

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
КАФЕДРА МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ
ЭНЕРГЕТИЧЕСКИХ СИСТЕМ

Сковородина Татьяна Валерьевна

Выпускная квалификационная работа магистра

Моделирование информационных процессов в
социальных сетях

Направление 01.04.02

Прикладная математика и информатика

Математическое и информационное обеспечение экономической
деятельности

Научный руководитель,
кандидат физ.-мат. наук,
доцент
Губар Е. А.

Санкт-Петербург
2019

Содержание

Введение	3
Постановка задачи и целей	8
Глава 1. Эпидемические модели	10
1.1 Модель SIS	10
1.2 Модель SAIS	13
1.3 Модель GEMF	16
Глава 2. Модель распространения информации о налоговых проверках	21
2.1 Описание модели SIAP	21
2.2 Процедура моделирования процесса распространения инфор- мации модели SIAP	27
Глава 3. Численное моделирование	31
3.1 Первая серия экспериментов	31
3.2 Вторая серия экспериментов	35
3.3 Третья серия экспериментов	38
3.4 Четвертая серия экспериментов	41
3.5 Выводы	45
Заключение	47
Список литературы	48
Приложение	51

Введение

Математическое моделирование уже давно применяется для изучения распространения инфекционных заболеваний. С его помощью были созданы эпидемические модели — упрощенный способ описания передачи инфекционных заболеваний через отдельных лиц. Эпидемическое моделирование помогает исследовать процессы распространения вирусных заболеваний, выявлять их тенденции к дальнейшему развитию, искать решения предотвращения эпидемий в будущем.

Самыми первыми эпидемическими моделями являются детерминированные. При их использовании предполагается, что популяция разделена на группы, которые соответствуют стадиям развития болезни и размер популяции фиксирован.

Первая детерминированная модель была представлена в 1927 году в статье «Contribution to the mathematical theory of epidemics» [1]. В ней W. O. Kermack и A. G. McKendrick описали модель SIR (susceptible — infected — recovered). Она легла в основу математического моделирования эпидемических процессов, так как позволяет описывать распространение простых инфекционных заболеваний, которые включают в себя три основные группы, соответствующие стадиям болезни: S — восприимчивые, I — инфицированные, R — иммунные. Но также SIR легко адаптируется под более сложные виды эпидемий.

В современных исследованиях большое внимание уделяется модели SIS (susceptible — infected — susceptible) [2] — модификации SIR модели — так как она может рассматриваться как простейшая модель эпидемии, в которой популяция разделена на две группы: инфицированные и восприимчивые.

В последнее время детерминированные модели применяются все реже, так как имеют некоторые недостатки. Один из самых главных заключается в том, что детерминированные модели не учитывают влияние отдельных индивидов на эпидемический процесс. Данный факт не позволяет получить полноценное представление об эпидемическом процессе. Это приводит к тому, что используемая модель далека от реальности. Поэтому сейчас изучаются новые методы и модели, учитывающие взаимосвязи индивидов в популяции. Это необходимо для того, чтобы оценивать и,

возможно, контролировать влияние каждого индивида в распространении инфекционного заболевания.

Модификация детерминированных моделей заключается в том, что распространение инфекционного заболевания рассматривается на графе с применением теории марковских процессов. Узлы графа обозначают агентов, а ребра графа — контакты. В этом случае говорят о точных или стохастических моделях. Такой метод исследования эпидемического процесса применяет Piet Van Mieghem в своей статье [3]. В ней автор получает дифференциальные уравнения для точных моделей SIR и SIS.

Модель SIR была тщательно изучена в работе «An individual-based approach to SIR epidemics in contact networks» [4]. В этой статье, на основе детерминированной модели SIR, авторы проводят сравнительный анализ нескольких видов моделей распространения инфекции на случайной сети. Также исследуют взаимосвязь топологии сети и спектра матрицы смежности с параметрами эпидемии. Еще одним примером исследования влияния топологии является статья «Networks and epidemic models» [5]. Однако, в ней авторы делают упор на определение взаимосвязи между динамикой эпидемического процесса и различными типами сетей: случайной сетью, решеткой, моделью малого мира. Понимание структуры сети является важным аспектом моделирования, так как учет реальных размеров сети обычно невозможен в силу ограниченности ресурсов.

Как правило, в задачах моделирования процессов сеть задается одним графом. Однако, при изучении эпидемий это не всегда может быть корректно, так как между индивидами в обществе возникают контакты разных видов. Поэтому следующей ступенью в изучении распространения инфекционных заболеваний встал вопрос об изменении одноуровневой сетевой модели. В связи с этим в задачах, требующих рассмотрения нескольких типов взаимодействий между агентами, необходимо использовать многоуровневые сети.

Авторы в работе «Generalized Epidemic Mean-Field Model for Spreading Processes Over Multilayer Complex Networks» [6] предлагают обобщенную модель GEMF (Generalized Epidemic Mean-Field), применимую к эпидемическим моделям, которые могут включать в себя любое число состояний и уровней сети. Таким образом, исследование процессов распространения становится более эффективным, при использовании такой модели.

К сожалению, анализ точных моделей довольно сложный, поэтому на практике используются аппроксимации точных моделей. Piet Van Mieghem в статье «The N-Intertwined SIS epidemic network model» [7] предложил модель NIMFA (N-Intertwined mean-field approximation) — аппроксимацию точной модели SIS. В работе «Virus spread in Networks» [8] в общем виде моделируется процесс распространения вируса на сети с применением теории Маркова для эпидемии, которая имеет два возможных состояния (группы). Найдена система обыкновенных дифференциальных уравнений для модели NIMFA, а также проведен сравнительный анализ этой модели и аналогичной ей точной модели.

На сегодняшний день имеется много научных исследований, посвященных изучению точных эпидемических моделей и их аппроксимаций. Причин такого интереса к данной теме может быть несколько.

Во-первых, конечно, с развитием науки люди смогли искоренить определенные виды инфекционных заболеваний, некоторые научились контролировать и не допускать их распространения среди населения. Однако, все еще существуют инфекционные заболевания, способные привести к летальному исходу, а также появляются новые, не изученные виды вирусов, которые могут спровоцировать вспышку эпидемии. К тому же у вирусов, вызывающих инфекционные заболевания, возникают мутации. Эта проблема была исследована в статье «Optimal Control of Influenza Epidemic Model with Virus Mutations» [9]. В ней рассматривается вариант эпидемии вируса гриппа среди населения, когда за один эпидемический сезон вирус гриппа мутирует, что приводит к циркуляции уже двух типов вируса в одной популяции. Классическая модель SIR, которая обычно применяется для изучения этого заболевания, не приспособлена для такого вида постановки задачи. Поэтому необходимо модернизировать ее, чтобы она учитывала динамику не только эпидемического процесса среди населения, но и мутации вируса.

Во-вторых, существуют различные аппроксимационные методы, которые, в большинстве случаев, могут быть применимы ко множеству эпидемических моделей. Однако, на сегодняшний момент не существует наилучшего метода аппроксимации, поэтому до сих пор возникают новые.

В работах [10] - [11] рассматриваются аппроксимации с использованием как методов теории среднего поля для модели SIS: N-intertwined

mean-field approximation (NIMFA), Heterogeneous mean-field method (HMF), Second-order NIMFA, так и методов Link percolation approach (LP), Message passing approach (MP). Каждый из методов имеет свои положительные и отрицательные стороны, поэтому необходимо выбирать тот, который применим к конкретной задаче.

В-третьих, на основе классических моделей создаются новые за счет исключения или добавления состояний, описывающих эпидемию. Поэтому, для любого процесса, поведение которого подобно эпидемическому, можно создать модель и изучить динамику этого процесса. Например, авторы F. D. Sahneh и C. Scoglio статьи «Epidemic Spread in Human Networks» [12] добавили новую группу A (alert - бдительный) к модели SIS и получили модель SAIS. Она предполагает, что восприимчивый индивидум может перейти в бдительное состояние A с вероятностью γ_A , если он окружен инфицированными соседями, а из этого состояния в инфицированное (I) с вероятностью $\gamma_I < \beta$, где β - вероятность перехода индивида из восприимчивого в инфицированное состояние.

Как уже говорилось, методы моделирования эпидемий могут также применяться вне эпидемиологии, расширяя тем самым область их применения и круг задач. Например, для изучения распространения информации в обществе, в задачах сетевой безопасности или в исследованиях вирусного маркетинга.

A. L.Hill, D. G. Rand, M. A. Nowak, N. A. Christakis провели интересное исследование, посвященное изучению распространения положительных и отрицательных эмоций у людей посредством социальных сетей. В статье «Emotions as infectious diseases in a large social network: the SISa model» [13] приводятся результаты этого исследования. Они предложили свою модель SISa, которая была получена на основе классической SIS модели, с помощью дополнительной стадии a, которая задает процесс спонтанной смены настроения у человека, который не зависит от контактов рассматриваемого индивида.

Например, в работе «Complete game-theoretic characterization of SIS epidemics protection strategies» [14] группа ученых используют аппроксимационную модель NIMFA для анализа процесса распространения вируса в полной сети. В этой статье формулируется задача о поиске оптимальной стратегии защиты игрока от вредоносного программного обеспечения. Для

постановки задачи применяется не только теория эпидемического моделирования, но и теория игр заполнения [15].

Рассмотрим еще один пример использования эпидемических моделей вне эпидемиологии. Для качественного налогового контроля необходимо осуществлять налоговые проверки, направленные на выявление недобросовестных налогоплательщиков, которые декларируют свои доходы не в полном объеме или не декларируют их вовсе. Проблема заключается в том, что проведение таких аудитов требует больших финансовых вложений. Поэтому для снижения расходов на проверки, а также для побуждения людей платить налоги, указывая весь доход, в работах [16] - [18] рассматривается возможность влияния на поведение налогоплательщиков с помощью распространения информации о предстоящем аудите.

Постановка задачи и целей

Динамические процессы, происходящие в окружающей среде, могут описываться с помощью математических моделей. Они задаются системой дифференциальных уравнений и моделируют эпидемические, социальные изменения в жизни общества, а также экономические процессы. В данной работе будем рассматривать принципы математического моделирования в задаче о системе налогообложения.

Таким образом, в рамках этой работы была определена цель: разработать модель, описывающую динамику распространения и влияния информации о налоговой проверке в популяции, в которой взаимодействие агентов описывается на многоуровневой сети.

Налоги и сборы составляют основу налоговой системы, а ее ключевой функцией является контроль уплаты налогов налогоплательщиками. Необходимость контроля состоит в том, что существуют налогоплательщики, которые предпочитают уклоняться от налогов. Однако, любой вид налогового аудита требует определённых затрат. Следовательно, при увеличении числа проверяемых налоговым органом, снижается эффективность проверки.

Необходимо ввести меры влияния на общество, позволяющие одновременно увеличивать налоговые выплаты и снижать затраты на налоговый контроль. Один из способов решения данной проблемы заключается в воздействии на общество с помощью распространения информации о предстоящей проверке.

Будем полагать, что в начальный момент времени рассмотрения процесса распространения часть налогоплательщиков обладает информацией о предстоящей налоговой проверке. Далее эта информация циркулирует среди агентов в рассматриваемом обществе. Так как время проведения налоговой проверки ограничено, то информация о ней может устаревать. Поэтому информированный агент, через определенный промежуток времени, может перестать обладать информацией.

Описанный процесс аналогичен распространению инфекции в обществе, который соответствует эпидемической модели SIS. Однако, в нашем случае, распространение информации о налоговой проверке имеет некоторые особенности, что приводит к усложнению этой модели.

Основной целью применения в сфере налогового контроля распространения информации в обществе является побуждение налогоплательщиков платить налоги и платить в полном объеме. В связи с этим, при создании модели, необходимо учесть влияние, которое оказывает информация на поведение человека, после того как он ее получил. Поэтому главное отличие от эпидемиологии состоит в том, что нужно не только изучить процесс распространения информации, но и проанализировать его влияние на общество.

Однако, каждый налогоплательщик принимает решение об уплате налога исходя из своих личных убеждений и в действительности все причины, оказывающие влияние на этот выбор неизвестны. Для того, чтобы приблизить наше исследование к реальности, мы будем моделировать распространение информации основываясь, на многоуровневой модели GEMF (Generalized Epidemic Mean-Field).

В соответствии с целью выпускной квалификационной работы основными задачами стали:

- Изучение обобщенной модели GEMF;
- Создание модели распространения информации о налоговом аудите в популяции;
- Разработка процедуры в пакете MATLAB для проведения численных экспериментов;
- Иллюстрация динамики распространения информации на многоуровневой сети модели SIAP (susceptible — infected — alert — pay);
- Проведение серий численных экспериментов, базирующихся на модели SIAP;
- Анализ полученных результатов.

Глава 1. Эпидемические модели

1.1 Модель SIS

Модель SIS (susceptible-infected-susceptible) [2] основана на предположении, что агенты, которые были инфицированы и вылечились, не приобрели иммунитет. В этом случае они возвращаются в первую группу восприимчивых. Эта модель задается уравнениями:

$$\frac{dS(t)}{dt} = -\frac{\beta S(t)I(t)}{N} + \gamma I(t); \quad (1)$$

$$\frac{dI(t)}{dt} = \frac{\beta S(t)I(t)}{N} - \gamma I(t). \quad (2)$$

В уравнении (1) описывается переход доли агентов, которые были инфицированы, из группы $S(t)$ в группу $I(t)$. Уравнение (2) описывает переход доли агентов из группы $I(t)$ снова в $S(t)$, т.е. агентов, которые были инфицированы, вылечились и не получили иммунитет. Данный процесс представлен на рис. 1.

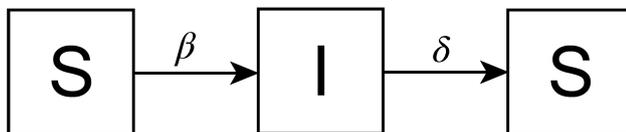


Рис. 1: Схема перехода агентов детерминированной модели SIS.

Если говорить об интерпретации динамики такой модели в сфере налогообложения, то под инфекцией может рассматривать информация, которая распространяется в популяции подобно эпидемии. Тогда состояние S может также соответствовать восприимчивому состоянию, а состояние I — информированному. Переход доли агентов из состояния S в I определяет процесс получения некоей экономической информации, например, о налоговом аудите. Обратный переход (из I в S) соответствует процессу устаревания информации, циркулирующей в популяции.

Уравнения (1) — (2) описывают процесс распространения (эпидемии, информации) детерминированной модели SIS, который с равной вероятностью может произойти с любым агентом в популяции вне зависимости

от его личных характеристик. Однако, игнорирование этого факта может привести к некорректным результатам моделирования. По этой причине следует рассматривать такие процессы на сети.

При помощи теории марковских процессов можно описать динамику распространения на сети. Для этого необходимо задать граф $G(N)$, который определяется симметричной матрицей смежности C и имеет N вершин. Узлы этого графа обозначают людей, а матрица C — контакты между ними.

Определение 1. Матрица смежности C графа $G(N)$ — это симметричная, бинарная матрица, элементы которой задаются следующим образом:

$$c_{ij} = \begin{cases} 1, & \text{если узел } i \text{ связан с узлом } j, \\ 0, & \text{иначе.} \end{cases}$$

В некоторых задачах могут быть использованы ориентированные, взвешенные графы. Однако, чаще всего эпидемические процессы моделируются на неориентированных и невзвешенных графах.

Применим вышеупомянутую теорию, например, к модели SIS. В этом случае каждый узел i в момент времени t будет находиться в одном из двух возможных состояний. Введем вектор, который будет характеризовать состояние узла i в момент времени t .

Определение 2. Вектор состояния узла i в момент времени t — это вектор $V_i(t) = [v_{i1}(t), v_{i2}(t), \dots, v_{iM}(t)]$, где $v_{ik}(t)$ — вероятность того, что узел i в момент времени t находится в состоянии k . Необходимым условием является: $v_{i1} + v_{i2} + \dots + v_{iM} = 1$.

Агенты, которые являются информированными, будем считать способными передавать информацию соседним агентам. Для узла i это состояние зададим так: $V_i(t) = [1, 0]$ (состояние инфицирования — I), где $v_{i1}(t) = 1, v_{i2} = 0$. Если же агент i в момент времени t не обладает информацией, то $V_i(t) = [0, 1]$ (восприимчивое состояние — S), где $v_{i1}(t) = 0, v_{i2} = 1$.

Для такой формулировки задачи запишем основное уравнение, соответствующее модели SIS, которое задает $v_{i1}(t)$ — вероятность того, что в

момент времени t узел i информирован [19]:

$$\frac{dv_{i1}(t)}{dt} = -\delta v_{i1}(t) + (1 - v_{i1}(t))\beta \sum_{j=1}^N a_{ij}v_{j1}(t). \quad (3)$$

Здесь β — вероятность получения информации восприимчивым агентом от информированного, δ — вероятность устаревания информации, которой обладает информированный агент. Вероятность того, что узел i в момент времени t восприимчив задается следующим образом:

$$v_{i2}(t) = 1 - v_{i1}(t). \quad (4)$$

Процесс изменения во времени состояний каждого узла популяции, которые задаются уравнениями (3) — (4), представляет собой цепь Маркова с 2^N состояниями для модели SIS [20].

Определение 3. Непрерывной цепью Маркова $\{X(t), t \geq 0\}$ называют процесс, для которого выполняется [21]:

$$\begin{aligned} P[X(t + \tau) = j | X(\tau) = i, X(u) = x(u), 0 \leq u < \tau] = \\ = P[X(t + \tau) = j | X(\tau) = i]. \end{aligned} \quad (5)$$

1.2 Модель SAIS

Если рассматривать динамику инфекционного заболевания в популяции, то при анализе такого процесса возникает вопрос о том, как влияет на распространение этой эпидемии поведение агентов в популяции. Имеют ли влияние индивидуальные реакции и решения членов популяции. Их можно описать следующим образом:

- 1) влияние изменения контактной топологии. Агенты популяции могут прекратить общение с другими людьми, которые являются инфицированными;
- 2) влияние профилактических мер на вероятность заражения. Для того, чтобы снизить вероятность заражения, индивид применяет различные профилактические меры, например, использует маски;
- 3) вакцинации на количество восприимчивых в системе. Количество вакцинированных в начальные моменты времени позволяет существенно снизить скорость и охват распространения инфекционных заболеваний.

Таким образом, индивидуальные реакции агентов являются защитной реакцией на возникновение инфекционного заболевания в популяции. Предполагается, что индивид не будет вредить себе и будет совершать действия, направленные на снижение вероятности инфицирования.

Рассмотрим модель SAIS (susceptible-alert-infected-susceptible), которая представляет собой модель SIS с добавленным состоянием A (alert) — предупрежден. Следовательно, каждый индивид может находиться теперь в одном из трех возможных состояний: S — восприимчив, A — предупрежден, I — инфицирован.

В этой модели индивид, находящийся в восприимчивом состоянии, может быть инфицирован с вероятностью β и восстановлен с вероятностью δ . Однако, в отличие от модели SIS, в этой модели возможны еще два перехода из восприимчивого состояния в предупрежденное с вероятностью γ_A , а также из предупрежденного в инфицированное — с вероятностью γ_I . Предполагается, что $0 \leq \gamma_I < \beta$.

Модель SAIS может быть применена к задаче о системе налогообложения. Тогда S и I будут соответствовать, как и в модели SIS, состояниям восприимчивости и информированности соответственно. В таком случае, для агентов переход из состояния S в I означает факт получения информации о предстоящей налоговой проверке и является благоприятным, в отличие от процесса инфицирования. Поэтому, если задавать состояние A в терминах распространения информации, то его также можно рассматривать как предупрежденное, а переход, осуществляемый из S в A , как получение другого вида информации. Например, дополнительная или неполная информация может спровоцировать активное взаимодействие агентов, а также стремление получить факты, удостовериться в подлинности информации об аудите. Поэтому вероятность γ_I , соответствующая переходу из состояния A в I , удовлетворяет условию: $0 \leq \beta < \gamma_I$.

Пусть $v_{i1}(t), v_{i2}(t), v_{i3}(t)$ — вероятности того, что узел i в момент времени t является восприимчивым, информированным или предупрежденным соответственно, а также выполняется равенство для узла $i \in \{1, \dots, N\}$:

$$1 = v_{i1}(t) + v_{i2}(t) + v_{i3}(t). \quad (6)$$

Таким образом, изменения вероятностей каждого узла i во времени задаются тремя дифференциальными уравнениями. Тогда для модели SAIS динамика системы описывается 3^N уравнениями. Однако, для каждого узла i одну из вероятностей, например, v_{i1} можно вычислить через остальные с помощью равенства (6). Тогда система будет иметь 2^N дифференциальных уравнений. Дифференциальные уравнения вероятностей v_{i2} и v_{i3} узла i модели SAIS имеют вид [12]:

$$\frac{dv_{i2}(t)}{dt} = \beta(1 - v_{i2} - v_{i3}) \sum_{j=1 \in N_i} c_{i,j} v_{j2} + \gamma_I v_{i3} \sum_{j=1 \in N_i} c_{i,j} v_{j2} - \delta v_{i2}; \quad (7)$$

$$\frac{dv_{i3}(t)}{dt} = \gamma_A(1 - v_{i2} - v_{i3}) \sum_{j=1 \in N_i} c_{i,j} v_{j2} - \gamma_I v_{i3} \sum_{j=1 \in N_i} c_{i,j} v_{j2}; \quad (8)$$

для $i \in \{1, \dots, N\}$.

Динамический процесс, который имеет три состояния и задается уравнениями (6) — (8), изображен на рис.2.

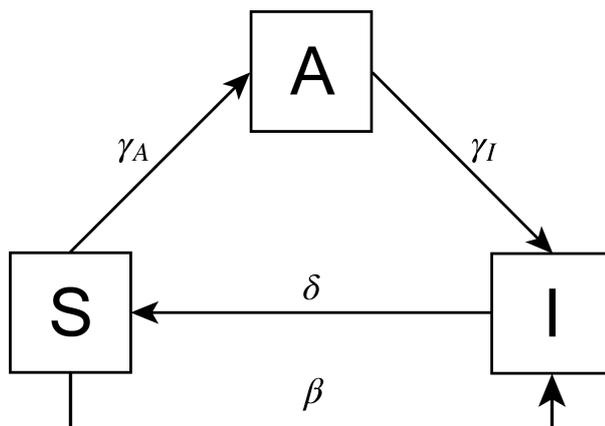


Рис. 2: Схема перехода агентов модели SAIS.

1.3 Модель GEMF

Рассмотрим обобщенную модель GEMF (Generalized Epidemic Mean-Field), которая позволяет моделировать любой процесс распространения на популяции, динамика которого подобна, например, модели SIS. Для этого необходимо задать структуру многоуровневой сети, количество возможных состояний агентов и правила их изменения.

Определение 4. Многоуровневая сеть — это сеть из L уровней, топология которой задается графами $G_l(N, Y_l)$. Каждый уровень имеет одинаковое количество узлов N и набор ребер Y_l . Множества Y_l различаются на каждом уровне l , где $l = \{1, \dots, L\}$. Заданная сеть изображена на рис.3.

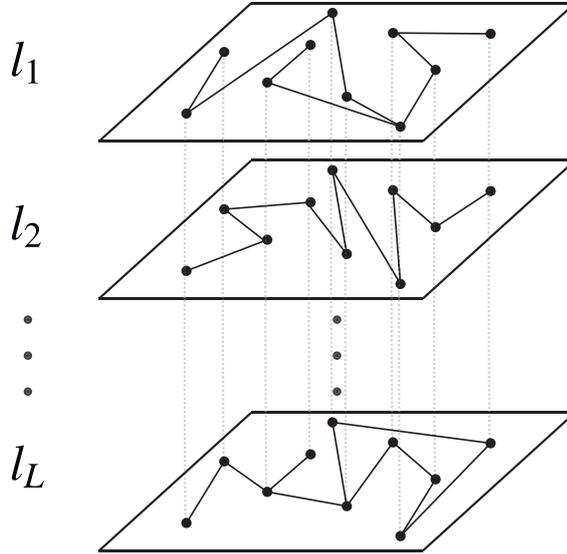


Рис. 3: Сетевые уровни в модели GEMF.

Использование многоуровневой сети позволяет учитывать различные типы взаимодействий агентов в популяции. Каждый уровень сети l соответствует определенному типу взаимодействия агентов и задается матрицей смежности C_l . Если узел i связан с узлом j на уровне l , то $c_{l,ij} = c_{l,ji} = 1$, иначе $c_{l,ij} = c_{l,ji} = 0$.

В общем случае будем считать, что в каждый момент времени t агент может находиться в одном из M состояний. Зададим правила их изменения. Рассмотрим два состояния $a, b \in \{1, M\}$. Пусть агент i , находясь в состоянии a , переходит в состояние b с вероятностью r_{ab} . Этот переход $a \rightarrow b$ может не зависеть от состояний соседей этого узла (процесс $I \rightarrow S$ в модели SIS, где $r_{IS} = \delta$) или наоборот, зависеть от взаимодействия с каждым узлом $j \in \{1, \dots, N\} \setminus \{i\}$, для которого выполняется условие $c_{ij} = c_{ji} = 1$

(процесс $S \rightarrow I$ в модели SIS, где $r_{SI} = \beta$). Дадим определения этим двум переходам.

Определение 5. Процесс, который происходит независимо от состояний других узлов будем называть узловой переход.

Определение 6. Процесс, который происходит в результате взаимодействия узлов будем называть сетевым переходом.

Далее будем считать, что если переход из состояния a в b является сетевым, то он происходит с вероятностью r_{ab} , а если узловым — то с вероятностью r_{ab}^δ .

Рассмотрим два состояния a и b , а также сетевой переход $a \rightarrow d$, где d — любое состояние, для которого выполняется: $d \in \{1, \dots, M\} \setminus \{a\}$. Предположим, что агент i находится в состоянии a , а агент j в состоянии b . Тогда: $v_{ia} = 1$, $v_{jb} = 1$, $c_{l,ij} = c_{l,ji} = 1$, где $l \in \{1, \dots, L\}$. Будем называть состояние b **влияющим состоянием** на уровне l , если при взаимодействии узлов i и j на этом уровне, узел i осуществляет сетевой переход $a \rightarrow d$.

Для каждого уровня l введем $q_l \in \{1, \dots, M\}$ — влияющее состояние, а также граф $\hat{\mathcal{G}}_l(\mathcal{M}, \mathcal{Y}_l)$. Влияющее состояние q_l будет задавать состояние, под действием которого происходит сетевой переход на данном уровне. Графы $\hat{\mathcal{G}}_l(\mathcal{M}, \mathcal{Y}_l)$ определим, как ориентированные, взвешенные графы, где $\mathcal{M} = \{1, \dots, M\}$ — набор узлов, который является постоянным для всех уровней сети и содержит возможные состояния системы. Набор ребер \mathcal{Y}_l определяется для каждого уровня l сети и задает на нем сетевые переходы.

Определение 7. Вес $\omega_l(a, b)$ ребра $(a, b)_l \in \mathcal{Y}_l$ графа $\hat{\mathcal{G}}_l(\mathcal{M}, \mathcal{Y}_l)$ — это вероятность, с которой происходит соответствующий переход из состояния a в b на уровне l .

Задавать граф $\hat{\mathcal{G}}_l(\mathcal{M}, \mathcal{Y}_l)$ будем двумя способами: с помощью матрицы смежности A_l и матрицы Кирхгофа Q_l . Матрицы являются квадратными, размерность, столбцы и строки которых, соответствуют количеству узлов графа, которые задают возможные состояния системы. Дадим определения этим понятиям.

Определение 8. Матрица смежности A_l взвешенного графа $\hat{\mathcal{G}}_l(\mathcal{M}, \mathcal{Y}_l)$ — это матрица, элементы которой задаются по следующему правилу: если на уровне l задан сетевой переход $a \rightarrow b$ с вероятностью $r_{l,ab}$, то $a_{l,ab} = r_{l,ab} = \omega_l(a, b)$, иначе $a_{l,ab} = 0$, где $a, b \in \{1, M\}$.

Определение 9. Матрица Кирхгофа Q_l — является одним из способов представления графа. Элементы этой матрицы задаются следующим образом:

$$q_{l,ab} = \begin{cases} \sum_{\substack{0 < b \leq \mathcal{M} \\ (a,b) \in \mathcal{Y}_l}} \omega(a,b) & \text{если } a = b; \\ -\omega(a,b) & (a,b)_l \in \mathcal{Y}_l; \\ 0 & \text{иначе.} \end{cases}$$

Для описания узловых переходов, необходимо задать один граф $\mathcal{G}^\delta(\mathcal{M}, \mathcal{Y}^\delta)$, не зависящий от уровня сети. Матрицы смежности A^δ и Кирхгофа Q^δ задаются аналогично соответствующим матрицам A_l и Q_l без учета сетевых уровней l .

С помощью матриц Кирхгофа и теории марковских процессов задается изменение вектора состояния V_i для узла $i \in \{1, \dots, N\}$ на многоуровневой сети обобщенной модели GEMF [6]:

$$\frac{dV_i(t)}{dt} = -(Q^\delta)^T V_i - \sum_{l=1}^L \left(\sum_{j=1}^N c_{l,ij} V_{j,q_l} \right) (Q_l)^T V_i; \quad (9)$$

для $i \in \{1, \dots, N\}$.

Дифференциальное уравнение (9) содержит два слагаемых. Первое задает изменение вектора состояния под влиянием узловых переходов, которые описываются матрицей Q^δ . Второе слагаемое задает сетевые переходы, возникающие на разных уровнях, которые описываются матрицами Q_l . Сумма $\sum_{j=1}^N c_{l,ij} V_{j,q_l}$ обозначает величину воздействия, которое оказывают узлы j , находящиеся во влияющем состоянии, на узел i на уровне l , где $j \in \{1, \dots, N\} \setminus \{i\}$, для которого выполняется условие $c_{l,ij} = c_{l,ji} = 1$.

На рис. 4 — 5 изображены примеры графов $\hat{\mathcal{G}}_l(\mathcal{M}, \mathcal{Y}_l)$ и $\mathcal{G}^\delta(\mathcal{M}, \mathcal{Y}^\delta)$ для сетевого перехода на уровне l и узлового соответственно. Для этого примера параметры следующие: $\mathcal{M} = \{1, 2, 3, 4, 5\}$, матрицы смежности A_l , A^δ и матрицы Кирхгофа Q_l , Q^δ имеют вид:

$$A_l = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & r_{l,42} & r_{l,43} & 0 & 0 \\ r_{l,51} & 0 & 0 & 0 & 0 \end{bmatrix}, Q_l = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -r_{l,42} & -r_{l,43} & r_{l,42} + r_{l,43} & 0 \\ -r_{l,51} & 0 & 0 & 0 & r_{l,51} \end{bmatrix},$$

$$A^\delta = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & r_{25}^\delta \\ 0 & 0 & 0 & 0 & 0 \\ r_{41}^\delta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, Q^\delta = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & r_{l,25}^\delta & 0 & 0 & -r_{l,25}^\delta \\ 0 & 0 & 0 & 0 & 0 \\ -r_{l,41}^\delta & 0 & 0 & r_{l,41}^\delta & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

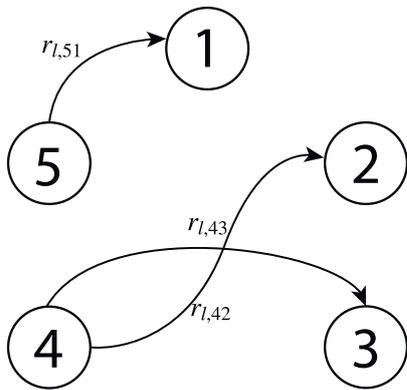


Рис. 4: Граф сетевого перехода.

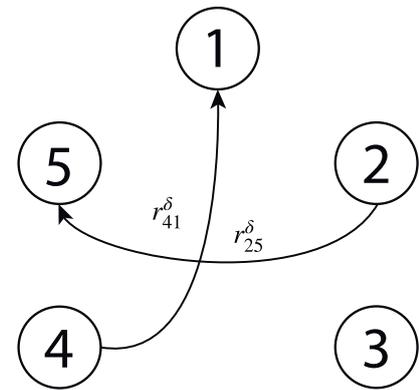


Рис. 5: Граф узлового перехода.

Для модели SIS рассмотрим пример контактной сети с одним уровнем $L = 1$ и $q_1 = I$, которая имеет $N = 5$ узлов.

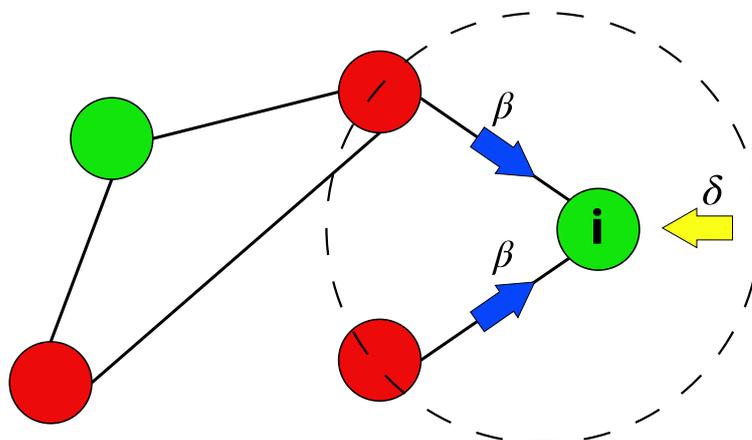


Рис. 6: Контактная сеть для модели SIS, представленная графом $G_1(5)$.

На рис. 6 красным обозначены инфицированные узлы, зеленым — здоровые. Рассмотрим узел i , у которого $V_i(t) = [0, 1] - S$. Этот узел имеет двух соседей, состояния которых совпадают с компонентной влияния этого уровня: $q_1 = I$. Поэтому для узла i возможен сетевой переход $S \rightarrow I$ с вероятностью $r_{SI} = \beta$, а также, после инфицирования, узловой переход $I \rightarrow S$ с вероятностью $r_{IS} = \delta$.

Глава 2. Модель распространения информации о налоговых проверках

2.1 Описание модели SIAP

В предыдущих исследованиях было показано, что процесс распространения информации в популяции подобен эпидемическому процессу [22]. Однако, способы передачи вируса и информации отличаются, так как распространение информации происходит не только через контактную сеть, но и с помощью СМИ. Поэтому будем рассматривать популяцию, в которой взаимодействие агентов осуществляется через многоуровневую сеть. Она имеет $L = 3$ уровня и задается с помощью неориентированных графов $G_l(N, Y_l)$, где N — множество узлов, Y_l — множество ребер графов каждого уровня, $l = \{1, 2, 3\}$.

Для описания распространения и влияния нескольких типов информации о предстоящей налоговой проверке введем модель SIAP (susceptible — infected — alert — pay), которая задает $M = 4$ состояния агентов популяции: S — восприимчивые, A — предупрежденные, I — инфицированные, P — платят налоги. На рис. 7 представлена общая схема данной модели.

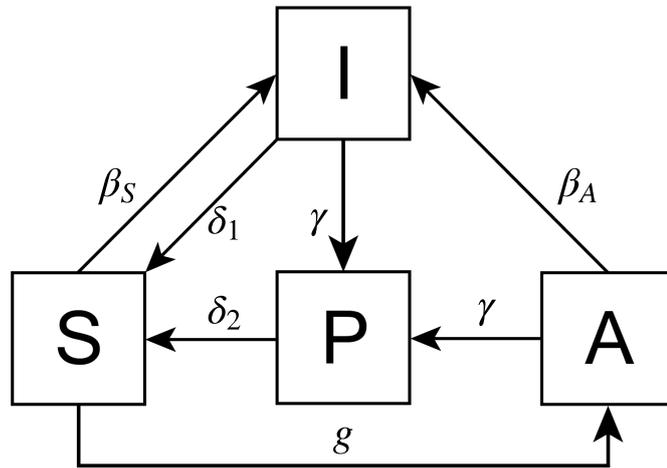


Рис. 7: Схема перехода агентов модели SIAP.

В данной модели каждый уровень имеет различные матрицы смежности C_l , влияющие состояния q_l и графы сетевых переходных вероятностей, которые задаются матрицами A_l и Q_l . Общим для всех уровней, помимо множества узлов N , являются число состояний $M = 4$ и два узловых перехода, определяемых при помощи матриц смежности A^δ и Кирхгофа

Q^δ . При описании матриц, задающих графы сетевых и переходных вероятностей на трех уровнях, будем полагать, что состоянию S соответствует первый столбец и строка, состоянию I — второй столбец и строка, а состояниям A и P — третий и четвертый соответственно.

Опишем каждый уровень сети отдельно. Пусть базовый уровень $l = 1$ обозначает отношения между ближайшим окружением агента, которые задаются матрицей смежности C_1 . На этом уровне влияющим состоянием q_1 является состояние A , а матрица сетевых переходных вероятностей имеет вид:

$$A_1 = \begin{bmatrix} 0 & 0 & g & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Таким образом, на данном уровне агент может перейти из восприимчивого состояния в предупрежденное за счет получения информации о проверке от своих друзей и знакомых — людей, с которыми он имеет личный контакт. Будем полагать, что такая информация является не подтвержденной и носит характер слухов.

Следующий уровень $l = 2$ имеет матрицу смежности C_2 , которая характеризует распространение информации через СМИ. Влияющим состоянием q_2 является состояние I . Здесь возможны два перехода: из S в I с вероятностью β_S и из A в I с вероятностью β_A . Запишем для них матрицу сетевых переходных вероятностей:

$$A_2 = \begin{bmatrix} 0 & \beta_S & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \beta_A & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Для этого уровня будем задавать сеть, структура которой отличается от контактной. В данной модели рассматриваются три источника официальной информации, такие как телевидение, радио и новостные порталы сети Интернет, которые на графе задаются узлами W_1, W_2, W_3 .

Так как в модели предполагается, что сетевые уровни имеют одинаковый набор узлов N , то узлы, обозначающие источники СМИ, распо-

лагаются на каждом уровне. Однако, они не имеют связей на первом и третьем уровнях с другими узлами, обозначающими агентов популяции. Это связано с тем, что наборы ребер Y_1 и Y_3 графов $G_1(N, Y_1)$ и $G_3(N, Y_3)$ соответственно задают между агентами взаимоотношения, определяемые личными контактами. Поэтому в графе второго уровня сохраняется количество узлов, однако меняется набор ребер. Ребра из множества Y_2 соединяют узлы СМИ с агентами, которые получают информацию из этих источников. Связи между узлами, не являющимися СМИ, отсутствуют.

Количество связей, которые генерируются для каждого источника информации, задается заранее. Так как при моделировании процесса распространения связи образуются случайным образом, некоторые узлы, обозначающие агентов, могут быть соединены сразу с несколькими источниками информации или совсем не иметь связей на этом уровне. Это связано с тем, что в реальном мире агент может получать информацию сразу из нескольких источников или не получать официальную информацию вовсе. Пример графа второго уровня модели SIAP приведен на рис. 8.

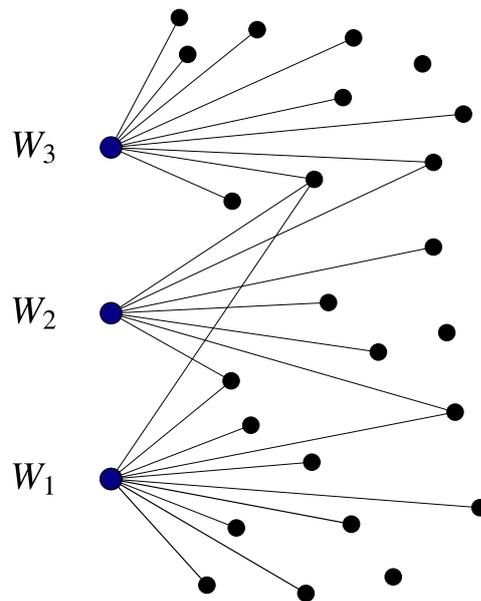


Рис. 8: Пример графа уровня $l = 2$.

Так как период проведения налогового аудита ограничен, то данный тип информации теряет свою актуальность через определенный промежуток времени.

На последнем уровне $l = 3$, который задается матрицей смежности первого контактного уровня C_1 , влияющим состоянием q_3 является состояние I . Однако, на этом уровне происходят два перехода из состояний A и

I в состояние P с вероятностью γ . Матрица A_3 имеет вид:

$$A_3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma \\ 0 & 0 & 0 & \gamma \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Этот уровень предполагает, что если агент, находясь в состоянии предупрежден или информирован, получает официальную информацию о предстоящей проверке от других агентов популяции, то он решает платить налог.

На рассматриваемом уровне, как и на предыдущем, распространяется официальная информация, поэтому необходимо учитывать устаревание информации в модели. В связи с этим зададим два перехода: $I \rightarrow S, P \rightarrow S$ с вероятностями δ_1 и δ_2 , соответственно. Для каждого агента популяции эти процессы не зависят от состояний его соседей, следовательно, являются узловыми и задаются матрицей:

$$A^\delta = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \delta_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \delta_2 & 0 & 0 & 0 \end{bmatrix}.$$

На основе матриц переходных вероятностей построим соответствующие матрицы Кирхгофа Q_l и Q^δ для сетевых и узловых переходов, где $l = 1, 2, 3$.

$$Q_1 = \begin{bmatrix} g & 0 & -g & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad Q_2 = \begin{bmatrix} \beta_S & -\beta_S & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -\beta_A & \beta_A & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad Q_3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma & -\gamma \\ 0 & 0 & 0 & 0 \end{bmatrix};$$

$$Q^\delta = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -\delta_1 & \delta_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\delta_2 & 0 & 0 & \delta_2 \end{bmatrix}.$$

На рис. 9 — 12 изображены графы сетевых и узловых переходных вероятностей обобщенной модели GEMF, рассматриваемой на модели SIAP.

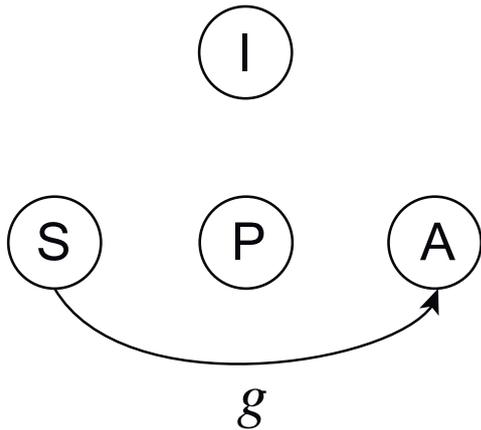


Рис. 9: Граф сетевых переходных вероятностей первого уровня.

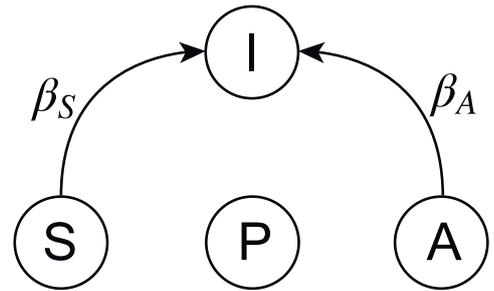


Рис. 10: Граф сетевых переходных вероятностей второго уровня.

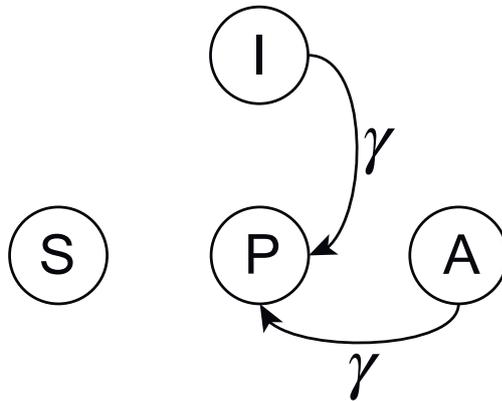


Рис. 11: Граф сетевых переходных вероятностей третьего уровня.

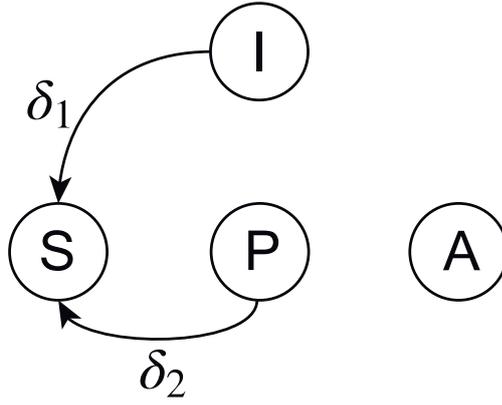


Рис. 12: Граф узловых переходных вероятностей.

Таким образом, запишем систему дифференциальных уравнений обобщенной модели GEMF на основе модели SIAP:

$$\left\{ \begin{array}{l}
 \frac{dV_1(t)}{dt} = -(Q^\delta)^T V_1 - \sum_{j=1}^N c_{1,1j} V_{j,A}(t) Q_1^T V_1 - \sum_{j=1}^N c_{2,1j} V_{j,I}(t) Q_2^T V_1 \\
 \quad - \sum_{j=1}^N c_{3,1j} V_{j,I}(t) Q_3^T V_1; \\
 \quad \vdots \\
 \frac{dV_i(t)}{dt} = -(Q^\delta)^T V_i - \sum_{j=1}^N c_{1,ij} V_{j,A}(t) Q_1^T V_i - \sum_{j=1}^N c_{2,ij} V_{j,I}(t) Q_2^T V_i \\
 \quad - \sum_{j=1}^N c_{3,ij} V_{j,I}(t) Q_3^T V_i; \\
 \quad \vdots \\
 \frac{dV_N(t)}{dt} = -(Q^\delta)^T V_N - \sum_{j=1}^N c_{1,Nj} V_{j,A}(t) Q_1^T V_N - \sum_{j=1}^N c_{2,Nj} V_{j,I}(t) Q_2^T V_N \\
 \quad - \sum_{j=1}^N c_{3,Nj} V_{j,I}(t) Q_3^T V_N;
 \end{array} \right. \quad (10)$$

где $i \in \{1, \dots, N\}$.

2.2 Процедура моделирования процесса распространения информации модели SIAP

В предыдущей главе была введена модель SIAP, описывающая поведение налогоплательщиков в популяции, в которой распространяется информация о предстоящей налоговой проверке. Для иллюстрации динамики данной модели, ее изучения, а также для дальнейшего использования в исследованиях в области аудита был написан скрипт в пакете MATLAB. Он позволяет моделировать процесс распространения информации, соответствующий заданной модели SIAP, поэтому количество уровней сети, число возможных состояний системы, а также правила их изменения заранее определены. Таким образом необходимо задать следующие параметры:

- 1 Множество узлов N ;
- 2 Параметры системы $\beta_S, \beta_A, \gamma, \delta_1, \delta_2, g$;
- 3 Начальный и конечный моменты времени моделирования процесса: t_0 и t_f ;
- 4 Доли агентов каждого состояния в начальный момент времени.

Сетевые уровни задаются с помощью матриц смежности C_l ($l = \{1, 2, 3\}$) в отдельной процедуре. Поэтому, в ней можно внести изменения для одной или нескольких матриц, если по условию задачи необходимо использовать матрицу определенного вида.

Конечно, использование реальных контактных сетей позволит эффективно изучить распространение информации в популяции и получить точные результаты. Однако, на практике могут возникнуть проблемы со сбором и обработкой данных.

В данном исследовании использовались несколько видов матриц смежности, что позволило провести сравнительный анализ результатов моделирования. В первой и во второй сериях экспериментов матрицы смежности генерировались случайным образом с $N = 100$ и $N = 1000$ узлами соответственно. В последующих сериях использовались три социальных графа, полученные в результате фактического сбора данных. Они представляют контактную сеть трех гуманитарных колледжей США в социальной сети Facebook: Amherst College, Oberlin College и Bowdoin College, которые

далее будем обозначать Facebook1, Facebook2, Facebook3 соответственно. Данные были взяты из открытого онлайн-репозитория — The Network Data Repository [23].

На рис. 13 представлены основные характеристики, которые были посчитаны для каждого графа: количество узлов — N , количество связей — E , минимальная степень узла — D_{min} , максимальная степень узла — D_{max} , средняя степень узла — D_{av} . Такие размеры графов позволили провести исследование модели SIAP на реальных данных с использованием персональной ЭВМ (ноутбук).

	Facebook1	Facebook2	Facebook3
N	2235	2920	2252
E	91000	89900	84400
D_{min}	1	1	1
D_{max}	467	478	670
D_{av}	81	61	74

Рис. 13: Основные характеристики графов Facebook1, Facebook2. Facebook3.

Определение 10. Степень узла графа — количество ребер, выходящих из этой вершины.

Для оценки состояния системы введем вектор — $Y(t) = [y_S(t), y_I(t), y_A(t), y_P(t)]$. Его компоненты являются функции, которые задают средние вероятности восприимчивости, информированности, предупрежденности и платы, соответственно:

$$\begin{aligned}
 y_S(t) &= \frac{\sum_{i=1}^N v_{i1}(t)}{N}, & y_I(t) &= \frac{\sum_{i=1}^N v_{i2}(t)}{N}, \\
 y_A(t) &= \frac{\sum_{i=1}^N v_{i3}(t)}{N}, & y_P(t) &= \frac{\sum_{i=1}^N v_{i4}(t)}{N}.
 \end{aligned} \tag{11}$$

Вектор $Y(0)$ — определяет доли состояний S, I, A, P в момент времени $t = t_0 = 0$. В данной модели предполагается, что в популяции налогоплательщиков распространение достоверной информации происходит из $k_I = 3$ узлов, которые обозначают официальные источники СМИ: W_1, W_2, W_3 .

Поэтому начальная доля информированных агентов равна:

$$y_I(t_0) = \frac{k_I}{N} 100\%. \quad (12)$$

Помимо использования реальных графов, сложность при моделировании заключается также в выборе начальных данных: параметров системы (вероятности перехода — $\beta_S, \beta_A, \gamma, g, \delta_1, \delta_2$) и начальных долей состояний системы ($Y(0) = [y_S(0), y_I(0), y_A(0), y_P(0)]$). В данной работе будем предполагать, что параметры модели SIAP следующие: $\beta_S = 0.2, \beta_A = 0.4, \gamma = 0.3, g = 0.5, \delta_1 = 0.4, \delta_2 = 0.2$. Выбор данных значений обусловлен структурой модели SIAP. Однако в процедуре "SIAP.m" возможно задание любого значения как для параметров, так и для начальных долей состояний системы.

Далее будем считать, что в начальный момент времени в системе присутствует приблизительно 3% предупрежденных агентов, а также порядка 5% тех, кто готов платить налоги в полном объеме. Остальная часть агентов в популяции будет находиться в восприимчивом состоянии — S . Использование примерных величин обусловлено тем, что начальное количество узлов, которое необходимо задать, является целым числом. Поэтому при моделировании используется целое число узлов, которое является ближайшим значением к заданным процентам.

Для эффективного распространения информации в популяции, перед моделированием самого процесса, происходит анализ контактной сети: формируется ранжированный список, содержащий вычисленные степени каждого узла системы. Далее узлы, которые имеют наибольшее количество ребер, в соответствии с заданной начальной долей, определяются как предупрежденные в момент времени $t = t_0$. Распространение на первом уровне модели SIAP недостоверной информации происходит из этих узлов.

Результатом исполнения скрипта при моделировании процесса распространения информации в популяции являются графики функций $y_S(t), y_I(t), y_A(t), y_P(t)$ и файл, в котором записываются следующие выходные данные:

- 1 Параметры системы;
- 2 Вектор средних вероятностей состояний системы в начальный момент

времени — $Y(0)$;

3 Вектор средних вероятностей состояний системы в момент времени $t = t_f$: $Y(t_f)$.

В репозитории на [GitHub.com](https://github.com) [24] представлен код скрипта и всех процедур, необходимых для численного моделирования поставленной задачи.

Глава 3. Численное моделирование

3.1 Первая серия экспериментов

Проиллюстрируем работу скрипта "SIAP.m". Для этого рассмотрим процесс распространения информации в популяции со следующими параметрами: $N = 100, t_f = 10, \beta_S = 0.2, \beta_A = 0.4, \gamma = 0.3, \delta_1 = 0.4, \delta_2 = 0.2, g = 0.9$. На рис. 14 приведен график этого моделирования.

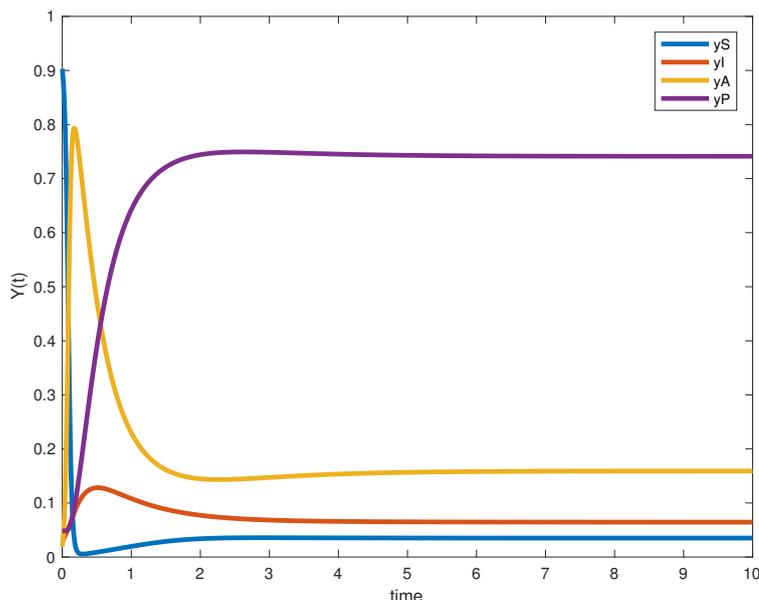


Рис. 14: Траектории функций $y_S(t), y_I(t), y_A(t), y_P(t)$ модели SIAP.

На графике представлены компоненты вектора $Y(t)$ — функции средних вероятностей состояний системы: $y_S(t), y_I(t), y_A(t), y_P(t)$ в зависимости от времени моделирования t . Таким образом, можно наблюдать начальные доли агентов каждого состояния и их последующие изменения.

В исследуемой модели SIAP предполагается, что в популяции налогоплательщиков имеется информация двух видов: официальная и слухи. Распространение официальной информации из достоверных источников соответствует сетевым переходам на втором уровне из групп S и A в состояние I с вероятностями β_S и β_A соответственно, а также на третьем уровне: из I и A в P с вероятностью γ . Переход из восприимчивого состояния в предупрежденное возможен на первом уровне с вероятностью g и происходит под действием недостоверной информации — слухов. Они распространяются через контактную сеть популяции.

С целью изучения влияния слухов на процесс распространения в модели SIAP была проведена серия экспериментов на случайном графе —

Rand100, в которых варьировался параметр g от 0.95 до 0.05 с шагом 0.05. Остальные параметры: $N = 100, t_f = 5, \beta_S = 0.2, \beta_A = 0.4, \gamma = 0.3, \delta_1 = 0.4, \delta_2 = 0.2$. В этой серии 19 экспериментов, поэтому приведена часть графиков, которые изображены на рис. 16 — 20. Для каждого графика записаны начальные и стационарные состояния функции $y_P(t)$. Данные по всем экспериментам представлены ниже.

Rand100						
№	g	S (%)	I (%)	A (%)	P (%)	
1	0,95	3,39	6,45	15,96	74,20	
2	0,9	3,51	6,46	15,91	74,12	
3	0,85	3,65	6,46	15,86	74,03	
4	0,8	3,80	6,47	15,81	73,93	
5	0,75	3,97	6,47	15,75	73,81	
6	0,7	4,16	6,48	15,67	73,68	
7	0,65	4,39	6,49	15,59	73,53	
8	0,6	4,66	6,50	15,50	73,35	
9	0,55	4,97	6,51	15,38	73,13	
10	0,5	5,35	6,52	15,25	72,88	
11	0,45	5,82	6,54	15,08	72,56	
12	0,4	6,40	6,56	14,87	72,16	
13	0,35	7,15	6,59	14,61	71,65	
14	0,3	8,17	6,63	14,25	70,95	
15	0,25	9,60	6,68	13,75	69,96	
16	0,2	11,79	6,77	13,00	68,44	
17	0,15	15,53	6,91	11,74	65,82	
18	0,1	23,39	7,22	9,31	60,08	
19	0,05	48,61	8,13	2,35	40,92	

Рис. 15: Данные экспериментов первой серии на случайном графе Rand100.

В этой серии экспериментов доля агентов группы S в момент времени $t = t_0$ составляла 92.72% от всей популяции. На рис. 16 можно видеть скачок функции y_A , который связан с тем, что с вероятностью $g = 0.95$ основная часть агентов популяции, которая находится в группе S , переходит в A .

Далее, при уменьшении параметра g , максимум функции y_A уменьшается. Более того, время достижения стационарных состояний функций $y_S(t), y_I(t), y_A(t), y_P(t)$ увеличивается.

В эксперименте 17 (рис. 18) стационарное состояние функции y_P не достигается при таком же времени моделирования, как у предыдущих экспериментов — $t = t_f = 5$. По этой причине для оставшихся экспериментов, изображенных на рис.18 — 20, время моделирования было увеличено —

$t = t_f = 10$. После этого все функции состояний системы достигли своих стационарных состояний.

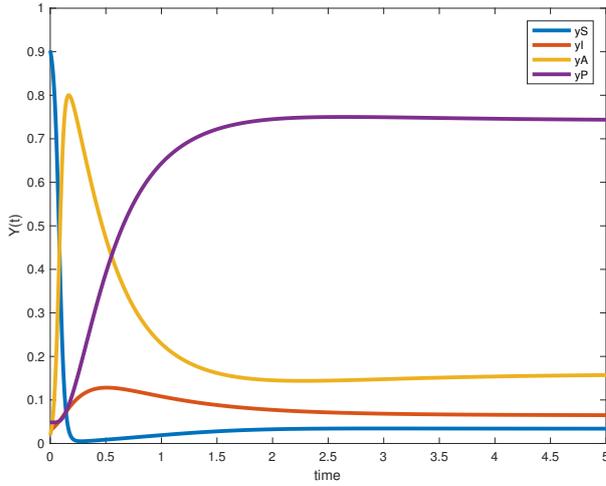


Рис. 16: Rand100. Эксперимент №1: $g = 0.95$; начальное состояние: $t = 0, y_P = 0.0485$; стационарное состояние: $t = 4.687, y_P = 0.742$.

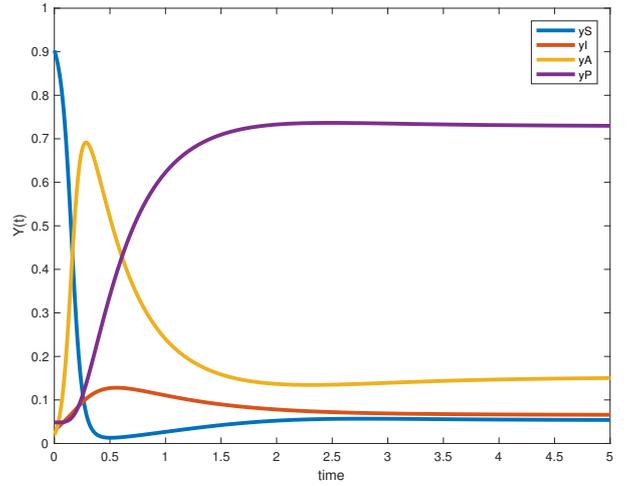


Рис. 17: Rand100. Эксперимент №10: $g = 0.5$; начальное состояние: $t = 0, y_P = 0.0485$; стационарное состояние: $t = 4.477, y_P = 0.7288$.

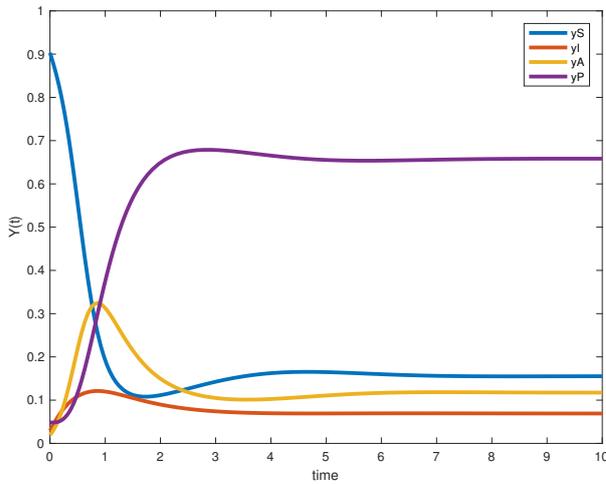


Рис. 18: Rand100. Эксперимент №17: $g = 0.15$; начальное состояние: $t = 0, y_P = 0.0485$; стационарное состояние: $t = 9.337, y_P = 0.6582$.

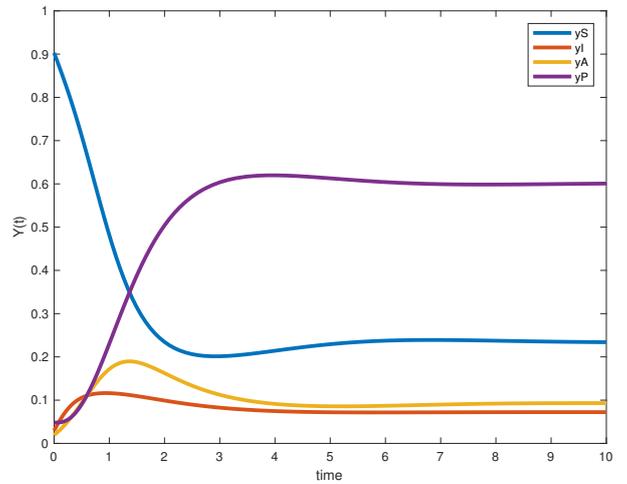


Рис. 19: Rand100. Эксперимент №18: $g = 0.1$; начальное состояние: $t = 0, y_P = 0.0485$; стационарное состояние: $t = 9.884, y_P = 0.6008$.

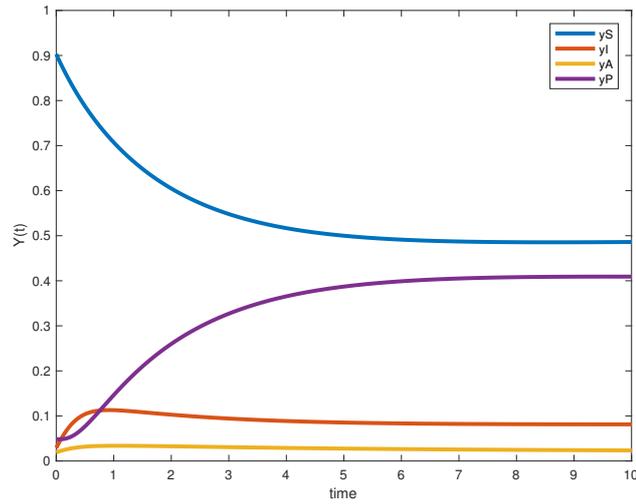


Рис. 20: Rand100. Эксперимент №19: $g = 0.05$; начальное состояние: $t = 0, y_P = 0.0485$; стационарное состояние: $t = 9.28, y_P = 0.4092$; $t = 9.76, y_S = 0.4861$.

Из результатов моделирования видно, что при уменьшении параметра g , изменяются конечные значения всех функций системы: значения функций $y_S(t)$ и $y_I(t)$ — увеличиваются, а значения функций $y_A(t)$ и $y_P(t)$ — уменьшаются, где $t = t_f$. Так как g уменьшается постепенно на 0.05, а остальные параметры не изменяются, то количество агентов, осуществляющих переход из S в A уменьшается. Следовательно, доля агентов группы P также уменьшается.

В экспериментах, где параметр g принимает значения от 0.95 до 0.1, большая часть системы находится в группе P в момент времени $t = t_f$. Однако в эксперименте на рис. 20, при $g = 0.05$, значения функции $y_P(t)$ не превышали значений функции $y_S(t)$ на протяжении всего времени моделирования.

3.2 Вторая серия экспериментов

По причине того, что реальные контактные сети популяции имеют большие размеры, была проведена еще серия экспериментов на случайном графе Rand1000 с такими же параметрами системы: $\beta_S = 0.2$, $\beta_A = 0.4$, $\gamma = 0.3$, $\delta_1 = 0.4$, $\delta_2 = 0.2$. Однако, размер популяции увеличился — $N = 1000$. Параметр g будем изменять в том же интервале от 0.95 до 0.05 с шагом 0.05. Предполагается, что динамика системы при изменении параметра g сохранится в популяции большего размера. Данные по всем экспериментам приведены ниже.

Rand1000						
№	g	S (%)	I (%)	A (%)	P (%)	
1	0,95	0,59	1,60	7,47	90,34	
2	0,9	0,62	1,60	7,46	90,32	
3	0,85	0,65	1,60	7,45	90,30	
4	0,8	0,69	1,60	7,45	90,27	
5	0,75	0,73	1,60	7,43	90,24	
6	0,7	0,78	1,60	7,42	90,20	
7	0,65	0,83	1,60	7,41	90,16	
8	0,6	0,90	1,60	7,39	90,11	
9	0,55	0,97	1,61	7,37	90,05	
10	0,5	1,07	1,61	7,35	89,98	
11	0,45	1,18	1,61	7,32	89,89	
12	0,4	1,32	1,62	7,28	89,79	
13	0,35	1,50	1,62	7,23	89,65	
14	0,3	1,75	1,63	7,17	89,46	
15	0,25	2,10	1,64	7,08	89,19	
16	0,2	2,62	1,65	6,94	88,78	
17	0,15	3,52	1,68	6,72	88,08	
18	0,1	5,40	1,73	6,26	86,60	
19	0,05	11,76	1,90	4,86	81,47	

Рис. 21: Данные экспериментов второй серии.

Из приведенных выше данных можно видеть, что тенденция изменения функций средних вероятностей состояний системы сохраняется на популяции большего размера. Время моделирования процесса и достижения стационарных состояний функциями $y_S(t)$, $y_I(t)$, $y_A(t)$, $y_P(t)$ уменьшилось незначительно. Это связано с тем, что с увеличением числа узлов в системе в 10 раз, также в 10 раз увеличилось среднее количество связей для каждого узла. Поэтому отношение среднего числа связей узла (D_{av}) к общему количеству узлов не изменилось.

На рис. 22 — 25 представлены некоторые графики из этой серии экс-

периментов. Если сравнить их попарно, то можно заметить, что изменения функций средних вероятностей состояний системы на рис.22 — 23 практически не происходят. Зато на рис. 24 — 25, которые соответствуют последним двум экспериментам с параметром $g = 0.1$ и $g = 0.05$ соответственно, заметны изменения функций $y_S(t)$, $y_I(t)$, $y_A(t)$, $y_P(t)$. Такая закономерность, только более явная, наблюдалась и в предыдущей серии экспериментов.

Однако в данной серии экспериментов имеется существенное отличие. Оно заключается в том, что при любом значении параметра g , больше 80% агентов системы находятся в состоянии P в момент времени $t = t_f$. Тогда как в эксперименте №19 предыдущей серии (рис. 20) доля агентов состояния P составляет 40.92%.

Таким образом, размер сети также влияет на время моделирования процесса распространения. Кроме того, рост числа узлов и среднего количества узлов в случайном графе приводит к снижению чувствительности модели к изменению параметров системы.

Далее выполним моделирование процесса распространения информации на социальных графах и проведем сравнительный анализ результатов моделирования.

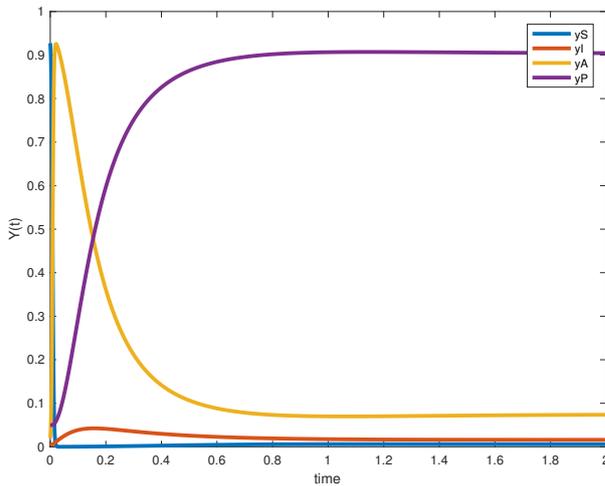


Рис. 22: Rand1000. Эксперимент №1: $g = 0.95$; начальное состояние: $t = 0, y_P = 0.0498$; стационарное состояние: $t = 1.452, y_P = 0.9034$.

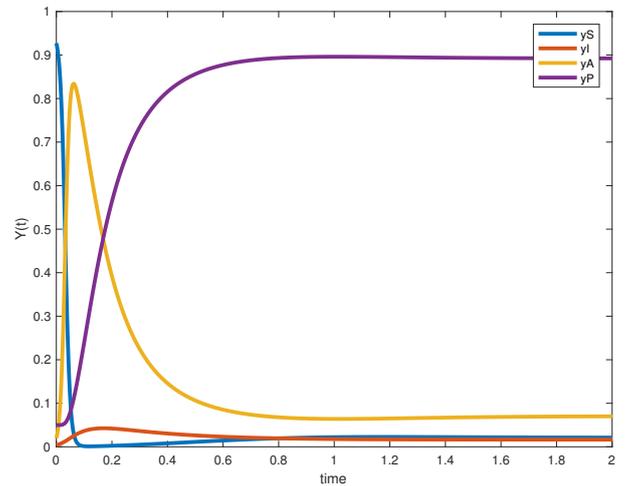


Рис. 23: Rand1000. Эксперимент №15: $g = 0.25$; начальное состояние: $t = 0, y_P = 0.0498$; стационарное состояние: $t = 1.892, y_P = 0.8919$.

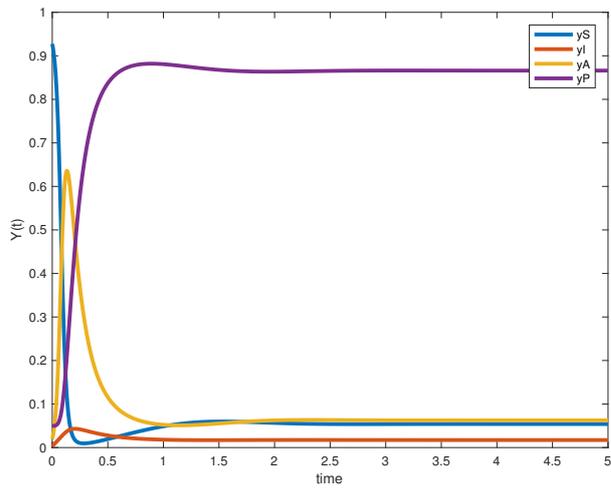


Рис. 24: Rand1000. Эксперимент №18: $g = 0.1$; начальное состояние: $t = 0, y_P = 0.0498$; стационарное состояние: $t = 3.952, y_P = 0.8660$.

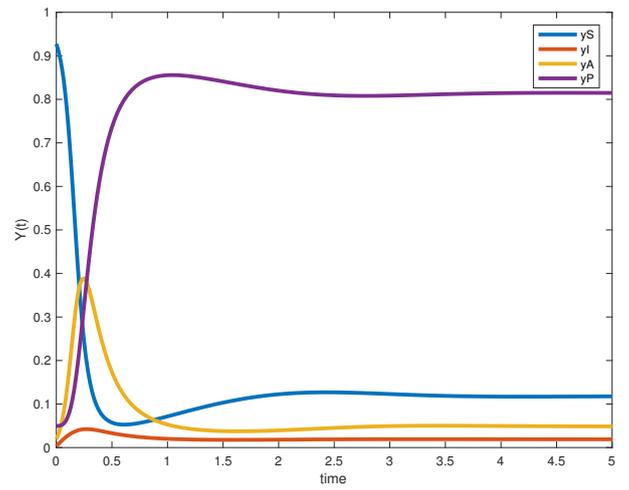


Рис. 25: Rand1000. Эксперимент №19: $g = 0.05$; начальное состояние: $t = 0, y_P = 0.0498$; стационарное состояние: $t = 4.662, y_P = 0.8147$.

3.3 Третья серия экспериментов

Как уже говорилось, написанный скрипт позволяет моделировать процесс распространения на любой сети. До этого использовались графы, которые генерировались случайным образом. Однако для того, чтобы получить результаты, приближенные к действительности, необходимо использовать реальные данные сетей. Поэтому был взят граф Facebook1 [23] и на нем рассмотрен процесс распространения модели SIAP.

Для того, чтобы сравнить результаты, полученные при моделировании процесса на случайном графе с результатами на реальной сети, будем варьировать параметр g от 0.95 до 0.05 с шагом 0.05. Зададим остальные параметры системы: $t_f = 7$, $\beta_S = 0.2$, $\beta_A = 0.4$, $\gamma = 0.3$, $\delta_1 = 0.4$, $\delta_2 = 0.2$. Граф Facebook1 имеет $N = 2235$ узлов. Начальные доли состояний системы следующие: $S - 92.94\%$, $I - 0.13\%$, $A - 1.97\%$, $P - 4.96\%$.

Данные всех экспериментов приведены на рис. 26. Ниже, на рис. 27 — 30, выборочно представлены графики данной серии.

Facebook1							
№		g		S (%)	I (%)	A (%)	P (%)
1		0,95		5,42	9,39	12,91	72,28
2		0,9		5,65	9,38	12,81	72,16
3		0,85		5,89	9,38	12,70	72,03
4		0,8		6,17	9,37	12,58	71,88
5		0,75		6,48	9,37	12,44	71,71
6		0,7		6,82	9,36	12,29	71,52
7		0,65		7,21	9,36	12,13	71,31
8		0,6		7,65	9,35	11,94	71,06
9		0,55		8,17	9,35	11,72	70,76
10		0,5		8,77	9,34	11,47	70,42
11		0,45		9,48	9,34	11,18	70,00
12		0,4		10,34	9,34	10,83	69,48
13		0,35		11,41	9,35	10,41	68,83
14		0,3		12,78	9,36	9,88	67,98
15		0,25		14,62	9,38	9,19	66,81
16		0,2		17,21	9,41	8,26	65,11
17		0,15		21,24	9,49	6,88	62,39
18		0,1		28,49	9,55	4,66	57,29
19		0,05		45,79	9,90	0,52	43,79

Рис. 26: Данные экспериментов третьей серии для графа Facebook1.

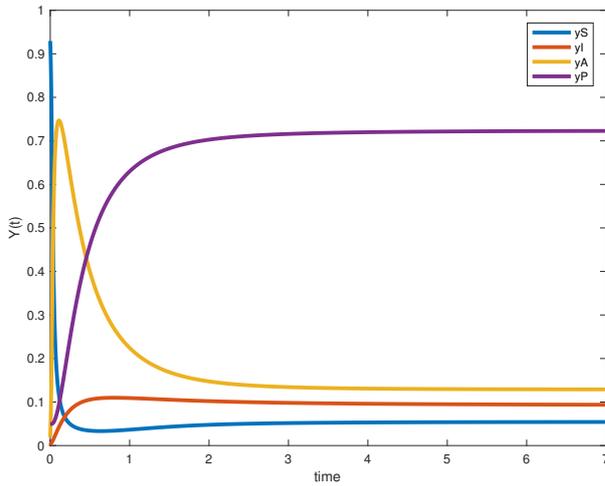


Рис. 27: Facebook1. Эксперимент №1: $g = 0.95$; начальное состояние: $t = 0, y_P = 0.0496$; стационарное состояние: $t = 6.956, y_P = 0.7228$.

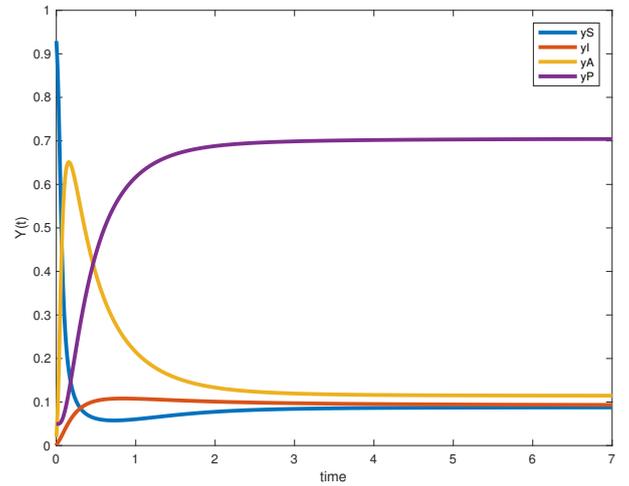


Рис. 28: Facebook1. Эксперимент №10: $g = 0.5$; начальное состояние: $t = 0, y_P = 0.0496$; стационарное состояние: $t = 6.971, y_P = 0.7042$.

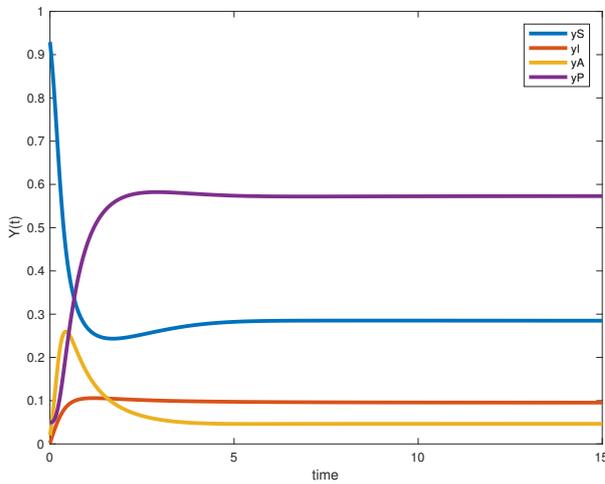


Рис. 29: Facebook1. Эксперимент №18: $g = 0.1$; начальное состояние: $t = 0, y_P = 0.0496$; стационарное состояние: $t = 14.91, y_P = 0.5729$; $t = 14.98, y_S = 0.2849$.

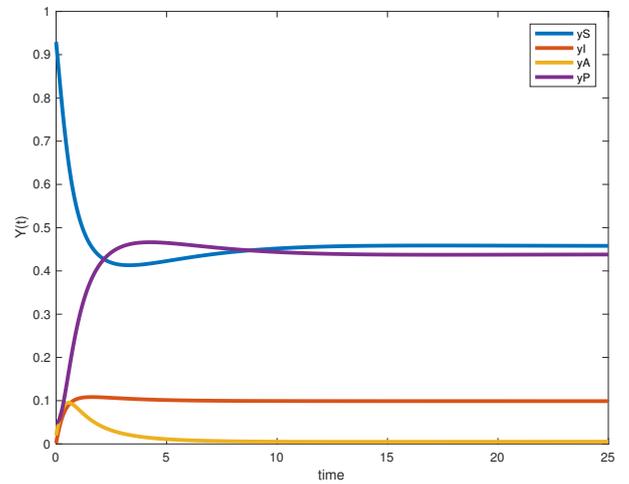


Рис. 30: Facebook1. Эксперимент №19: $g = 0.05$; начальное состояние: $t = 0, y_P = 0.0496$; стационарное состояние: $t = 23.49, y_P = 0.4379$; $t = 23.74, y_S = 0.4579$.

В данной серии экспериментов граф Facebook1 имеет 2235 узлов, однако среднее число связей равно 81 (рис. 13). Поэтому отношение D_{av} к общему числу узлов N составляет $\frac{81}{2235} = 0.0362$. Поэтому время достижения стационарных состояний в этой серии экспериментов увеличилось. В связи с этим потребовалось увеличить время моделирования до $t_f = 7$ единиц, однако остальные параметры были взяты без изменений.

Сравнивая результаты экспериментов (рис. 21 и рис. 26), можно заметить, что тенденция изменения долей агентов для каждого состояния системы сохранилась. Однако, есть ряд существенных различий.

Во-первых, отличаются финальные доли состояний системы. В экспериментах на случайном графе, например, для $g = 0.95$ они составили: S — 0.59%, I — 1.60%, A — 7.47%, P — 90.34 %. Тогда как на графе Facebook1: S — 5.42%, I — 9.39%, A — 12.91%, P — 72.28 %.

Во-вторых, в данной серии с уменьшением параметра g конечные доли состояний системы изменяются быстрее, чем в предыдущей. Например, на рис. 26 видно, что конечная доля агентов состояния P уменьшилась с 72.28% до 45.50%, тогда как во второй серии экспериментов изменения составили порядка 10%.

В-третьих, результаты эксперимента №19 (рис. 30) скорее напоминают результаты эксперимента №19 из первой серии (рис. 20), так как в момент времени $t = t_f$ большая часть агентов популяции перешла в восприимчивое состояние. Тогда как на данных из второй серии (рис. 25) видно, что больше 80% находятся в состоянии P .

При детальном рассмотрении результатов видно, что разница во второй и третьей серии экспериментов является существенной. Это связано с тем, что уравнения системы (10) стохастической модели SIAP учитывают связи каждого узла. В данном случае, узел сети Facebook1 в среднем имеет 81 связь, в то время как узел сгенерированного графа размером $N = 1000$ — 500 связей.

Таким образом, структура сети, на которой происходит процесс моделирования, сильно влияет на получаемые результаты. Поэтому при исследовании распространения информации в популяции важно учитывать не только размер сети, но и свойства графа этой сети.

3.4 Четвертая серия экспериментов

В заключительной серии экспериментов выполним сравнительный анализ результатов моделирования, проведенных на трех социальных графах: Facebook1, Facebook2, Facebook3 [23].

Моделирование распространения информации будем проводить со следующими параметрами: $t_f = 7, \beta_S = 0.2, \beta_A = 0.4, \gamma = 0, 3, \delta_1 = 0.4, \delta_2 = 0.2$. Вероятность перехода из восприимчивого состояния в предупрежденное — g будем изменять с шагом 0.05 от 0.95 до 0.05. Начальные доли состояний системы приблизительно составляют: $S - 93.00\%, A - 2.00\%, P - 5.00\%$. Доля информированных агентов в момент времени $t = t_0$ зависит от количества узлов $N_1 = 2235, N_2 = 2252, N_3 = 2920$ графов Facebook1, Facebook2, Facebook3 соответственно и вычисляется по формуле (12).

На рис. 31 представлены данные, разделенные в соответствии с графами Facebook1, Facebook2, Facebook3.

№	g	Facebook1				Facebook2				Facebook3			
		S (%)	I (%)	A (%)	P (%)	S (%)	I (%)	A (%)	P (%)	S (%)	I (%)	A (%)	P (%)
1	0,95	5,42	9,39	12,91	72,28	6,82	11,10	13,32	68,77	5,52	9,48	12,82	72,19
2	0,9	5,65	9,38	12,81	72,16	7,10	11,09	13,19	68,62	5,74	9,47	12,72	72,07
3	0,85	5,89	9,38	12,70	72,03	7,40	11,08	13,05	68,46	6,00	9,47	12,61	71,93
4	0,8	6,17	9,37	12,58	71,88	7,74	11,08	12,90	68,28	6,28	9,46	12,49	71,78
5	0,75	6,48	9,37	12,44	71,71	8,12	11,07	12,73	68,08	6,59	9,46	12,35	71,61
6	0,7	6,82	9,36	12,29	71,52	8,54	11,06	12,54	67,85	6,93	9,45	12,20	71,41
7	0,65	7,21	9,36	12,13	71,31	9,02	11,05	12,34	67,59	7,33	9,45	12,03	71,19
8	0,6	7,65	9,35	11,94	71,06	9,56	11,05	12,10	67,29	7,78	9,45	11,84	70,93
9	0,55	8,17	9,35	11,72	70,76	10,18	11,04	11,83	66,94	8,30	9,44	11,63	70,63
10	0,5	8,77	9,34	11,47	70,42	10,91	11,04	11,52	66,53	8,92	9,44	11,37	70,27
11	0,45	9,48	9,34	11,18	70,00	11,78	11,03	11,16	66,03	9,64	9,44	11,08	69,83
12	0,4	10,34	9,34	10,83	69,48	12,82	11,03	10,73	65,41	10,53	9,45	10,73	69,30
13	0,35	11,41	9,35	10,41	68,83	14,10	11,04	10,21	64,64	11,63	9,46	10,30	68,62
14	0,3	12,78	9,36	9,88	67,98	15,74	11,05	9,57	63,64	13,04	9,47	9,76	67,73
15	0,25	14,62	9,38	9,19	66,81	17,90	11,07	8,75	62,28	14,94	9,50	9,05	66,51
16	0,2	17,21	9,41	8,26	65,11	20,90	11,12	7,65	60,34	17,64	9,54	8,09	64,73
17	0,15	21,24	9,49	6,88	62,39	25,42	11,20	6,10	57,28	21,85	9,63	6,67	61,85
18	0,1	28,49	9,55	4,66	57,29	33,20	11,23	3,69	51,88	29,48	9,72	4,34	56,46
19	0,05	45,79	9,90	0,52	43,79	48,35	11,51	0,01	40,13	47,21	10,11	0,13	42,55

Рис. 31: Данные экспериментов четвертой серии для графов Facebook1, Facebook2 и Facebook3.

На рис. 32 — 35 и рис. 36 — 39 расположены графики для Facebook2 и Facebook3 соответственно. Описание каждого графика содержит название соответствующего графа, номер эксперимента, значение параметра g , а также начальные и стационарные состояния функции $y_P(t)$.

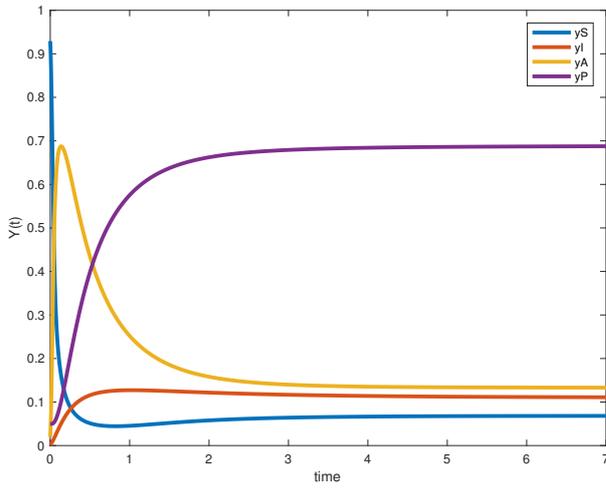


Рис. 32: Facebook2. Эксперимент №1: $g = 0.95$; начальное состояние: $t = 0, y_P = 0.0499$; стационарное состояние: $t = 6.34, y_P = 0.6877$.

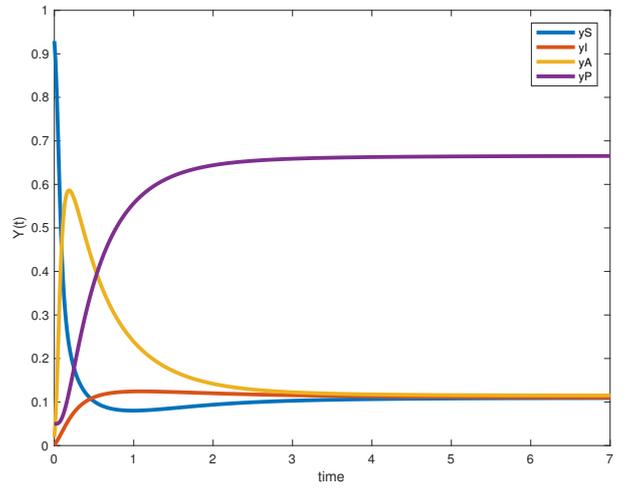


Рис. 33: Facebook2. Эксперимент №10: $g = 0.5$; начальное состояние: $t = 0, y_P = 0.0499$; стационарное состояние: $t = 6.53, y_P = 0.6653$.

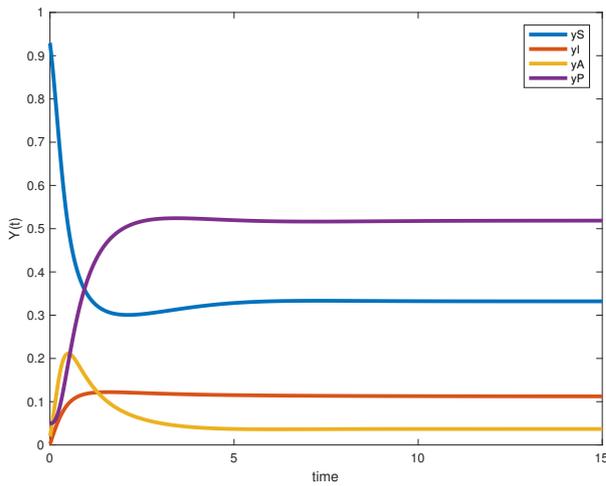


Рис. 34: Facebook2. Эксперимент №18: $g = 0.1$; начальное состояние: $t = 0, y_P = 0.0499$; стационарное состояние: $t = 14.98, y_P = 0.5188; t = 14.47, y_S = 0.3320$.

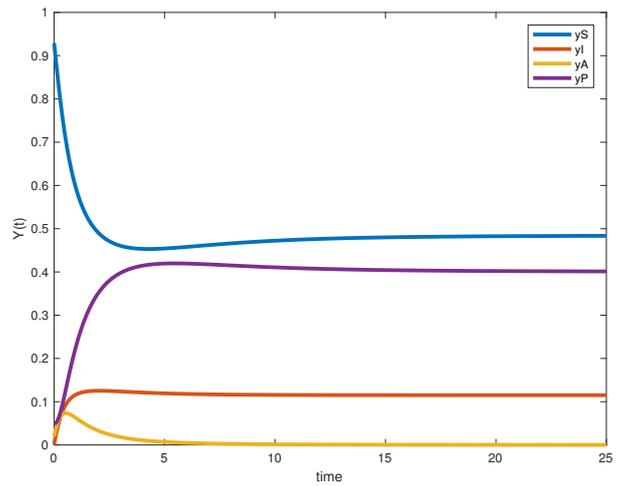


Рис. 35: Facebook2. Эксперимент №19: $g = 0.05$; начальное состояние: $t = 0, y_P = 0.0499$; стационарное состояние: $t = 23.98, y_P = 0.4013; t = 23.83, y_S = 0.4835$.

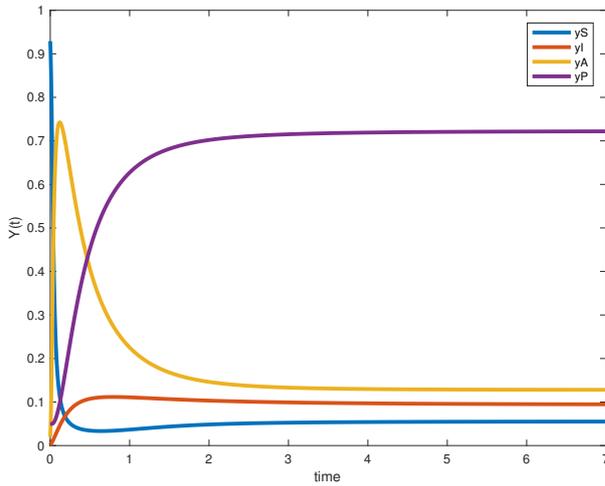


Рис. 36: Facebook3. Эксперимент №1: $g = 0.95$; начальное состояние: $t = 0, y_P = 0.0497$; стационарное состояние: $t = 6.45, y_P = 0.7219$.

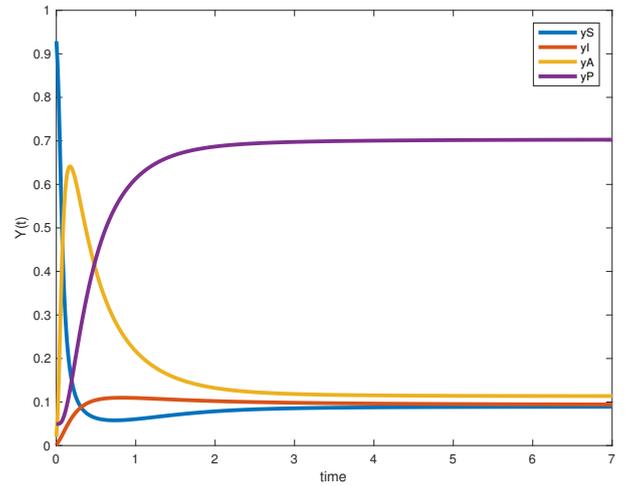


Рис. 37: Facebook3. Эксперимент №10: $g = 0.5$; начальное состояние: $t = 0, y_P = 0.0497$; стационарное состояние: $t = 6.78, y_P = 0.7027$.

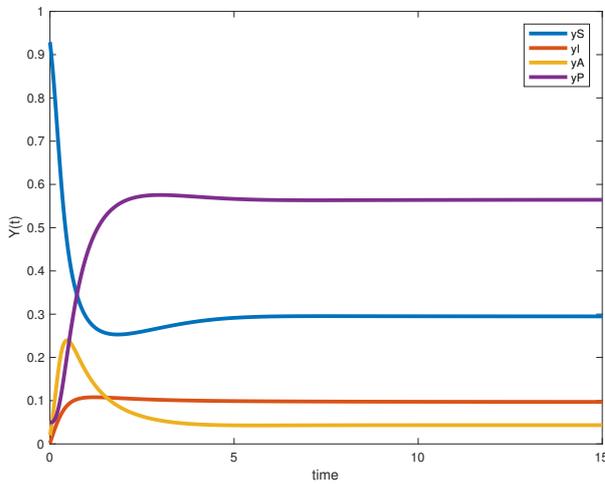


Рис. 38: Facebook3. Эксперимент №18: $g = 0.1$; начальное состояние: $t = 0, y_P = 0.0497$; стационарное состояние: $t = 14.86, y_P = 0.5646; t = 14.89 y_S = 0.2948$.

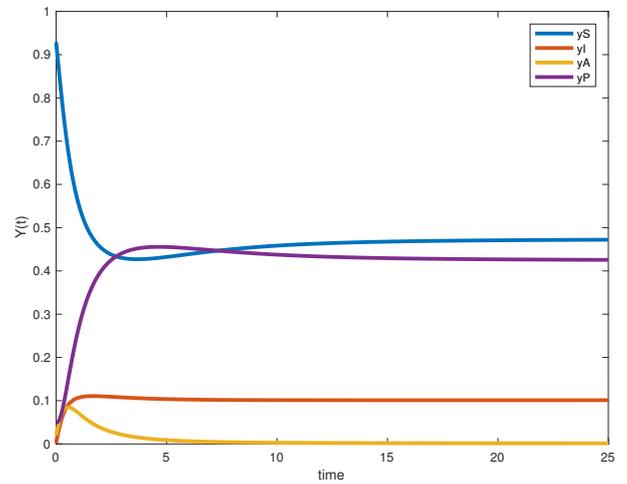


Рис. 39: Facebook3. Эксперимент №19: $g = 0.05$; начальное состояние: $t = 0, y_P = 0.0497$; стационарное состояние: $t = 23.86, y_P = 0.4255; t = 23.93 y_S = 0.4721$.

Из данных на рис. 26 и рис. 31, а также графиков третьей и четвертой серии видно, что результаты моделирования экспериментов на разных социальных графах имеют слабые отличия. Это связано с тем, что исследуемые графы Facebook1, Facebook2, и Facebook3 имеют приблизительно схожие величины характеристик (рис. 13). Поэтому описанные выше особенности динамики системы для третьей серии экспериментов также применимы и в этой серии экспериментов для графов Facebook2, и Facebook3.

Однако, эксперимент №19, изображенный на рис. 35, имеет отличие от остальных экспериментов третьей и четвертой серии, соответствующих параметру $g = 0.05$. Разница заключается в том, что на протяжении всего

моделирования значения функции $y_P(t)$ не превышают $y_S(t)$. Предполагается, что это связано с размером контактной сети Facebook2, который превосходит Facebook1 и Facebook3 на 668 и 685 узлов соответственно. А также с отношением средней степени узла $D_{av}^2 = 61$ к общему количеству узлов в системе $N_2 = 2920$. Для Facebook2 это отношение равно 0.0209 , тогда как для Facebook1 и Facebook3 0.04 и 0.03 соответственно.

3.5 Выводы

На рис. 40 приведены результаты моделирования для всех рассмотренных типов графов: случайный граф $N = 100$ — Rand100, случайный граф $N = 1000$ — Rand1000, социальный граф — Facebook1. Так как эксперименты третьей и четвертой серии показали, что результаты моделирования на социальных графах Facebook1, Facebook2 и Facebook3 не имеют существенных отличий, то был выбран граф Facebook1.

№	g	Rand100				Rand1000				Facebook1			
		S (%)	I (%)	A (%)	P (%)	S (%)	I (%)	A (%)	P (%)	S (%)	I (%)	A (%)	P (%)
1	0,95	3,39	6,45	15,96	74,20	0,59	1,60	7,47	90,34	5,42	9,39	12,91	72,28
2	0,9	3,51	6,46	15,91	74,12	0,62	1,60	7,46	90,32	5,65	9,38	12,81	72,16
3	0,85	3,65	6,46	15,86	74,03	0,65	1,60	7,45	90,30	5,89	9,38	12,70	72,03
4	0,8	3,80	6,47	15,81	73,93	0,69	1,60	7,45	90,27	6,17	9,37	12,58	71,88
5	0,75	3,97	6,47	15,75	73,81	0,73	1,60	7,43	90,24	6,48	9,37	12,44	71,71
6	0,7	4,16	6,48	15,67	73,68	0,78	1,60	7,42	90,20	6,82	9,36	12,29	71,52
7	0,65	4,39	6,49	15,59	73,53	0,83	1,60	7,41	90,16	7,21	9,36	12,13	71,31
8	0,6	4,66	6,50	15,50	73,35	0,90	1,60	7,39	90,11	7,65	9,35	11,94	71,06
9	0,55	4,97	6,51	15,38	73,13	0,97	1,61	7,37	90,05	8,17	9,35	11,72	70,76
10	0,5	5,35	6,52	15,25	72,88	1,07	1,61	7,35	89,98	8,77	9,34	11,47	70,42
11	0,45	5,82	6,54	15,08	72,56	1,18	1,61	7,32	89,89	9,48	9,34	11,18	70,00
12	0,4	6,40	6,56	14,87	72,16	1,32	1,62	7,28	89,79	10,34	9,34	10,83	69,48
13	0,35	7,15	6,59	14,61	71,65	1,50	1,62	7,23	89,65	11,41	9,35	10,41	68,83
14	0,3	8,17	6,63	14,25	70,95	1,75	1,63	7,17	89,46	12,78	9,36	9,88	67,98
15	0,25	9,60	6,68	13,75	69,96	2,10	1,64	7,08	89,19	14,62	9,38	9,19	66,81
16	0,2	11,79	6,77	13,00	68,44	2,62	1,65	6,94	88,78	17,21	9,41	8,26	65,11
17	0,15	15,53	6,91	11,74	65,82	3,52	1,68	6,72	88,08	21,24	9,49	6,88	62,39
18	0,1	23,39	7,22	9,31	60,08	5,40	1,73	6,26	86,60	28,49	9,55	4,66	57,29
19	0,05	48,61	8,13	2,35	40,92	11,76	1,90	4,86	81,47	45,79	9,90	0,52	43,79

Рис. 40: Данные экспериментов моделирования на случайном графе — Rand100, на случайном графе — Rand1000, на социальном графе — Facebook1.

Опираясь на результаты, полученные при проведении четырех серий экспериментов, сформулируем предположения о динамике состояний системы при распространении двух видов информации о предстоящей налоговой проверке на многоуровневой сети.

Во-первых, на динамику состояний системы помимо параметров модели — $\beta_S, \beta_A, \gamma, \delta_1, \delta_2, g$, также оказывают влияние характеристики графа: количество узлов — N , средняя степень узла — D_{av} , а также отношение средней степени узла к количеству узлов в сети — $\frac{D_{av}}{N}$.

Во-вторых, для любого вида контактного графа при уменьшении параметра g — вероятность перехода из состояния S в состояние P — увеличивается время достижения стационарных состояний функций средних

вероятностей состояний системы $y_S(t), y_I(t), y_A(t), y_P(t)$.

В-третьих, чувствительность динамики состояний системы повышается в социальных графах или специально сгенерированных, то есть в тех, где отношение средней степени узла к общему количеству узлов в системе крайне мало.

В-четвертых, наблюдаются нелинейные изменения динамики состояний системы при малых значениях параметра g .

Заключение

В выпускной квалификационной работе была описана стохастическая модель SAIS и изучена обобщенная модель GEMF, позволяющая моделировать распространение процессов на многоуровневой структуре сети. Эти модели были адаптированы для задачи о системе налогообложения и на их основе была создана модель SIAP, которая описывает распространение двух видов информации в обществе налогоплательщиков и позволяет исследовать влияние этой информации на решение налогоплательщиков об уплате налогов.

Для изучения введенной модели SIAP, была разработана программа и процедуры в пакете MATLAB. С их помощью были проведены четыре серии экспериментов, с целью иллюстрации и исследования динамики распространения информации в популяции модели SIAP.

Таким образом, написанная программа вместе со вспомогательными процедурами образуют инструмент для изучения динамики распространения двух видов информации в популяции, которая может быть представлена многоуровневой сетью.

Задачи, поставленные в начале работы и сформулированные в соответствии с целью, были успешно реализованы. Полученные результаты можно использовать в дальнейших исследованиях в области налогового аудита, используя данную модель SIAP и инструмент моделирования. Созданный инструмент позволяет осуществлять изучение и подбор параметров системы, а также начальных долей состояний системы для модели распространения SIAP, в соответствии с реальными данными о выплатах налогов налогоплательщиками в Российской Федерации. Таким образом, разработанные модель SIAP и инструмент моделирования, можно использовать для корректирования процесса налогового аудита с целью повышения его эффективности.

Список литературы

- [1] Kermack W. O., McKendrick A. G. A Contribution to the mathematical theory of epidemics // Proceedings of the Royal Society. 1927. P. 700—721.
- [2] Allen L. J. S. An introduction to stochastic epidemic models // Mathematical epidemiology, Springer. 2008. P. 81—130.
- [3] Mieghem P. V. Exact Markovian SIR and SIS epidemics on networks and an upper bound for the epidemic threshold // Decision and Control (CDC), IEEE 53rd Annual Conference. 2014. P. 1179—1184.
- [4] Youssef M., Scoglio C. An individual-based approach to SIR epidemics in contact networks // Journal of theoretical biology. 2011. Vol. 283, No. 1. P. 136—144.
- [5] Keeling M. J., Eames K. T. D. Networks and epidemic models // Journal of the Royal Society Interface. 2005. Vol. 2, No. 4. P. 295—307.
- [6] Sahneh F.D., Scoglio C., and Mieghem P. V. Generalized epidemic mean-field model for spreading processes over multilayer complex networks // IEEE/ACM Transactions on Networkin. 2013. Vol. 21, No. 5. P. 1609—1620.
- [7] Mieghem P. V. The N-intertwined SIS epidemic network model // Computing. 2011. Vol. 93, No. 2. P. 147—169.
- [8] Mieghem P. V., Omic J., Kooij R. Virus spread in Networks // IEEE/ACM Transactions on Networking. 2009. Vol. 17, No. 1. P. 1—14.
- [9] Gubar E., Zhu Q. Optimal control of influenza epidemic model with virus mutations // Control Conference (ECC), 2013 European. 2013. P. 3125—3130.
- [10] Devriendt K., Mieghem P. V. Unified mean-field framework for susceptible-infected-susceptible epidemics on networks, based on graph partitioning and the isoperimetric inequality // PHYSICAL REVIEW E. 2017. Vol. 96, No. 5. P. 052314-1—052314-18
- [11] Wang W., Tang M., Stanley H.E., Braunstein L. A. Unification of theoretical approaches for epidemic spreading in complex networks // Reports on Progress in Physics. 2017. Vol. 80, No.3. P. 1—16

- [12] Shane F. D., Scoglio C. Epidemic Spread in Human Networks // Decision and Control and European Control Conference (CDC -ECC), 50th IEEE Conference. 2011. P. 3008—3013.
- [13] Hill A. L., Rand D. G., Nowak M. A., Christakis N. A. Emotions as infectious diseases in a large social network: the SISa model // Proceedings of the Royal Society. 2010. Vol. 277, No. 1701. P. 3827—3837.
- [14] Hayel Ye., Trajanovski S., Altman E., Wang H., Miegheem P. V. Complete game-theoretic characterization of SIS epidemics protection strategies // Decision and Control (CDC), IEEE 53rd Annual Conference. 2014. P. 1179—1184.
- [15] Monderer D., Shapley L. S. Potential games // Games and economic behavior. 1996. Vol. 14, No. 1. P. 124—143.
- [16] Gubar E.A., Kumacheva S.Sh., Zhitkova E.M., Kurnosykh Z., Skovorodina T. Modelling of information spreading in the population of taxpayers: evolutionary approach // Contributions to Game Theory and Management. 2017. T. 10. P. 100—128.
- [17] Gubar E.A., Kumacheva S.Sh., Zhitkova E.M., Kurnosykh Z. Evolutionary behavior of taxpayers in the model of information dissemination // 2017 Constructive Nonsmooth Analysis and Related Topics (dedicated to the memory of VF Demyanov)(CNSA). 2017. P. 1—4.
- [18] Gubar E.A., Kumacheva S.Sh., Zhitkova E.M., Porokhnyavaya O. Impact of propagation information in the model of tax audit // Recent Advances in Game Theory and Applications. 2017. P.91—110
- [19] Губар Е. А., Сквородина Т. В., Моделирование эпидемического процесса с использованием модели NIMFA // Процессы управления и устойчивость. 2017. Т. 4., №. 1. С. 676—680.
- [20] Сквородина Т. В. Моделирования эпидемического процесса с учетом структуры популяции // Архив открытого доступа Санкт-Петербургского государственного университета. URL: <http://hdl.handle.net/11701/11200>
- [21] Miegheem P. V. Performance Analysis of Communications Networks and Systems. Cambridge University Press. 2006. P. 543

- [22] Meisel M., Pappas V., Zhang L. A taxonomy of biologically inspired research in computer networking // *Computer Networks*. 2010. Vol. 54, No. 6. P. 901–916.
- [23] Rossi R. A., Ahmed N. K., The Network Data Repository with Interactive Graph Analytics and Visualization // *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. URL: <http://networkrepository.com>. 2015
- [24] SIAP // GitHub.com URL: <https://github.com/githubstv/SIAP>.

Приложение

SIAPR.m — основной скрипт для численного моделирования распространения информации на многоуровневой сети, каждый уровень которой задается случайным графом.

```
clear all; close all;

fid = fopen('my_file.txt', 'w', 'native', 'UTF-8');
if fid == -1
    error('File is not opened');
end

t0=0;
tf=1;

n = 100;
partI = 5;
partA = 30;
partP = 2;
levels = 3;
states = 4;
N = states * n;
fprintf(fid, 'Nodes: %d.\n', n);
fprintf(fid, 'Levels: %d.\n', levels);
fprintf(fid, 'States: %d.\n', states);

betaS = 0.2;
betaA = 0.4;
k = 0.5;
gamma = 0.3;
delta1 = 0.4;
delta2 = 0.3;

fprintf(fid, 'betaS: %.1f, ', betaS);
fprintf(fid, 'betaA: %.1f, ', betaA);
```

```

fprintf(fid , 'k: %.1f, ', k);
fprintf(fid , 'gamma: %.1f, ', gamma);
fprintf(fid , 'delta1: %.1f, ', delta1);
fprintf(fid , 'delta2: %.1f.', delta2);
fprintf(fid , '\n');

%


---


A = generateMatrixA(n, levels);

[stA, Adelta, A11, A12, A13]=
generateMatrixStA(delta1, delta2, betaA, betaS, k, gamma);
[Q, Qdelta, Q11, Q12, Q13] =
generateMatrixQ(Adelta, A11, A12, A13, states);

QdeltaT = Qdelta';
Q11T = Q11';
Q12T = Q12';
Q13T = Q13';
Q1T= [Q11T, Q12T, Q13T];

initialV = generateInitVFixed(partI, partA, partP, n);
vec = zeros(states, 1);
statesChar = ['S', 'I', 'A', 'P'];
for i = 1 : states : N
    tmp = initialV(i : i + states - 1);
    vec = vec + tmp;
end
for i = 1:states
    fprintf(fid , '%s: %d, ', statesChar(i), vec(i));

end
fprintf(fid , '\n');
%


---



```

```
[t,y] = ode45(@(t,y) odeFunc(t, y, A, levels, Q1T, n,
states, N, QdeltaT),[t0 tf], initialV);
```

```
mS=y(:,1:states:N);
mI=y(:,2:states:N);
mA=y(:,3:states:N);
mP=y(:,4:states:N);
mSs=mean(mS');
mIs=mean(mI');
mAs=mean(mA');
mPs=mean(mP');
```

```
grid on
plot(t,mSs, 'LineWidth', 3)
hold on
plot(t,mIs, 'LineWidth', 3)
plot(t,mAs, 'LineWidth', 3)
plot(t,mPs, 'LineWidth', 3)
xlabel('time');
ylabel('Y(t)');
legend('yS', 'yI', 'yA', 'yP');
```

```
statesCharPercent = ['S', 'I', 'A', 'P'];
```

```
resultPercents = zeros(states + 1,1);
```

```
stepsInOde45 = size(y,1);
```

```
for i = 1:states
```

```
    resultPercents(i) =
```

```
    (100 * sum(y(stepsInOde45,i:states:N))) / n;
```

```
end
```

```
resultPercents(states + 1) = 100 - sum(resultPercents);
```

```

for i = 1:states

fprintf(fid , '%s: %.2f %%', ',
statesCharPercent(i), resultPercents(i));

end
fclose(fid);

```

В процедуре generateMatrixA.m генерируются матрицы смежности C_l , где $l \in \{1, 2, 3\}$.

```

function A = generateMatrixA(N, levels)

for i=1:levels
    TEMP = randi(2,N) - ones(N);
    temp= zeros(N) + triu(TEMP, 1) + triu(TEMP,1)';
    A(1:N,((i-1)*N+1):i*N)=temp;
end
end

```

В процедуре generateMatrixStA.m задаются матрицы смежности A^δ и A_l ($l \in \{1, 2, 3\}$), соответствующие графам узловых и сетевых переходных вероятностей соответственно.

```

function [stA, Adelta, A11, A12,A13] =
    generateMatrixStA(delta1, delta2, betaA, betaS, k,
    gamma)

Adelta=[0 0 0 0; delta1 0 0 0; 0 0 0 0; delta2 0 0 0];
A11=[0 0 k 0; 0 0 0 0; 0 0 0 0; 0 0 0 0];
A12 = [0 betaS 0 0; 0 0 0 0; 0 betaA 0 0; 0 0 0 0];
A13 = [0 0 0 0; 0 0 0 gamma; 0 0 0 gamma; 0 0 0 0];

stA=[Adelta, A11, A12,A13];
end

```

В соответствии с заданными в процедуре generateMatrixStA.m матрицами смежности A^δ и A_l , задаются матрицы Кирхгофа Q^δ и Q_l ($l \in \{1, 2, 3\}$) в процедуре generateMatrixQ.m.

```
function [Q, Qdelta, Q11, Q12, Q13] =
generateMatrixQ(Adelta, A11, A12, A13, states)

Qdelta = zeros(states);
Q11 = zeros(states);
Q12 = zeros(states);
Q13 = zeros(states);

sAdelta=sum(Adelta,2);
s11=sum(A11,2);
s12=sum(A12,2);
s13=sum(A13,2);

for i=1:states
    for j=1:states
        if i==j
            Qdelta(i,j) = sAdelta(i)-Adelta(i,j);
            Q11(i,j) = s11(i)-A11(i,j);
            Q12(i,j) = s12(i)-A12(i,j);
            Q13(i,j) = s13(i)-A13(i,j);
        else
            Qdelta(i,j) = -Adelta(i,j);
            Q11(i,j) = -A11(i,j);
            Q12(i,j) = -A12(i,j);
            Q13(i,j) = -A13(i,j);
        end
    end
end

Q = [Qdelta, Q11, Q12, Q13];
```

В процедуре `generateInitVFixed` задается вектор начальных долей состояний системы $S, I, A, P - Y(0)$.

```
function answer = generateInitVFixed(I,A,P,n)

answer = zeros(n*4,1);

for i = 1 : I
    cell = zeros(4,1);
    cell(2) = 1;
    answer((i-1) * 4 + 1: i * 4) = cell;
end

for i = I + 1 : I + A
    cell = zeros(4,1);
    cell(3) = 1;
    answer((i-1) * 4 + 1: i * 4) = cell;
end

for i = I + A + 1 : I + A + P
    cell = zeros(4,1);
    cell(4) = 1;
    answer((i-1) * 4 + 1: i * 4) = cell;
end

for i = I + A + P + 1 : n
    cell = zeros(4,1);
    cell(1) = 1;
    answer((i-1) * 4 + 1: i * 4) = cell;
end
end
```

В процедуре odeFunc.m осуществляется поиск решения системы дифференциальных уравнений модели SIAP (10).

```
function dydt = odeFunc(t, y, A, levels, Q1T, n,
states, N, QdeltaT)

dydt = zeros(N,1);

for i = 1 : states : N

    dydt(i:i + states - 1) = [
        ( -QdeltaT - getSumsFori(y, fix(i/states)
+ 1, A, levels, Q1T, n, states)) *
        y(i : i + states - 1) ];
end
end
```

Следующие три процедуры getSumsFori.m, getV.m и getq.m используются в odeFunc.m.

```
getSumsFori.m

function finalSum =
getSumsFori(y,i, A, levels, Q1T, n, states)

finalSum = zeros(states);
for k = 1 : levels
    currentLevel = A(i, (k-1)*n + 1:k*n);
    currentV = getV(y,k, states, n);
    finalSum = finalSum +
    (currentLevel * currentV) * Q1T(:, ((k-1)*states +
1 : k * states));
end
end
```

getV.m

```
function newV = getV(y,level , states , n)

newV = zeros(n,1);

for j = 1 : n
    newV(j) = y((j-1) * states + getq(level));
end
end
```

getq.m

```
function q = getq(level)

if (level == 1 )
    q = 3;
else
    q=2;
end
end
```