Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра математической
кибернетики и компьютерных наук

РАСЧЕТ ХАРАКТЕРИСТИК ГРАФОВ КОРРЕЛЯЦИОННЫХ ВЗАИМОСВЯЗЕЙ ФИНАНСОВЫХ АКТИВОВ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 451 группы
направления 09.03.04 — Программная инженерия
факультета КНиИТ
Глазова Романа Владимировича

Научный руководитель
к. ф.-м. н. _____________ С. В. Миронов

Заведующий кафедрой
к. ф.-м. н. _____________ С. В. Миронов

Саратов 2018

ВВЕДЕНИЕ

Одной из наиболее важных проблем в современных финансах является поиск эффективных способов обобщения и визуализации данных фондового рынка, которые позволят получить полезную информацию о поведении рынка. В настоящее время на фондовых рынках торгуется большое количество акций; кроме того, это число неуклонно увеличивается. Объем данных, генерируемых на фондовом рынке каждый день, огромен. Эта данные обычно визуализируются тысячами графиков, отражающих цену каждого запаса в течение определенного периода времени. Анализ этих участков становится все более сложным по мере роста количества акций.

Альтернативный способ обобщения данных о ценах на акции, основан на представлении фондового рынка в виде неориентированного невзвешенного графа. Этот граф называется графом корреляционных взаимосвязей финансовых активов. Следует отметить, что подход представления набора данных в виде графа становится все более и более используемым в различных практических приложениях, одним из которых являются финансы. Эта методология позволяет визуализировать набор данных, представляя его элементы в виде вершин и наблюдать определенные соотношения между ними. Изучение структуры графа, представляющего набор данных, важно для понимания внутренних свойств рынка, которые он представляет, а также для улучшения организации хранения и поиска информации.

Основной целью настоящей работы является разработка приложения для выявления динамики изменений структурных свойств графа корреляционной взаимосвязи финансовых активов с течением времени. В работе рассматриваются графы, построенные на основе данных о ценах на акции для различных периодов времени в течение 2011-2015 гг. и изучение эволюции некоторых характеристик этих графов.

Итак, в рамках выпускной квалификационной работы поставлены следующие задачи:

- разработать программу для преобразования исходных данных о динамике цен финансовых активов в граф корреляционных взаимосвязей;
- реализовать методы нахождения различных характеристик графов, таких как: плотность, размер максимальной клики, размер максимального независимого множества;

— протестировать реализованные алгоритмы на реальных данных;
— визуализировать и проанализировать полученные данные.

1 Постановка проблемы

Проблема, решаемая в данной работе формулируется следующим образом: базу данных, хранящую изменение котировок финансовых активов за определенный промежуток времени необходимо преобразовать в граф корреляционных взаимосвязей, показывающий «схожесть» финансовых активов. Граф корреляционной взаимосвязи финансовых активов имеет следующую структуру: каждому финансвовому активу соответствует ровно одна вершина графа, а пара вершин смежна, если значение критерия корреляции Пирсона между финансовыми представляющими эти вершины больше заведомо определенного значения θ.

Полученный граф необходимо проанализировать, измерив его характеристики, такие как: плотность, размер максимальной клики, размер максимального независимого множества, распределение степеней вершин.

1.1 Граф

Формально граф представляет собой пару множеств (V, E), где V — множество вершин, а E — множество ребер, образованное парами вершин. В данной работе речь пойдет только о неориентированных графах, поэтому пары вершин в множестве E являются неупорядоченными. [1–3]

1.2 Коэффициент корреляции Пирсона

Коэффициент корреляции Пирсона – коэффициент параметрической статистики, позволяющий определить наличие или отсутствие линейной связи между двумя количественными показателями, а также оценить ее тесноту и статистическую значимость [4–6]. Другими словами, коэффициент корреляции Пирсона позволяет определить, есть ли линейная связь между изменениями значений двух переменных X и Y. В статистических расчетах и выводах коэффициент корреляции обычно обозначается как C_{xy}.

В общем виде формула для подсчета коэффициента корреляции такова:

$$C_{xy} = \frac{\sum\limits_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2 \sum\limits_{i=1}^{n}(y_i - \overline{y})^2}} \qquad (1)$$

где

x_i – значения переменной X,

y_i – значения переменной Y

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Также предполагается, что переменные X и Y распеределены нормально [7].

Недостатком данной формулы является то, что из каждого значения x_i, нужно вычитать среднее значение переменной X. Поэтому равносильными преобразованиями приводим формулу к следующему виду:

$$C_{xy} = \frac{\sum\limits_{i=1}^{n}(x_i y_i) - \frac{1}{n}(\sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i)}{\sqrt{S_x S_y}} \tag{2}$$

где

$$S_x = \sum_{i=1}^{n} x_i^2 - \frac{(\sum\limits_{i=1}^{n} x_i)^2}{n}$$

$$S_y = \sum_{i=1}^{n} y_i^2 - \frac{(\sum\limits_{i=1}^{n} y_i)^2}{n}$$

или модификацию этой формулы:

$$C_{xy} = \frac{n\sum\limits_{i=1}^{n}(x_i y_i) - (\sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i)}{\sqrt{(n\sum\limits_{i=1}^{n} x_i^2 - (\sum\limits_{i=1}^{n} x_i)^2)(n\sum\limits_{i=1}^{n} y_i^2 - (\sum\limits_{i=1}^{n} y_i)^2)}} \tag{3}$$

C_{xy} находится в отрезке $[-1, 1]$. Положительный знак говорит о наличии прямой связи между X и Y, отрицательный – о наличии обратной связи [8, 9].

1.3 Создание графа

Естественное графическое представление фондового рынка основано на кросс-корреляциях ценовых коллебаний финансовых активов. Граф корреляционной взаимосвязи финансовых активов имеет следующий вид: каждому финансовому активу соответствует вершина, а две вершины связанный ребром, если коэффициент корреляции между ними (рассчитанный за определенный период времени) превышает заданный порог $\theta \in [-1, 1]$.

Формальная процедура построения графа корреляционной взаимосвязи финансовых активов следующая. Обозначим через $P_i(t)$ цену инструмента i в день t. Тогда

$$R_i(t) = \ln \frac{P_i(t)}{P_i(t-1)} \tag{4}$$

определяет натуральный логарифм отношения стоимости инструмента i в день t к стоимости в день $t-1$. Пусть

$$C_{ij} = PCC(R_i(1), R_i(2), ..., R_i(k), R_j(1), R_j(2), ..., R_j(k)) \tag{5}$$

где PCC – коэффициент корреляции Пирсона.

Ребро, соединяющее вершины i и j добавляется в граф, если $C_{ij} \geq \theta$, что означает, что цены на эти два инструмента ведут себя одинаково с течением времени, и степень этого сходства определяется выбранным значением θ.

1.4 Реберная плотность

Реберная плотность простого неориентированного графа G – отношение количества ребер графа к максимально возможному количеству ребер в нем [10–12].

$$D = \frac{2|E|}{|V|(|V|-1)} \tag{6}$$

где

V — количество вершин графа,

E — количество ребер графа.

Реберная плотность является важной характеристикой графа корреляционной взаимосвязи финансовых активов. Увеличение реберной плотности говорит о некой «глобализации» фондового рынка, т.е. о том, что все больше и больше финансовых активов существенно влияют на друг на друга и изменение котировок одного финансового актива влечет за собой изменение котировок других финансовых активов.

1.5 Размер максимальной клики

Данный раздел содержит необходимые теоретические сведения о кликах в неориентированных графах.

1.5.1 Теоретические сведения

Задача о поиске клики максимального размера относится к классу NP-полных задач в области теории графов [13, 14].

Кликой в неориентированном графе называется подмножество вершин такое, что любые две вершины из этого подмножества соединены ребром [15, 16].

Поскольку клика представляет собой набор полностью взаимосвязанных вершин, любой финансовый актив, принадлежащий клике, сильно коррелирует со всеми остальными финансовыми активами в этой клике. Поэтому финансовый актив привязывается к определенной клике только если он демонстрирует поведение, подобное всем остальным финансовым активам в этой группе. Очевидно, что размер максимальной клики является важной характеристикой фондового рынка, поскольку она представляет собой максимально возможную группу аналогичных объектов (т. е. взаимно коррелирующих финансовых активов).

1.6 Распределение степеней

Согласно ранее проведенным исследованиям, для значения порога корреляции $\theta \geq 0.2$, распределение степеней вершин хорошо аппроксимируется степенной моделью. Согласно этой модели вероятность того, что вершина x имеет степень k (т.е. k исходящих ребер) равна

$$P(x) \propto k^{-\gamma} \tag{7}$$

или

$$logP(x) \propto -\gamma logk \tag{8}$$

1.7 Максимальное независимое множество вершин

Данный раздел содержит необходимые теоретические сведения о максимальных независимых множествах вершин в неориентированных графах.

1.7.1 Теоретические сведения

Задача о поиске максимального независимого множества не разрешима за полиномиальное время.

Независимым множеством в неориентированном графе называется подмножество вершин такое, что любые две вершины из этого подмножества не связаны ребром [17].

1.7.2 Алгоритм Брона — Кербоша для поиска минималного независимого множества вершин

Для поиска минималного независимого множества вершин необходимо слегка модифицировать алгоритм Брона-Кербоша для поиска клик. Модифицированная версия алгоритма рассмотрена ниже.

В ходе работы алгоритма поддерживаются три множества вершин (R, P, X), такие что:

1. R — множество вершин, содержащее на каждом шаге текущее максимальное независимое множество вершин.
2. P — множество вершин, которые могут быть добавлены к множеству R на текущем шаге.
3. X — множество вершин, которые уже использовались для расширения множества R на предыдущих шагах работы алгоритма.

На каждом шаге работы алгоритма поддерживается следующий инвариант: множества P и X не пересекаются, т.е. $P \cap X = \varnothing$. Так же, объединение этих множеств состоит из вершин, которые образуют МНМВ при добавлении в R. Таким образом, любая пара вершин (u, v), такая что $u \in R, v \in P \cup X$ не связана ребром. Когда множества P и X одновременно пусты, нет никаких дополнительных вершин, которые могут быть добавлены в R, а значит R является максимальным независимым множеством вершин, и алгоритм завершается, выдавая в результате своей работы множество R.

В начале работы рекурсивного алгоритма множества R и X устанавливаются пустыми, а множество P содержит все вершины графа G. Внутри каждого рекурсивного вызова вершины в P рассматриваются по очереди. Далее рассматриваются два случая. Если множество X пусто, то алгоритм завершается и возвращает множество R. Иначе алгоритм возвращается назад. Для каждой вершины $v \in P$ алгоритм делает рекурсивный вызов, в котором вершина v добавляется в множество R, а из множеств P и X удаляется, если она там есть. Если это рекурсивынй вызов не нашел независимое множество вершин макисмального размера, вершина v удаляется из множества P и добавляется к множеству X, чтобы избежать её повторного рассмотрения в следующих рекурсивных вызовах.

Стоит отметить, так на ГКВФА алгоритм Брона-Кербоша для поиска МНМВ не завершается за разумное время(из-за их малой плотности), количество рекурсивных вызовов алгоритма ограничено константой `MAXCALL`. В данной работе константе `MAXCALL` присвоено значение 100000. При увеличении константы `MAXCALL` точность алгоритма увеличивается, но в то же время увеличивается и время его работы. При уменьшении константы уменьшается как точность алгоритма, так и его время работы.

Тестирование на реальных ГКВФА показало, что для достижения лучшего результата работы алгоритма все отведенное на него время нужно потратить на один вызов рекурсивной функции поиска.

1.8 Распределение коэффициентов корреляции

Согласно ранее проведенным исследованиям, коэффициенты корреляции распределны нормально.

2 Тестирование

База данных для построения и анализа ГКВФА была взята с ресурса [18].

Для исследования динамики графа корреляционной взаимосвязи финансовых активов, был выбран 1000-дневний промежуток торговых дней в 2011-2015 годах, который, в свою очередь, поделен на 11 последовательных 500-дневных периодов. Каждый период, кроме первого получается сдвигом предыдущего на 50 дней. Таким образом, два соседних периода имеют 450 общих дней. Даты, соответствующие каждому периоду, представлены в таблице.

Период	Дата начала	Дата завершения
1	01.02.2011	19.03.2013
2	20.03.2013	01.06.2013
3	02.06.2013	12.08.2013
4	13.08.2013	26.10.2013
5	27.10.2013	08.01.2014
6	09.01.2014	22.03.2014
7	23.03.2014	04.06.2014
8	05.06.2014	18.08.2014
9	19.08.2014	30.10.2014
10	31.10.2014	11.01.2015
11	12.01.2015	24.03.2015

2.1 Распределение коэффициентов корреляции

Распределение коэффициентов корреляции симметрично и имеет форму, похожую на нормальное с математическим ожиданием порядка 0.25 [19, 20]. Положительное математическое ожидание означает, что в среднем финансовые активы коррелируют в незначительной степени.

2.2 Размер максимальной клики

Размер максимальной клики увеличился в полтора раза за время измерений. Этот факт говорит о том, что все большее количество финансовых активов демонстрируют поведение, подобное другим финансовым активам.

2.3 Реберная плотность

Реберная плотность увеличилась от первого измерения к одиннадцатому на сорок процентов. Увеличение реберной плотности говорит о «глобализации» фондового рынка, т.е. о том, что все больше и больше финансовых активов существенно влияют на друг на друга и изменение котировок одного финансового актива влечет за собой изменение котировок других финансовых активов, а структура рынка образуется не чисто случайным образом.

2.4 Распределение степеней вершин

Подтвердилось наблюдение о характере распределения степеней вершин, сделанное ранее. А именно то, что распределение степеней вершин аппроксимируется степенной моделью.

ЗАКЛЮЧЕНИЕ

В рамках выполнения выпускной квалификационной работы была сформулирована проблема обобщения и визуализации данных фондового рынка при помощи графа коррелляционной взаимосвязи финансовых активов.

Была разработана программа для преобразования исходных данных о динамике цен финансовых активов в граф корреляционных взаимосвязей.

Были реализованы методы нахождения различных характеристик графов, таких как: плотность, размер максимальной клики, размер максимального независимого множества.

Реализованные алгоритмы были протестированы на реальных данных.

Полученные данные были визуализированы и проанализированы.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 *Ruohonen, K.* Graph Theory / K. Ruohonen. — 2013.

2 *Trudeau, J.* Introduction to Graph Theory / J. Trudeau. — 1993.

3 *Saoub, R.* A Tour Through Graph Theory / R. Saoub. — 2004.

4 *Jones, K.* Text Book Of Correlations And Regression / K. Jones, M. Beazley. — Discovery Publishing House, 2005.

5 Correlation Coefficient: Simple Definition [Электронный ресурс]. — URL: http://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/ (Дата обращения 20.05.2018). Загл. с экр. Яз. англ.

6 Pearson Correlations – Quick Introduction [Электронный ресурс]. — URL: https://www.spss-tutorials.com/pearson-correlation-coefficient/ (Дата обращения 20.05.2018). Загл. с экр. Яз. англ.

7 *Read, B.* Handbook of the Normal Distribution / B. Read, K.Patel. — Discovery Publishing House, 2008.

8 Pearson's Correlation Coefficient [Электронный ресурс]. — URL: http://www.statisticssolutions.com/pearsons-correlation-coefficient/ (Дата обращения 20.05.2018). Загл. с экр. Яз. англ.

9 Pearson's Correlation [Электронный ресурс]. — URL: http://academic.sun.ac.za/emergencymedicine/TRRM/module10/BS1-11.htm (Дата обращения 20.05.2018). Загл. с экр. Яз. англ.

10 *Diestel, R.* Graph theory / R. Diestel. — Springer, 2006.

11 An edge density definition of overlapping and weighted graph communities [Электронный ресурс]. — URL: https://arxiv.org/abs/1301.3120 (Дата обращения 20.05.2018). Загл. с экр. Яз. англ.

12 What is the definition of the density of a graph? [Электронный ресурс]. — URL: https://math.stackexchange.com/questions/1526372/what-is-the-definition-of-the-density-of-a-graph (Дата обращения 20.05.2018). Загл. с экр. Яз. англ.

13 *Кормен, Т.* Алгоритмы. Построение и анализ / Т. Кормен, Ч. Лейзерсон, Р. Ривест, К. Штайн. — Москва: Вильямс, 2016.

14 *Peyton, W.* An Introduction to Chordal Graphs and clique trees / W. Peyton, R.S. Blair. — New York: Name Springer, 1992.

15 *Chartand, G.* A First Course in Graph Theory / G. Chartand, P. Zhang. — Dover Publications, 2006.

16 What's maximal clique? [Электронный ресурс]. — URL: https://math.stackexchange.com/questions/758263/whats-maximal-clique (Дата обращения 20.05.2018). Загл. с экр. Яз. англ.

17 *Balakrishnan, R.* A Textbook of Graph Theory / R. Balakrishnan, K. Balakrishnan. — Springer, 2012.

18 Kaggle [Электронный ресурс]. — URL: https://www.kaggle.com (Дата обращения 20.05.2018). Загл. с экр. Яз. англ.

19 *Klenke, A.* Probability Theory: A Comprehensive Course / A. Klenke. — Springer, 2014.

20 *Jaynes, T.* Probability Theory: The Logic of Science / T. Jaynes. — Cambridge University Press, 2001.