

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
**«Крымский федеральный университет имени В. И.
Вернадского»**
(ФГАОУ ВО «КФУ им. В. И. Вернадского»)
Таврическая академия (структурное подразделение)
Факультет математики и информатики
Кафедра прикладной математики

Консманов Алексей Витальевич

**Моделирование распространения
инфекционных заболеваний**

Выпускная квалификационная работа

Обучающегося	<u>4</u> курса
Направления подготовки	<u>01.03.04. Прикладная математика</u>
Форма обучения	<u>очная</u>

Научный руководитель
доцент кафедры прикладной математики,
кандидат физико-математических наук Ю.Ю. Дюличева

Симферополь 2019

Оглавление

Введение	3
1 Подходы к моделированию процесса распространения инфекционных заболеваний	5
1.1 Агентный подход к имитационному моделированию процесса распространения заболевания	7
1.2 Применение случайных графов для распознавания и анализа инфекционных заболеваний	9
1.3 Модели инфекционной динамики на основе предфрактальных графов	13

Введение

Несмотря на значительные достижения в области борьбы с инфекционными заболеваниями и порождаемыми ими эпидемиями, данная проблема и связанные с ней задачи все ещё актуальны. Одной из задач, возникающих в связи с борьбой с инфекционными заболеваниями, является исследование путей и способов передачи инфекций. В данной области возникает задача прогнозирования распространения конкретного инфекционного заболевания в некоторой заданной среде, то есть задача построения математической модели, описывающей скорость и масштабы распространения заболевания и его последствия: количество инфицированных и умерших, пространственные масштабы заражения, затраты средств и ресурсов, необходимые для изоляции больных и последующего излечения.

Математическое моделирование является мощным и гибким инструментом для исследования реальных процессов и объектов, а также связей между ними. Математическое моделирование удобно применять в тех ситуациях, когда проведение эксперимента с реальными субъектами затруднено по любой причине, что делает его подходящим способом исследования распространения инфекционных заболеваний. Заметим, что учитывая специфическую природу заболеваний и путей их передачи, которые будут изучены и исследованы далее, важным фактором для проверки адекватности построенной модели является наличие статистической информации, позволяющей тестировать соответствие модели реальным прецедентам.

Важно заметить, что модели, разработанные в данной области, не являются идеальными и универсальными. Основные допущения, принимаемые при построении модели, и недостатки этих моделей будут подробно рассмотрены далее.

Целью данной работы является анализ существующих моделей распространения инфекционных заболеваний и создание нескольких различных собственных моделей, позволяющих прогнозировать распространение конкретных заболеваний на основе существующих статистических данных, реализовать полноценное приложение для использования специалистами-эпидемиологами. Для этого необходимо решить нижеизложенный ком-

плекс задач:

- Проанализировать существующие подходы к моделированию, изучить сильные и слабые стороны этих подходов, принимаемые допущения. На основании этого анализа выбрать подход, который будет использоваться при построении собственной модели.
- выбрать заболевание для тестирования, подробно изучить его и собрать статистические данные о нем, найти статистику, позволяющую сравнить результаты моделирования с реальными;
- выбрать среду моделирования;
- разработать и реализовать базовую модель;
- на основе базовой модели создать модель повышенной точности и детализации;
- итеративно тестировать и улучшать результаты модели повышенной точности;
- разработать отдельное приложение для ввода данных в модель и визуализации полученных моделированием результатов для конечного пользователя – специалиста-медика или эпидемиолога.

Для решения поставленного комплекса задач использовались методы математической статистики и теории вероятности, дискретной математики, математического анализа. Разработанные модели основываются на методах агентного / мультиагентного моделирования, их реализации опираются на объектно-ориентированное программирование и современные средства программного анализа и визуализации данных.

Практическая ценность результатов работы: проведено исследование существующих подходов к моделированию, разработана модель и программный продукт, реализующий эту модель, также разработан программный продукт для ввода данных в модель специалистом и динамической визуализации результатов моделирования. Разработанные приложения позволяют дать количественный и качественный прогноз распространения инфекционного заболевания для любых достаточно точно и полно описанных популяции и заболевания.

1 Подходы к моделированию процесса распространения инфекционных заболеваний

Эпидемии всегда являлись одной из наиболее острых проблем медицины. Внезапно возникающие эпидемии формируют сложную ситуацию, динамически изменяющуюся во времени. В таких условиях медики и связанные чрезвычайные службы зачастую не могут принять адекватного решения и только порождают дополнительный хаос, неэффективно расходуют ресурсы, не способны в целом положительно повлиять на ход эпидемии или реализовать меры по борьбе с заболеванием.

Логично предположить, что одним из способов снижения угрозы для населения и уменьшения хаоса среди медиков и прочих ответственных служб являются меры предупредительного характера. Предупредительные меры могут иметь двойственный характер:

- 1 Во-первых, на основании регулярно обновляемых статистических данных, описывающих эпидемиологическую обстановку, возможно создавать кратковременную прогнозирующую модель, что позволит обнаружить эпидемию ещё до её начала;
- 2 Во-вторых, учет данных о прошлых подобных вспышках заболеваемости позволит составить прогноз для ответственных служб, описывающий характер необходимых ресурсов, их количество и область применения, что в свою очередь позволит уменьшить порождаемый вспышкой хаос.

Подобная идея не является революционно новой и к моделированию распространения заболеваний существует достаточно много подходов, основанных на разных принципах: полигамные модели, цепи Маркова, агентный подход, стохастические эксперименты, дифференциальные уравнения на графах, различные алгоритмы на графах. Далее будут подробно рассмотрены некоторые из этих подходов.

Некоторые общие допущения для всех моделей:

- Равномерность и стационарность в смысле неизменяемости во времени распределения возрастов, то есть все в популяции живут до

некоторого возраста L и затем умирают и для каждого возраста (включительно до L) количество людей в этой возрастной группе примерно или строго равно. Этот подход хорошо подходит для экономически развитых стран и стран, находящихся на постиндустриальном этапе экономического развития, где детская смертность мала и большинство людей доживают до ожидаемого возраста. Данное предположение может быть отвергнуто для стран, не подходящих под условия выше, например, при моделировании распространения лихорадки Эбола в странах Центральной и Южной Африки;

- Гомогенность или однородность перемещений внутри популяции, то есть индивидуумы в популяции *under scrutiny assort* и контактируют случайно, не замыкаясь в более мелких подгруппах. Данное допущение является спорным, т.к. социальная структура широко-масштабная и сложная, то есть индивидуумы внутри одной большой группы могут находиться в таких социальных отношениях, что большинство контактов будет приходиться на их собственную подгруппу, в то время как количество контактов вне группы будет предельно малым. Однако, такое допущение имеет место быть для упрощения построения модели и понимания результатов моделирования.

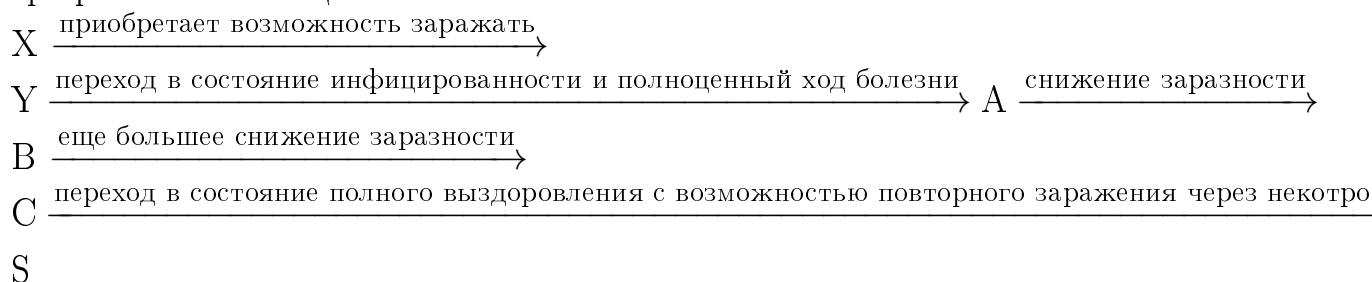
Также опишем один общий принцип разделения моделей по признаку случайности. По такому признаку модели можно разделить на стохастические и детерминистические. Стохастические модели предполагают наличие случайных величин и являются инструментом для оценки распределения вероятности потенциальных исходов посредством допущения случайных изменений в одной или нескольких переменных с ходом времени. Такие модели зависят от случайных изменений рисков, связанных с продолжительностью экспозиции, вероятности заболевания и прочих динамических изменяющихся параметров заболевания. Для описания различных этапов заболевания часто используют буквы M(maternally-derived immunity), S(suspicious), E(exposed), I(infected), R(recovered).

1.1 Агентный подход к имитационному моделированию процесса распространения заболевания

Существует несколько понятий определения "агент" и "агентный подход" в целом. Обобщая их, можно выделить, что *агент* – это некоторая сущность, имеющая активность, автономное поведение, способная самостоятельно принимать решения в соответствии с некоторой совокупностью правил, взаимодействовать с окружающей средой, если такая предусмотрена, и другими агентами, если такие существуют. Основное поле применения таких моделей – децентрализованные системы, динамика функционирования которых не является следствием некоторых, зачастую внешних, правил и законов, но наоборот, такие правила и законы являются внутренним неотъемлемым результатом работы множества агентов. Такие модели обычно являются дискретно-событийными или дискретными с непрерывными элементами, то есть гибридными.

Данный подход удобно иллюстрировать на примере моделирования распространения гриппа (обычно, гриппа А), т.к. грипп вносит весомый вклад в смертность при инфекционных заболеваниях и при этом хорошо известны его способы передачи и формы течения болезни. Также по данному заболеванию собрана достаточно обширная статистика, которую можно использовать для проверки точности модели и последующей калибровки. Такая модель достаточно подробно рассмотрена в работе М.А. Кондратьева, Р.И. Ивановского, Л.М. Цымбалова «Применение агентного подхода к имитационному моделированию процесса распространения заболевания». В данной работе ставится цель разработки модели, способной определять число больных гриппом в каждый день определенного краткосрочного периода (2-3 недели) во время сезонных вспышек, то есть построение краткосрочного прогноза. Для этого популяция разбивается на возрастные группы, позволяющие характеризовать контактов с другими людьми в сутки и, что важно, потенциально посещаемые места. В работе вводятся два основных типа объектов: локации и агенты и связанные с ними события перемещения агентов между локациями, протекания заболевания у агентов и событие «контакта». Также в данной работе вводятся правила, описывающие поведение агентов с помощью UML Statecharts (далее «стейтчарты»). С помощью таких стейтчартов

описывается переход между различными состояниями индивидуума как вне течения болезни, так и после, и может быть обозначен с помощью графической нотации как:



где X – этап инкубационного периода, когда агент еще не заразен, Y – также этап инкубационного периода, где агент становится заразным, A – начало этапа болезни с явным выражением симптомов, этапы B и C представляют собой постепенный спад заразности и после C наступает выздоровление, когда индивидуум более не заразен, однако снова может оказаться больным с течением времени, если не выработался иммунитет.

Данная модель обладает рядом преимуществ: высокая скорость разработки в силу выбора готового имитационного ПО (специфично для данной рассматриваемой работы, хотя в целом верно для разработки любой агентной модели в силу простоты инкапсуляции агентов и построения связей между объектами); высокая скорость работы, так как выбранная модель допускает высокий параллелизм в рамках одной итерации, когда вычисления для каждого объекта-места допускают распараллеливание; свобода от дифференциальных уравнений значительно упрощает расчеты, в том числе исчезает необходимость использовать численные методы и связанные с ними понятия «сходимости» и «устойчивости», при этом последнее особенно важно, так как входные статистические данные зачастую имеют шумы; сам агентный подход допускает быструю модификацию модели и возможность учета применяемых административных мер и их влияния на ход распространения инфекции.

Недостатки: в силу собственной стохастичности, модель требует многократного запуска и оценкой экспертом полученных результатов для получения некоторого «среднего результата»; достаточно сложная модель может потребовать много времени на разработку даже при использовании готового имитационного ПО, при этом такая модель может оказаться ресурсозатратной в терминах машинных мощностей (время, память);

невозможно точно определить, насколько точна данная модель по отношению к реальным, кроме как посредством многократного сравнения результатов работы модели и собранных существующих данных, при этом невозможно убедиться, что данная модель и текущие параметры могут быть применимы для прогнозирования, а не только откалиброваны для соответствия предыдущим результатам; стохастическая природа модели при компьютерной реализации опирается на генерацию случайных чисел, что требует использования мощных и проверенных генераторов.

1.2 Применение случайных графов для распознавания и анализа инфекционных заболеваний

В качестве другого подхода к моделированию распространения инфекционных заболеваний можно рассмотреть переносчиков заболеваний как саморазмножающиеся сущности, т.е. такие сущности, которые способны самореплицироваться, при этом свойство саморепликации передается не только между «родителями» и «потомками» (вертикальный перенос), но возможен и горизонтальный перенос, т.е. передача свойства саморепликации и, возможно, но не обязательно, других свойств, объектам, которые так или иначе соседние или контактируют с данным. Примерами являются размножение компьютерных вирусов и сетевых червей или инфицирование организмов вирусами и инфекциями. В данных моделях обычно применяются графы, используемые для моделирования эпидемий «мобильных червей» (т.е. вредоносных программ, распространяющихся между устройствами при помощи беспроводных протоколов типа Bluetooth или Wi-Fi, широко применяемых в мобильных устройствах, отсюда и происходит название) или воздушно-капельных инфекций среди высших животных. Согласно К.Е. Климентьеву, для подобных эпидемий характерны следующие черты:

- Постоянные изменения топологии среды моделирования в связи с высокой мобильностью объектов;
- ограниченный радиус «заражения», как для биологических агентов, так и для компьютерных вирусов, обусловленный физическими свойствами оных.

Для моделирования среды существования таких объектов используется класс сетей, называемый «специальным» («Ad hoc»), представляющий собой множество случайных графов с разнообразной топологией. При этом нетрудно выделить общие для таких графов характеристики:

- Вероятностное распределение степеней вершин, обозначаемое k_i ;
- также вероятностное распределение для локальных коэффициентов кластеризации вершин, обозначаемое c_i .

При этом под «степенью вершины» x подразумевается количество вершин графа G , инцидентных вершине x , т.е. степень вершины указывает количество вершин, непосредственно соединенных только один ребром с данным. «Локальный коэффициент кластеризации вершины» понимается в том же смысле, что и у К.Е. Климентьева и Б. Хогана, т.е. «коэффициент кластеризации» рассматривается как вероятность того, что два ближайших соседа этого узла сами являются ближайшими соседями, а ЛККВ является мерой того, насколько хорошо связаны между собой соседи данного узла и рассчитывается как

$\frac{\text{число связей между соседями данного узла}}{\text{возможное число связей между соседями}}$ (по Хогану) или, менее формально, по К.Е. Климентьеву: «доля «треугольников» данной вершины, образованных из «соседей» данной вершины и являющихся «соседями» друг для друга в общем количестве потенциально возможных треугольников, где $c_{max} = k_i \frac{k_i - 1}{2}$ ».

В качестве модели пространства, используемых для имитации развития и распространения эпидемии, применяются неориентированные, маркированные графы (в смысле частного случая сети Петри, в которой каждая позиция входом и выходом точно для одного перехода), являющиеся частным случаем дистанционных графов, при этом дистанционный граф понимается как $G = (V, E)$ – n -мерный дистанционный граф (граф расстояний), если $V \subseteq \mathbb{R}^n, E \subseteq \{\{\bar{x}, \bar{y}\} : \bar{x}, \bar{y} \in V, |\bar{x} - \bar{y}| = a, a \in \mathbb{R} > 0\}$, то есть множество вершин M является подмножеством или совпадает с n -мерным пространством, а множество ребер E является подмножеством или совпадает с множеством всевозможных пар вершин \bar{x}, \bar{y} , таких, что евклидово расстояние между \bar{x}, \bar{y} равно некоторому фиксированному вещественному положительному заранее заданному a .

Топология таких графов связана с принципом их построения, т.е. каждая вершина графа имеет некоторые пространственные координаты (обычно в \mathbb{R}^2 или \mathbb{R}^3) и «соседними вершинами» считаются только те, расстояние до которых меньше заданного или равно заданного r_0 (расстояние передачи вирусного воздействия).

Допущения:

- Инфекция передается только между инцидентными вершинами за некоторое время;
- вершины графа активно перемещаются, постоянно изменяя его конфигурацию.

Входные параметры модели:

- Начальная конфигурация графа, т.е. количество и координаты вершин вместе с их начальным состоянием, обычно в виде перечисления «здоров», «вакцинирован», «болен», однако допускается и большее число состояний, подобное применяемому в агентных подходах;
- максимальное расстояние инфицирующего воздействия r_0 ;
- правила или порядок перемещения вершин в процессе моделирования.

Способы построения исходного графа, исходя из «геометрических» и «географических» по К.Е Климентьеву соображений подробно рассмотрено в [1]. При этом для построения геометрического графа по заранее заданным параметрам типа k_i и c_i используется случайный граф Радона (RRG) или случайный геометрический граф (RGG) и в этой же работе выводится функция распределения между двумя случайными точками в единичном квадрате и математическое ожидание расстояния.

Намного более интересным свойством данных моделей являются правила перемещения вершин графа: если в описанных ранее подходах перемещения вершин присутствуют скорее в силу необходимости для динамичности во времени и модель учитывает только их положение по отношению к другим вершинам и, иногда, скорость движения этих, а топология графа не является предметом исследования, то в данном подходе

топологию графа и правила перемещения являются ключевым элементом при построении модели. В результате построения множества моделей, использующих данный подход, можно сделать следующие выводы, описывающие правила перемещения вершин в графе:

- Классические модели блуждания, описывающие, например, броуновское движение, мало пригодны для описания передвижения людей, особенно движущихся вне рутинного маршрута, а высокоподвижных агентов, в силу своей слабой стационарности. Следовательно, при анализе движений по плоскости, направление движения распределено равномерно на интервале $[0; 2\pi]$, что показано в ;
- скорость и продолжительность движения распределены по Леви. При этом «перемещение» или «движение» понимается как самый длинный прямолинейный переход объекта с одного места в другое без паузы или изменения направления ;
- тем не менее, перечисленные правила перемещения актуальны только для перемещений на ограниченном участке плоской (или приближенной к плоской) поверхности без учета препятствий, конфигураций помещения, уставившихся маршрутов, столкновений с другими телами и необходимости их избегать. Для внесения корректировок, учитывающих такие особенности, необходим сбор статистики на множестве реальных наблюдений с последующим внесением в некоторую ГИС.

Базируясь на работах , , , можно сделать следующие выводы, касающиеся основных свойств данного подхода к моделированию:

- 1 Данный подход хорошо подходит для моделирования распространения инфекционных заболеваний среди групп людей и животных, особенно блуждающих, в т.ч. по Леви, в некотором постоянном данном ареале;
- 2 использование данных моделей улучшает понимание того, какие графы лучше описывают топологию быстро движущихся источников инфекции. Так, в работах и четко прослеживается зависимость

между «Ad hoc» графами и графами Радо и той картиной, которую дают реальные перемещения мобильных устройств при вспышке компьютерных инфекций;

3 ДОПИСАТЬ

Достоинства данного подхода: высокая скорость работы моделей при правильном выборе инструментария (готовые фреймворки для работы с графами); высокая точность предсказания протекания вспышки инфекции на начальном этапе этой вспышки; наличие уже разработанных алгоритмов, приборов и готовых собранных данных для занесения в многослойные ГИС; возможность модификации алгоритмов для учета рельефа местности и возможных столкновений инфицируемых сущностей; математический аппарат блужданий Леви допускает модификации, позволяющие легко настраивать и изменять принципы перемещения сущностей; сопоставление результатов моделирования с реальными процессами позволяет изучать зависимости между некоторыми специальными графами, описанными выше, и путями распространения инфекций.

Недостатки: в силу специфики требований, предъявляемых к моделям данного типа, точное моделирование возможно только в течении небольшого количества итераций модели, т.к. предполагается характер «вспышки», то есть быстрого бурного увеличения числа зараженных индивидуумов с достаточно быстрым падением числа больных при изобретении лекарства или разработке антивируса; для построения точных моделей требуется не только точно знать координаты физического пространства, на котором будет происходить распространения заболевания, но и иметь детальные модели самих поверхностей или рельефов, что необходимо для точного построения блужданий Леви; используемые графы высокодинамичные и могут требовать значительных затрат как по памяти, так и по процессорному времени.

1.3 Модели инфекционной динамики на основе предфрактальных графов

Для моделирования распространения инфекционных заболеваний можно также использовать предфрактальные графы. В основу данного подхода

к моделированию положена идея о том, что графы, использующие операцию "замены вершины затравкой" (ЗВЗ), которая будет рассмотрена далее, «прирастают» не отдельными вершинами, а самоподобным или приближенно самоподобным графом, т.е. фрактальным, хорошо описывают системно «прирастающие» структуры: телекоммуникационные, инфраструктурные, социальные и технические сети, некоторые популяции, в которых существует разбиение на группы, в которых связей внутри группы больше, чем между отдельными группами. При этом показано, что такие модели могут быть использованы для моделирования распространения инфекционных заболеваний, а именно, что процессы разрастания графов релевантны динамике распространения инфекционных заболеваний и распад таких графов соответственно релевантен спаду эпидемий. Математико-алгоритмическая идея построения фрактальных графов – операция замены вершины затравкой (ЗВЗ). Для этого вводится несколько понятий, рассмотренных ниже. По определению, динамический граф G_D – последовательность обычных графов G_l , не имеющих параллельных ребер и петель. Затравка – какой-либо произвольный связный граф, т.е. граф, в котором между любой парой вершин существует хотя бы один путь, сама затравка обозначается как $H = (W, Q)$. Предфрактальный граф обозначается как $G_L = (V_L, E_L)$, определяется рекуррентно, заменяя каждый раз в построенном на предыдущем этапе $l = \overline{1, L-1}$ графе $G_l = (V_l, E_l)$ каждую вершину затравкой H . Вершины могут соединяться случайно или по заданному правилу или в некоторой заранее заданной последовательности. На этапе $l = 1$ предфрактальному графу соответствует затравка $G_1 = H$. Говорят, что предфрактальный граф G_L *порожден* затравкой H , при этом процесс построения предфрактального графа G_L по сути является процессом построения последовательности других предфрактальных графов $\overline{G_1, G_L}$, называемый «траекторией».

Суть операции замены вершины затравкой (ЗВЗ): в графе $G = (V, E)$ из выбранной для замещения затравкой вершины $\tilde{v} \in V$ выделяется множество $\tilde{V} = \{\tilde{v}_j\} \subseteq V$, где $j = \overline{1, |\tilde{V}|}$, смежных ей, вершине $\tilde{v} \in V$, вершин; затем из графа G удаляется вершина \tilde{v} и все инцидентные ребра, т.е. ребра, для которых эта вершина общая; после $\forall \tilde{v}_j \in \tilde{V}, j = \overline{1, |\tilde{V}|}$ соединяется ребром с одной из вершин затравки H , при этом соединение может про-

исходить случайно или по заранее заданному правилу или в заданной последовательности.

Под «распознаванием предфрактального графа» в таких подходах к моделированию обычно понимается определение траектории построения предфрактального графа при условии, что заданы виды и типы затраток. При этом задача распознавания различных типов предфрактальных графов хорошо изучена и отработана в практических алгоритмах, что дает широкую теоретико-практическую базу для решения данной задачи.

В предфрактальном графе G рёбра, порожденные на этапе l , где $l \in 1, 2, 3, \dots, L$ называются ребрами ранга l , при этом «новыми» ребрами называют ребра ранга L в предфрактальном графе G_L , все остальные называются «старыми». При этом, под рангом вершины графа понимается наименьший ранг l , выбираемый среди всех рёбер, инцидентных для данной вершины. Вершина ранга l обозначается как $v^{(l)}$, $l \in \{1, 2, \dots, L\}$. Также вводится понятие очага заражения ранга l , который обозначается как $\tilde{v}^{(l)}$ и является, по сути, вершиной ранга l .

Одной из самых важных задач данного подхода к моделированию является задача распознавания фрактального графа, при этом различают два вида распознавания: неявное – определение фрактальности графа и установление некой n -вершинной затратки, на которой он базируется; явное – представление множества рёбер для всякого ранга в явном виде или же представление порождающей траектории в явном виде для некоторого заданного графа G . Для этого используется несколько алгоритмов, показанных в и обозначаемых $\alpha_1 \dots \alpha_4$.