уровне. Также можно реализовать произвольно сложные распределения вероятностей для времени, проведенного в инфекционном периоде.

Разрабатываемая программа предназначена для построения различных моделей распространения эпидемий на графах. Среди **моделей** рассматриваются следующие:

- Классическая SIR модель на графах;
- Модификация SEIR для эпидемий с инкубационным периодом (E — Exposed);
- Модификация SEIQR, учитывающая изоляцию вершин (Q — Quarantined).
- Модификация SEIQRS, реализующая понятие временного иммунитета.

Модели рассматриваются как для равномерных, так и для нелинейных вероятностей перехода из одной группы в другую. В качестве эксперимента SEIQR расширена до учета вакцинации (SEIQVR, где V — Vaccinated) со сравнением эффективности случайной и выборочной стратегий вакцинации.

Предполагается возможность запуска моделей на следующих видах **графов**:

- Полный граф (однородно смешанный);
- Случайный граф в соответствии с моделью Эрдёша-Реньи;
- Случайный граф в соответствии с моделью Уоттса-Строгаца («Мир тесен»);
- Случайный граф в соответствии с моделью Барабаши-Альберт;
- Выбранные графы из Стэнфордской коллекции больших сетевых наборов данных.

Программа формально описана. Приведены цель разработки, основные требования к системе, порядок разработки, ограничения системы, а также используемые методы и инструменты.

Каждый раздел полученной программы подобно описан с приведением соответствующих прокомментированных кусков кода.

Отдельное внимание уделено генерации случайных графов и приведению их характеристик друг к другу для справедливого сравнения различных запусков, а также масштабированию графов при сохранении их линейной плотности d_l . При этом доказано, что топологические свойства реальной сети наиболее полно описывает именно модель Барабаши-Альберт. Таким образом, этот вид графов из всех представленных наиболее близко подходит к описанию реальных графов.

Все пять видов реализованных моделей и их характерные отличия друг от друга подробно описаны. Для каждой из них приведены упрощающие понимание схемы и представлены используемые параметры. Дополнительно проясняется разница между «simple» и «sojourn» моделями. Вторые отражают более реалистичные модели, так как вводят нелинейную вероятность перехода из одного состояния в другое.

В разделе с результатами приводятся запуски основных рассматриваемых моделей на четырёх сгенерированных графах и проанализированы по-

лученные результаты. Первые симуляции были произведены для графов с 1 000 вершинами и 0,5% первоначально зараженных (I_0), чтобы оценить работоспособность моделей и графов друг против друга. Отдельно сравнивается эффективность различных стратегий вакцинирования. Для моделей проверяется их устойчивость и масштабируемость. Затем полученные результаты сравнивались с запусками на SNAP-графах и историческими данными по COVID-19.

Было доказано, что представляется возможным производить симуляции на графах различных размеров, но аналогичных свойств и масштабировать полученный результат на сети реальных размеров. Не было получено однозначное доказательство преимущества выборочной вакцинации над случайным, несмотря на присутствующий в графах парадокс дружбы. Было получено, что распространение по графам, построенным в соответствии с моделью Барабаши-Альберт, более близко отражает распространение по реальным сетям, однако всё еще не описывает его в полной мере. В качестве дальнейших шагов исследования представляется актуальным усложнение модели распространения эпидемий на графах и рассмотрение случайных графов, учитывающих большее число характеристик реальных сетей.

В приложениях представлены свойства случайных графов, на которых проводились финальные запуски, листинги некоторых разделов программы, визуализация динамики изменения состояний на графе по модели Барабаши-Альберт для всех реализованных схем распространения, а также визуализация распространения эпидемии на SNAP-графе по части сети Facebook.

Заключение. По итогу выполнения выпускной квалификационной работы была написана программа для моделирования распространения эпидемий на графах.

В ходе написания магистерской работы было совершено следующее:

- Был проведен анализ существующих эпидемиологических подходов как в классическом варианте, так и в сетях;
- Были изучены характеристики сетей, значительно влияющих на скорость распространения эпидемий и характер этого распространения;
- Были рассмотрены методы контролирования и сдерживания эпидемий;

- Были определены основные теоретические аспекты выбранного подхода к моделированию;
- Было написано 12 функций, реализующих различные модели распространения эпидемий;
- Был проведен ряд симуляций, а их результаты сравнены против данных распространения по реальным сетям.

В конце каждого раздела были сделаны важные выводы. Все поставленные цели были в полной мере выполнены.

Таким образом, была реализована программа на платформе Kaggle на языке R, дающая возможность смоделировать эпидемию и её распространение. Рассмотренный подход предоставляет широкие возможности по дальнейшему усложнению моделей и оптимизации, однако уже сейчас показывает хорошие результаты, имеющие смысл в реальной жизни.