

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

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

Кафедра теории функций и стохастического анализа

**АЛГОРИТМЫ АНАЛИЗА РАСПРЕДЕЛЕНИЯ СТЕПЕНЕЙ В
СЕТЕВЫХ СТРУКТУРАХ**

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

Студентки 2 курса 248 группы

направления 01.03.02 — Прикладная математика и информатика

механико-математического факультета

Тряпкиной Татьяны Сергеевны

Научный руководитель

доцент, к. ф.-м. н.

С. П. Сидоров

Заведующий кафедрой

д. ф.-м. н., доцент

С. П. Сидоров

Саратов 2019

ВВЕДЕНИЕ

Современные фондовые рынки генерируют огромное количество информации, анализ которой невозможен без применения технологий моделирования. Одной из успешно применяющихся моделей фондового рынка является представление рынка в виде графа, вершинами которого являются акции, торгующиеся на бирже, а вес ребра определяет степень зависимости двух акций друг от друга. Совокупность характеристик графа, представляющего фондовый рынок, может помочь инвестору и другим акторам оценить состояние рынка.

Цель магистерской работы заключается в изучении методов анализа сетевых структур, а также исследование структуры и топологических свойств фондовой сети, построенной по данным фондового рынка России за период с 2012 по 2019 гг. Также в рамках работы был разработан программный продукт на языке R, позволяющий провести анализ топологических свойств сети фондового рынка.

Достижение поставленной цели требует решения следующих **задач**:

1. Сбор и подготовка данных для проведения анализа
2. Визуализация данных – построение фондовой сети на основе данных ММВБ
3. Исследование топологических свойств построенной сети
4. Анализ полученных результатов

Данная работа состоит из трех разделов:

- Сетевые структуры в современном мире,
- Теория графов в изучении сетевых структур,
- Анализ топологических свойств сети, построенной на основе данных ММВБ.

Работа прошла апробацию на различных конференциях, в частности,

- в XIX Международной Саратовской зимней школе «Современные проблемы теории функций и их приложения», посвященной 90-летию со дня рождения академика П. Л. Ульянова, январь 2018 года
- на ежегодной студенческой конференции "Актуальные проблемы математики и механики", которую проводил механико-математический факультет СГУ в апреле 2019 года, в секции "Анализ данных",
- в VII Международной молодежной научно-практической конференции «Математическое и компьютерное моделирование в экономике, страховании и управлении рисками», ноябрь 2018 года.

Результаты работы опубликованы в статье Analysis for Correlation Dependence Network of Financial Assets, издательство Atlantis Press.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел включает в себя описание сетевых структур в современном мире, а именно – их виды, сетевой подход в изучении, а также анализ фондовых рынков с помощью сетевого подхода.

Сетевые структуры (или просто сети) по-разному понимаются в литературе. Многие исследователи интерпретируют сети предельно широко: сеть как любая система из узлов, соединенных связями (ребрами). На базе данной интерпретации разработан математический понятийный аппарат, в рамках которого созданы концептуальные критерии центральности узлов в сети, равномерного или неравномерного распределения связей между ними (коэффициент кластеризации) и др. [2]. В противовес кратко обозначенному широкому пониманию сетевых структур существует более узкое истолкование сетей, также имеющее немало влиятельных сторонников в мире, особенно среди представителей социальных и гуманитарных наук [4].

В более узкой интерпретации сеть есть структура, где отсутствует единый центр (лидер, доминант), так что ее поведение является результатом кооперативных взаимодействий между элементами (узлами), среди которых могут быть несколько частичных лидеров с ограниченным воздействием на систему.

Термин «сложные сети» возник в начале этого века и относится к сетям с более сложной архитектурой, чем, скажем, классические случайные сети с заданным числом узлов и связей или решетки в кристаллах. Обычно в таких сетях имеется небольшое число узлов с большим числом связей – т. н. хабы (от английского hub – ядро, концентратор), которые в значительной степени и определяют свойства этих сетей. При этом оказалась, что большинство реальных сетей (биологических, технических, социальных) являются сложными.

В последние годы активно развивающимся направлением в изучении сложных систем является сетевой анализ. Каждый объект системы может быть представлен узлом сети (графом), а взаимодействия между этими объектами – соответствующими ребрами.

Считается, что активный интерес к исследованиям в области сложных сетей появился благодаря работам, посвященным анализу сетей «тесного мира» Уоттса-Строгатца [32] и «безмасштабных сетей» Барабаши-Альберт [2].

Впервые сетевой подход для анализа рынка ценных бумаг применил Р. Мантегна [14, 15]. Он предложил использовать матрицу корреляций для вычисления «расстояний» между парами акций, а затем выбирать такие из них, которые удовлетворяют критерию минимального остовного дерева. В получившемся дереве «расстояния» между акциями представлены ребрами, а узлы – ценными бумагами. Впоследствии многими авторами были исследованы похожие и изобретены новые сетевые модели фондового рынка.

Во втором разделе рассматриваются теоретические аспекты теории графов, применяемые при сетевом анализе фондовых рынков.

При построении сетевой модели, как и в большинстве других моделей, использующих матрицу попарных корреляций, в качестве меры близости доходностей акций используются коэффициенты корреляции Пирсона.

Для вычисления корреляций между всеми парами акций используется временной ряд цен $P_i(t)$ для каждой компании s_i в момент времени t . Для сглаживания колебаний используется переход к лог-доходностям $Y_i(t)$ компаний s_i в период времени $[t-\Delta t, t]$, определяемый следующей формулой:

$$Y_i(t) = \ln P_i(t) - \ln P_i(t - \Delta t), \quad (1)$$

где $\Delta t = 1$ для цен каждого дня. Затем был найден коэффициент корреляции Пирсона для каждой пары компаний s_i и s_j по следующей формуле:

$$p_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2) - (\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}}, \quad (2)$$

где $\langle \cdot \rangle$ означает среднее значение.

Также рассматриваются следующие топологические свойства сетей:

- Распределение степеней,
- Кластеризация,
- Центральность вершины,
- Средняя длина пути,
- Минимальное остовное дерево.

В третьем разделе изученные теоретические аспекты применяются для анализа сети, построенной на основе данных индекса ММВБ.

В течение последних десяти лет российская экономика пережила ряд «шоков», таких как резкое падение цен на нефть в 2014 году, а также санкции, введенные правительствами стран Европейского союза и США. Оба «шока» повлекли за собой широкомасштабные и продолжительные негативные последствия, а также привели к замедлению роста российской экономики. Эти негативные события незамедлительно отразились на состоянии российского фондового рынка. Таким образом, изучений свойств российского фондового рынка посредством обработки и представления данных в цифровой форме, а также применения сетевых подходов представляет собой особый интерес. В данном разделе работы исследуется влияние двух «шоков» на динамику изменения топологических свойств сети, построенной на основе данных российского фондового рынка за последнее десятилетие.

Россия является страной с развивающейся экономикой и занимает шестое место среди всех стран мира по объему ВВП. Индекс фондовой биржи РТС и индекс Московской фондовой биржи являются основными индикаторами российского фондового рынка. Разница между ними заключается в том, что РТС рассчитывается в долларах, а ММВБ - в рублях. Для проведения анализа были собраны данные о ценах для каждой за каждый день за период с 10.01.2012 по 09.04.2019 (1664 торговых дня). Данные были подготовлены по сведениям из открытых источников на Yahoo Finance. Аналогичным образом данные по индексу ММВБ были взяты с Московской фондовой биржи за тот же период. Всего рассматривается 194 компании, среди которых 32 компании входят в индекс ММВБ.

Были построены графики цен закрытия и лог-доходностей по данным за 1664 дня в период с 10 января 2012 по 09 апреля 2019.

Некоторые компании (CBOM, DSKY, FIVE, LNTA, MOEX, POLY, RNFT, RUAL, SFIN), которые принимают участие в торгах на ММВБ, не

были включены в набор данных. Это связано с тем, что они имеют очень короткую историю торговли (всего 2-3 года), что недостаточно для построения соответствующей модели. Так же их анализа были исключены наборы данных компании, у которых было более 250 пропущенных торговых дней подряд. Таким образом, только 194 эмитента остались в наборе данных из первоначального списка 278 эмитентов, торгуемых на Московской фондовой бирже. Компания YNDX начала торговаться на Московской фондовой бирже с 02.06.2014, но она торгуется на NASDAQ с 02.05.2011, откуда были взяты недостающие данные. После очистки и процесса отбора в набор данных для анализа были включены 194 акции, торгуемые на Московской бирже. Исследуемые котировки акций относятся к разным областям экономики. Как показано в таблице 1, была проведена их классификация по сферам деятельности среди десяти отрасли промышленности.

Таблица 1 – Разбиение 194 акций по секторам экономической деятельности

Сектор	Количество акций
Сырье	36
Потребительские товары (циклического характера)	11
Потребительские товары (нециклического характера)	9
Энергетический	22
Финансовый	16
Здравоохранение	3
Промышленный	24
Технологии	3
Телекоммуникации	11
Коммунальные услуги	59

Для вычисления корреляций между всеми парами акций используется

временной ряд цен $P_i(t)$ для каждой компаний s_i в момент времени t . Для сглаживания колебаний используется переход к лог-доходностям $Y_i(t)$ компаний s_i в период времени $[t-\Delta t, t]$, определяемый следующей формулой:

$$Y_i(t) = \ln P_i(t) - \ln P_i(t - \Delta t), \quad (3)$$

где $\Delta t=1$ для цен каждого дня. Затем был найден коэффициент корреляции Пирсона для каждой пары компаний s_i и s_j по следующей формуле:

$$p_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2) - (\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}}, \quad (4)$$

где $\langle \cdot \rangle$ означает среднее значение.

Значение коэффициента p_{ij} варьируется в промежутке от -1 до 1. Отрицательное значение $p_{ij} < 0$ означает, для данной пары акций при падении стоимости первой стоимость второй – повышается. А при положительном значении коэффициента $p_{ij} > 0$ обе акции испытывают изменения одинакового характера. Если $p_{ij} \approx 0$, то пара акций не коррелируема, а в случае $|p_{ij}| \approx 1$ пара акций подвержена практически полностью идентичным изменения. Данное явление часто свойственно акциям компаний, принадлежащим одному экономическому сектору.

Основываясь на значениях коэффициентов корреляции Пирсона была построена матрица расстояний $N \times N = 194 \times 194$ по формуле:

$$d_{ij} = \sqrt{2(1 - p_{ij})} \quad (5)$$

по данным индексов Московской биржи была построена сеть с 194 вершинами. Используя формулу (5), были построены матрицы расстояний для каждого окна. Следует отметить, что все отрицательные значения коэффициентов корреляции были преобразованы в положительные

расстояния. Все вершины в сети фиксированы и представляют собой акции. Однако ребра меняются в каждом скользящем окне по мере изменения расстояний. Исходя из этого, были усреднены все матрицы расстояний по всем скользящим окнам за изучаемый период. Ниже рассматриваются статистические свойства фондовой сети ММВБ.

Максимальное количество ребер в усредненной матрице расстояния равно $N(N-1)/2 = 18721$ и это довольно большое количество. Для упрощения сети можно применить фильтрацию ребер с небольшими весами. В этом случае график рынка строится на основе порогового подхода: для фиксированного порогового значения удаляются те ребра, вес которых больше, и сохраняются оставшиеся ребра. Другими словами, фильтрация проводится согласно следующему правилу:

$$e_{ij} = \begin{cases} 1, & d_{ij} < \theta, \\ 0, & d_{ij} \geq \theta. \end{cases} \quad (6)$$

В таблице 2 представлены основные свойства сети, построенной на основе индексов ММВБ для различных значений $0.5 \leq \theta \leq 1.5$.

Таблица 2 – Основные свойства сети для различных значений θ с шагом 0,1.

θ	$ e _{max}$	$ e $	$ e _{density}$	$\langle d \rangle$	$\langle d_{ij} \rangle$	d_{ij}^{max}	σ
0.5	18721	0	0.0000	0.0000	0.0000	0.0000	0.0000
0.6	18721	1	0.0001	0.0103	0.5458	0.5458	0.0000
0.7	18721	2	0.0001	0.0206	0.5848	0.6238	0.0552
0.8	18721	2	0.0001	0.0206	0.5848	0.6238	0.0552
0.9	18721	4	0.0002	0.0412	0.7201	0.8969	0.1631
1	18721	17	0.0009	0.1753	0.8957	0.9994	0.1263
1.1	18721	80	0.0043	0.8247	1.0235	1.0992	0.0915
1.2	18721	376	0.0201	3.8763	1.1351	1.1999	0.0756
1.3	18721	1996	0.1066	20.5773	1.2393	1.2998	0.0648

1.4	18721	14260	0.7617	147.0103	1.3477	1.4000	0.0551
1.5	18721	18721	1.0000	193.0000	1.3638	1.4748	0.0563

Максимально возможное количество вершин – 18721. Можно видеть, что число ребер равно нулю, если пороговое значение $\theta = 0,8$, а график становится полным при $\theta = 1,5$.

В процессе анализа было проведено исследование распределения степеней для сетей ММВБ, построенных с различными значениями θ . Распределения степеней являются шумными как для малых значений θ , так и для больших. Для достаточно малого интервала $\theta = (1.15; 1.2)$ распределения степеней близки к степенному закону.

ЗАКЛЮЧЕНИЕ

В данной работе были использованы некоторые методы теории сложных сетей для изучения сети российского финансового рынка. Комплексный анализ сети позволяет исследовать структуру сети российского финансового рынка, основанную на корреляционных отношениях между компонентами системы.

В ходе выполнения работы были решены все поставленные задачи, такие как сбор и подготовка данных для проведения анализа, визуализация данных, исследование топологических свойств построенной сети, а также анализ полученных результатов. Таким образом, была выполнена основная цель магистерской работы.