ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ

ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ**

**«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

**Факультет информатики, математики и компьютерных наук**

Программа подготовки магистров по направлению

01.04.02 Прикладная математика и информатика

**Интеллектуальный анализ данных**

**КУРСОВАЯ РАБОТА**

Тема:

**«Исследование динамики связей на рыночном графе»**

|  |  |
| --- | --- |
|  | **Выполнил:** студент группы 22 МАГ ИАД  Рухович Игорь Владимирович  **Руководитель:** профессор,  ведущий научный сотрудник  Колданов Александр Петрович |

1. Оглавление

[1. Оглавление 2](#_Toc137329393)

[2. Введение 3](#_Toc137329394)

[2.1. Актуальность выбранной темы 3](#_Toc137329395)

[3. Теоретическая часть 4](#_Toc137329396)

[3.1. Граф рынка 4](#_Toc137329397)

[3.1.1. Описание рыночного графа 4](#_Toc137329398)

[3.1.2. Определение рыночного графа 4](#_Toc137329399)

[3.1.3. Другие варианты определения графа рынка 5](#_Toc137329400)

[3.2. Клики в рыночном графе 6](#_Toc137329401)

[4. Практическая часть 7](#_Toc137329402)

[4.1. Введение 7](#_Toc137329403)

[4.2. Источник данных 7](#_Toc137329404)

[4.3. Преобразование данных 7](#_Toc137329405)

[4.3.1. Разбиение 7](#_Toc137329406)

[4.3.2. Фильтрация 8](#_Toc137329407)

[4.3.3. Подсчёт корреляций 8](#_Toc137329408)

[4.4. Поиск и выделение максимальных клик 9](#_Toc137329409)

[5. Анализ результатов 11](#_Toc137329410)

[5.1. Общие характеристики клик 11](#_Toc137329411)

[5.1.1. Размеры максимальных клик 11](#_Toc137329412)

[5.1.2. Размеры объединений максимальных клик 12](#_Toc137329413)

[5.1.3. Количество максимальных клик 12](#_Toc137329414)

[5.2. Объединения максимальных клик 13](#_Toc137329415)

[6. Заключение 17](#_Toc137329416)

[7. Список литературы 18](#_Toc137329417)

[8. Приложения 20](#_Toc137329418)

[8.1. Исходный код для выгрузки банковских данных 20](#_Toc137329419)

[8.2. Исходный код для преобразования данных доходности 20](#_Toc137329420)

[8.3. Исходный код для формирования временных периодов 21](#_Toc137329421)

[8.4. Исходный код для построения рыночных графов и поиска макс. клик 21](#_Toc137329422)

[8.5. Исходный код для построения диаграмм вкладов в макс. клики 23](#_Toc137329423)

1. Введение
   1. Актуальность выбранной темы

С момента своего появления и по сей день биржи вызывают к себе большой интерес. Для крупных компаний присутствие на бирже – это способ получить инвестиции без лишних обязательств, иметь дополнительное внешнее управление, разделять владение, и даже награждать отличившихся сотрудников без налоговых обременений. Для простых людей, а также для банков, биржи – источник дополнительного, или даже основного заработка, способ сохранить и преумножить капитал. Особенную популярность торговля на биржах приобрела в эпоху цифровых технологий, когда покупать и продавать ценные бумаги стало возможно онлайн, а порог входа на рынок снизился до наличия стабильного доступа в Интернет и подписания специального договора с одним из участников рынка.

Впрочем, не все инвесторы на рынке что-то зарабатывают. Котировки акций на бирже постоянно меняются, моментально реагируя на различные события в компаниях, а также на глобальные политические изменения, курсы валют. Риск потерять вложения всегда довольно существенный, особенно для неопытных игроков, причем сбывается этот риск достаточно часто как у частных инвесторов, так и у крупных инвестиционных компаний.

Анализ рыночных котировок в последнее время привлекает к себе колоссальное внимание всех заинтересованных сторон. Ведь если какой-то человек или компания узнает, пускай с некоторой погрешностью, что будет с курсом той или иной ценной бумаги на заданном интервале времени, он будет обладать существенным преимуществом перед другими игроками, а значит и возможностью увеличить свой капитал.

1. Теоретическая часть
   1. Граф рынка
      1. Описание рыночного графа

Одной из структур, которой активно пользуются аналитики для понимания и описания рынков являются графы. Более формально, графом:

|  |  |  |
| --- | --- | --- |
|  |  | ф. (3.1) |

называют пару – множество вершин и – множество пар из , означающих рёбра в графе. Если за множество вершин взять множество ценных бумаг на каком-либо рынке, а за множество рёбер – индикаторы связей между курсами этих бумаг, то мы получим концепцию рыночного графа [1, 2, 3]. В графе рынка рёбра между вершинами и проводятся при условии, что коэффициент корреляции между этими вершинами не меньше заданного порога .

* + 1. Определение рыночного графа

Более точно, пусть – цена бумаги на конец торгового дня . Тогда:

|  |  |  |
| --- | --- | --- |
|  |  | ф. (3.2) |

называют доходность акции за период в один день с по . Средней доходностью акции за период в дней называют:

|  |  |  |
| --- | --- | --- |
|  |  | ф. (3.3) |

Дисперсией доходности акции за период в дней называют:

|  |  |  |
| --- | --- | --- |
|  |  | ф. (3.4) |

Тогда коэффициент корреляции Пирсона между акциями и за промежуток времени в дней определяется как:

|  |  |  |
| --- | --- | --- |
|  |  | ф. (3.5) |

Посчитав коэффициенты корреляции для каждой пары акций , можно получить матрицу корреляции размера . Эту матрицу будем интерпретировать как матрицу смежности полного взвешенного графа, вершинами которого являются ценные бумаги. Из определения корреляции Пирсона заметим, что матрица симметрична относительно главной диагонали, поскольку корреляция коммутативна. Это значит, что полученный из неё граф будет неориентированным.

Полный взвешенный граф затруднительно анализировать, поэтому из него принято выделять подмножество рёбер, являющееся в той или иной степени самым важным. Возьмём некоторый порог и построим граф

|  |  |  |
| --- | --- | --- |
|  |  | ф. (3.6) |

Такой граф и принято называть графом рынка [4].

* + 1. Другие варианты определения графа рынка

В литературе также встречаются [5, 6] и другие способы определения рыночного графа. Они аналогичны до этапа получения матрицы смежности, но дальнейшую фильтрацию рёбер можно, например, делать с помощью MST (Maximum Spanning Tree).

Алгоритм построения максимального (или минимального) остовного дерева довольно прост – рёбра графа сортируются по невозрастанию (или неубыванию) весов, а затем поочередно добавляются в пустой граф по условию: добавляем только те рёбра, которые уменьшают число компонент связности. По-другому – если добавление нового ребра создаёт в графе цикл – такое ребро не попадает в MST.

С помощью остовных деревьев можно так же анализировать зависимости между ценными бумагами: “тяжелые” рёбра – высокая зависимость, рёбра “лёгкие” – зависимость между курсами этой пары акций низкая. Если поочерёдно удалять из MST наиболее лёгкие рёбра, можно получить разделение ценных бумаг на связные кластеры.

Дальнейшее развитие этой теории привело [6] к пониманию, что, оставляя от полного графа лишь его остов, мы теряем много информации. Поэтому было предложено добавлять в MST оставшиеся рёбра в порядке невозрастания их весов пока граф остаётся планарным. Планарный граф – граф, рисунок которого убирается на плоскость без пересечения рёбер. Более формально – граф не содержит подграфов, стягивающихся в или [7] (теорема Понтрягина – Куратовского). Такой граф стали называть PMFG (Planar Maximally Filtered Graph).

В данной работе будет рассмотрен рыночный граф только в классическом его представлении – на основе порогов.

* 1. Клики в рыночном графе

Для того, чтобы получить полезную информацию из рыночного графа, предлагается рассматривать его максимальные клики. Клика в графе – это подмножество его вершин, такое, что между любой парой вершин имеется ребро. Максимальная клика – это клика графа, которая при добавлении любой новой вершины, перестаёт быть кликой. К примеру, в PMFG максимальный размер клики по определению планарного графа равен 4 (при наличии в исходном графе хотя бы четырех вершин). Клик максимального размера в графе может быть несколько. Задача поиска максимальной по включению клики в графе относится к классу NP-полных.

Максимальные клики – это то, на что обычно обращают внимание при анализе рыночного графа. На клику можно смотреть как на множество акций, стоимость которых изменяется похожим образом. При этом важно понимать, что хотя попарная корреляция между каждой парой вершин достаточно велика, ещё не значит, что соответствующие им случайные величины зависимы вместе. Для проверки этого факта требуется одновременное выполнение гипотез о попарной зависимости этих величин с заданным уровнем значимости. Впрочем, при анализе рыночных графов этот факт часто опускают.

1. Практическая часть
   1. Введение

Экспериментальная часть работы была частично вдохновлена статьей [8], опубликованной в 2014 году. В ней рассматриваются рыночные графы, построенные на данных об акциях Московской Биржи за промежуток времени с середины 2007 по середину 2011 года. Это время интересно тем, что в середине отрезка (2008 год) произошел глобальный финансовый кризис, безусловно повлиявший на котировки многих акций.

В нашей же работе появился интерес сравнить полученные исследователями данные с ситуацией на текущее время и попытаться отследить похожие моменты и динамику.

* 1. Источник данных

Для нашей работы биржевую информацию было решено брать из Интернет-сервиса Tinkoff Invest API [9]. В сервисе доступны данные по всем акциям, торгующимся на Московской бирже за продолжительный период времени. Мы взяли промежуток, начиная с 12 июня 2015 года и по текущее время.

* 1. Преобразование данных
     1. Разбиение

За эти семь с половиной лет на бирже прошло ровно 2050 торговых дней. Как было предложено в предыдущем исследовании, для удобства сравнения, мы разделили общие данные на равные промежутки по 500 торговых дней со сдвигом в 50 дней. Получилось ровно 32 временных периода, где каждая соседняя пара имеет ровно 450 общих торговых дней и по 50 уникальных дней (рис. 1).

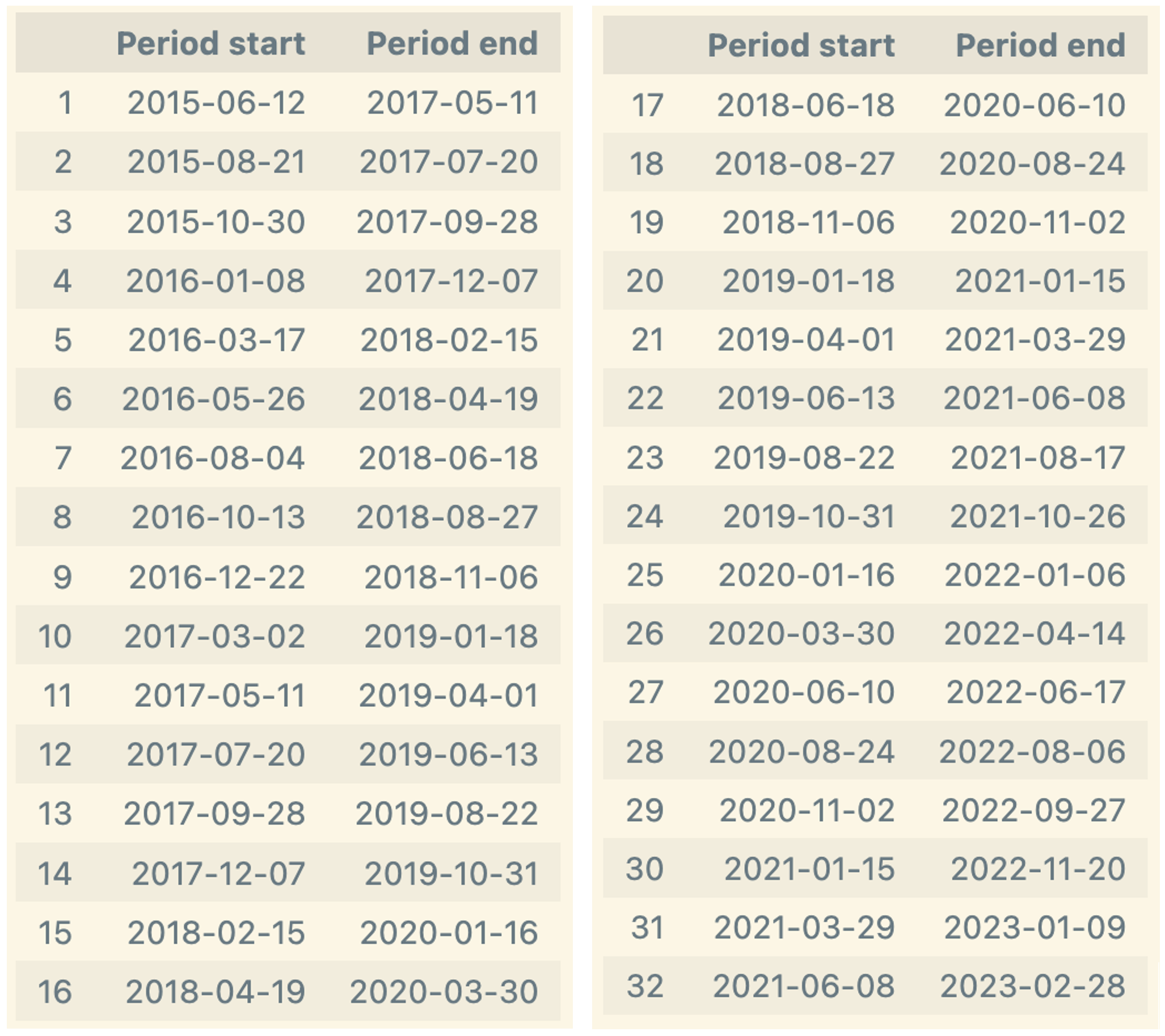


Рисунок 1 – Временные промежутки, взятые в рассмотрение.

Очевидно, что не все компании, акции которых сейчас торгуются на Мосбирже, имели листинги на ней восемь лет назад. В оригинальной статье доходности акций, ещё не вышедших на биржу в момент исследования, принимаются нулевыми. В нашем же эксперименте, в силу большого временного охвата данных, встречается много ценных бумаг, вышедших на биржу позже окончания самого раннего 500-дневного промежутка.

* + 1. Фильтрация

Корреляция Пирсона не определена, если дисперсия одной из случайных величин равна 0, поэтому было принято решение отбросить компании, вышедшие на публичные торги позднее 10 мая 2017 года (окончания первого 500-дневного периода).

Второе дополнение, которое было произведено – это принятие нулевой доходности по всем акциям, которые в силу биржевых ограничений не торговались в определённые торговые дни.

И последняя фильтрация – удаление “дублирующихся” акций. Дело в том, что многие крупные компании выпускают разные типы акций по своим собственным соображениям. В России принято делать “обычные” и “привилегированные” ценные бумаги. Для нашего эксперимента такие типы акций будем считать эквивалентными, поскольку их курсы всегда сильно коррелируют, и, по сути, представляют одни и те же компании на рынке. Итак, если компания на Московской бирже представлена несколькими типами акций, то оставляем только “обычные” ценные бумаги.

После всех фильтраций для экспериментов у нас осталось 87 позиций из 139. Полный список их кратких наименований представлен ниже:



Рисунок 2 – Список кратких наименований ценных бумаг, взятых для эксперимента.

* + 1. Подсчёт корреляций

Для подсчёта попарных созависимостей ценных бумаг было использовано два способа. Во-первых, как и в оригинальной статье, коэффициент корреляции Пирсона – именно с помощью него принято строить рыночный граф. Над исходной формулой была произведена небольшая модификация – использовано абсолютное значение корреляции Пирсона, поскольку отрицательные значения коэффициента так же сигнализируют о зависимости (правда об обратной). Для нашего эксперимента вид зависимости не принципиален. Мы строили графы рынка с порогами вида

|  |  |  |
| --- | --- | --- |
|  |  | ф. (4.1) |

Нижнее и верхнее значения были выбраны исходя из крайних случаев. При все полученные графы полные, и размер максимальной клики равен количеству вершин в графе. При все графы пустые, максимальная клика имеет размер 1.

Второй способ – использование Кенделла. Поскольку коэффициент корреляции Пирсона интерпретируется только для класса эллиптических распределений, к которому может не относиться доходность акций на бирже, использование этой меры может быть не эффективно. В работе [10] была предложена новая мера зависимости доходностей, основанная на вероятности совпадения знаков. Эта мера линейно зависит от коэффициента корреляции знаков Фехнера, который в свою очередь похож на известную всем меру Кенделла. Совпадение рангов, при возможности такового, искусственно избегалось. Графы рынка строились с порогами вида

|  |  |  |
| --- | --- | --- |
|  |  | ф. (4.2) |

При все графы пустые, и размер максимальной клики составляет 1.

* 1. Поиск и выделение максимальных клик

Поиск максимальных клик осуществлялся с помощью аппроксимального метода, поскольку полный перебор всех подмножеств вершин среди 90 штук занял бы значительное время, в то время как эвристические методы [11, 12] работают достаточно точно для наших целей.

Максимальные клики тяжело анализировать, если в графе таковых больше одной. Для решения этой проблемы авторы статьи объединяли множества вершин, входящие в максимальные клики, в одно большое множество, которое уже проще анализировать.

Нами было замечено, что при таком подходе теряется часть важной информации. Дело в том, что некоторые вершины входят почти во все максимальные клики, в то время как другие встречаются только в единичных. В этой связи было решено немного модифицировать метод и учитывать в итоговом множестве вершины из максимальных клик с некоторым весом, равным вероятности встретить эту вершину на каждой позиции в любой максимальной клике (при условии, что хотя бы в одну из максимальных клик эта вершина попала). Другими словами, количество вхождений вершины в клики максимального размера, по отношению к общему числу вхождений любых вершин в максимальные клики.

Таким образом, из каждого графа рынка мы получали описывающее его множество вершин из максимальных клик следующего вида:

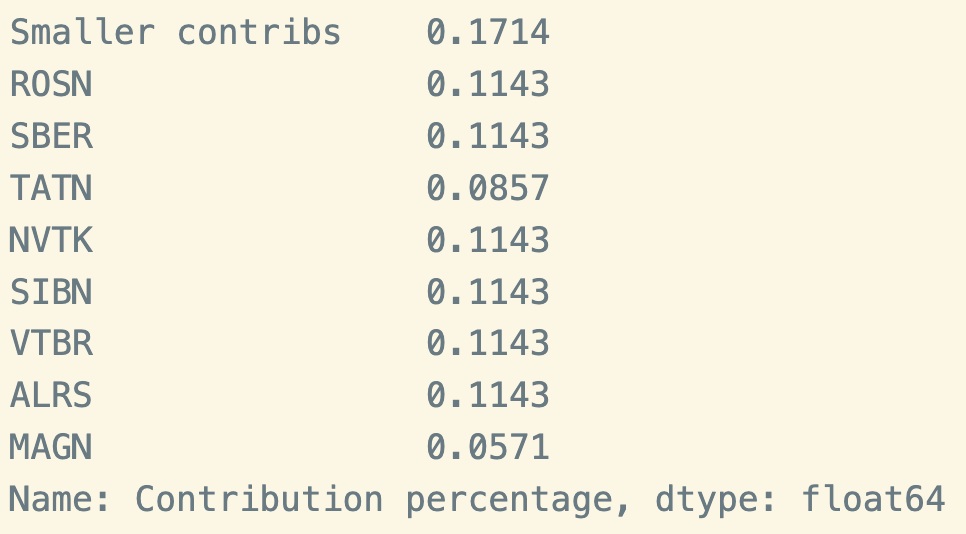


Рисунок 3 – Проценты вкладов каждой из вершин в максимальные клики рыночного графа.

В графу “Smaller contribs” для визуализаций, начиная с гл.5.2, отнесены вершины, процент присутствия которых в максимальных кликах менее заданного порога (в данном примере порог равен 5%). Очевидно, что вклад каждой вершины не может превышать , (в данном примере, 1/7). Сумма вкладов всех вершин равна единице.

1. Анализ результатов
   1. Общие характеристики клик

Для каждого рыночного графа был вычислен размер максимальной клики, количество клик максимального размера, а также размер объединения вершин из максимальных клик.

* + 1. Размеры максимальных клик

На рисунках ниже представлены размеры максимальных клик в рыночных графах, построенных через корреляции Пирсона и Кендалла с разными порогами за различные периоды. Стоит заметить, что с 16 периода в мире началась эпидемия коронавирусной инфекции 2019-nCoV, а с 26 периода – острая фаза политических разногласий на Украине.

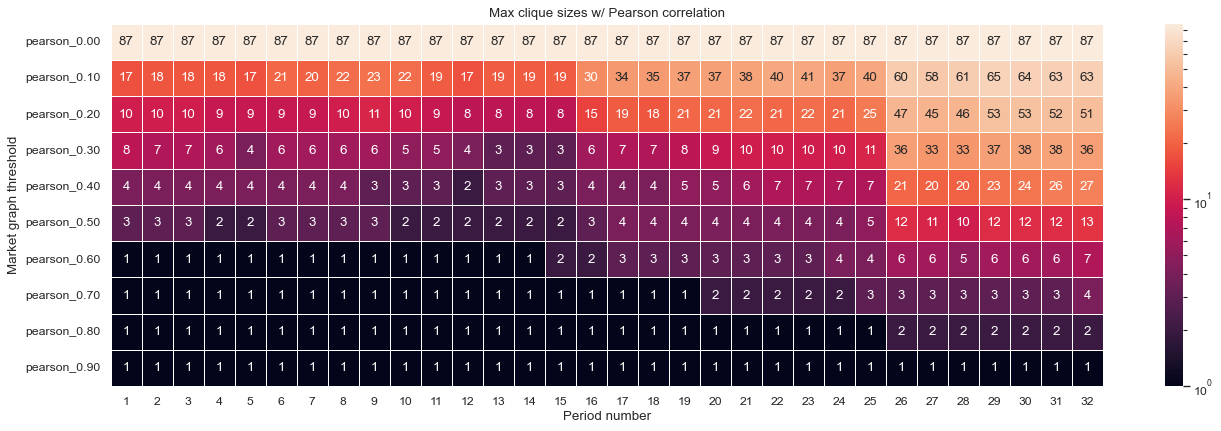


Рисунок 4 – Размеры максимальных клик в рыночных графах с различными порогами корреляции Пирсона

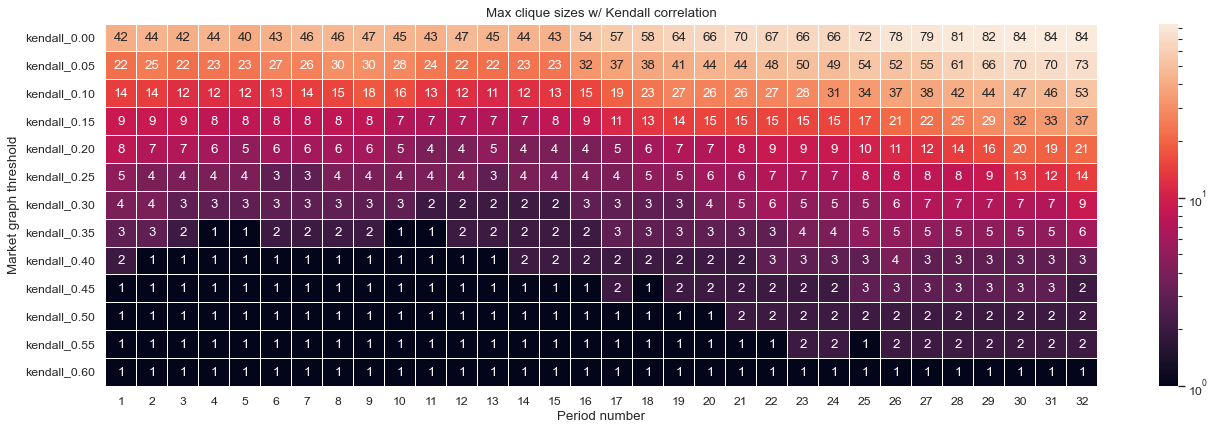


Рисунок 5 – Размеры максимальных клик в рыночных графах с различными порогами корреляции Кендалла

На рис. 4 отчетливо видна разница до 16 периода, до 26 и после. С началом любых потрясений в мире российский рынок акций теряет сосредоточенность вокруг узкого круга акций и начинает вести себя менее предсказуемо. Причём конфликт на Украине, исходя из изображений выше, отразился на рынке намного серьёзнее эпидемии.

* + 1. Размеры объединений максимальных клик

Ниже представлены размеры объединений максимальных клик за те же периоды. Наблюдаем аналогичную картину. Причём для корреляции Пирсона картинка намного более наглядная:

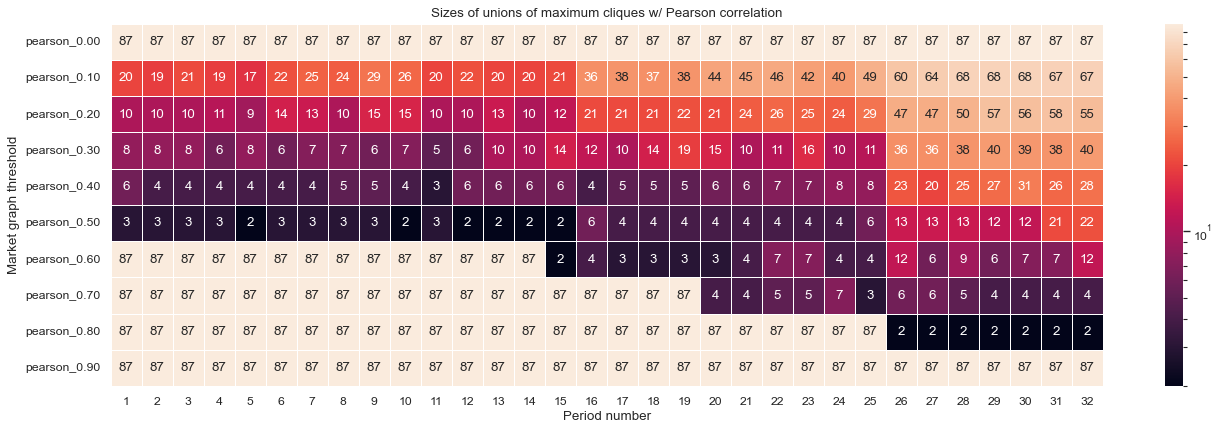


Рисунок 6 – Размеры объединений макс. клик в рыночных графах с различными порогами корреляции Пирсона

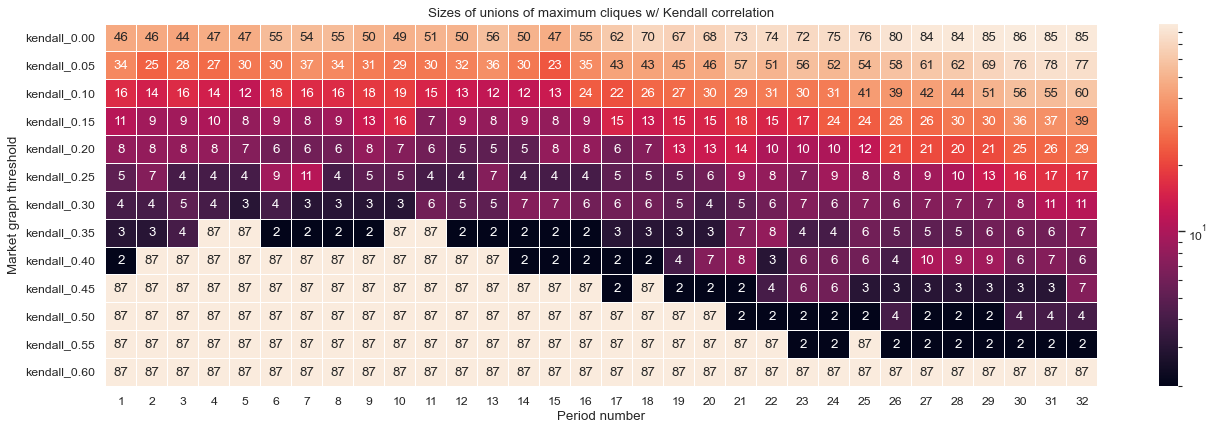


Рисунок 7 – Размеры объединений макс. клик в рыночных графах с различными порогами корреляции Кендалла

* + 1. Количество максимальных клик

Ну и, наконец, картины количества максимальных клик. По рисункам ниже тяжело делать какие-либо выводы. Тенденция на увеличение количества максимальных клик со временем едва прослеживается.

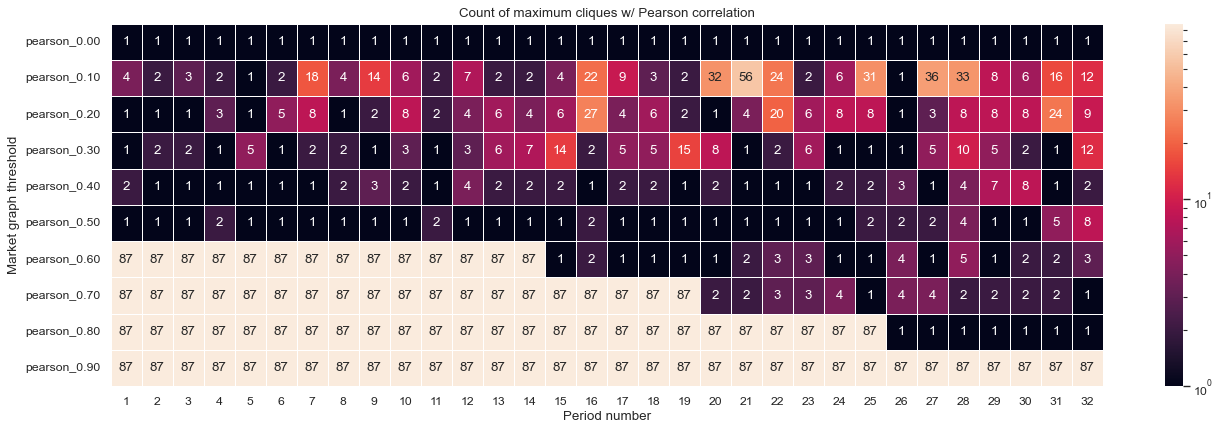


Рисунок 8 – Количества клик макс. размера в рыночных графах с различными порогами корреляции Пирсона

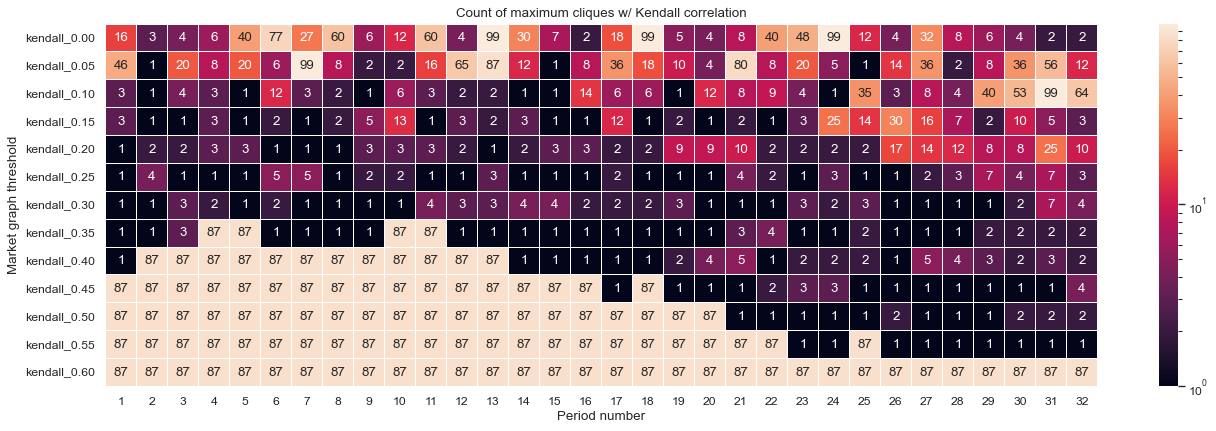


Рисунок 9 – Количества клик макс. размера в рыночных графах с различными порогами корреляции Кендалла

В целом анализ размеров клик не противоречит результатам 2007-2011 годов. В периоды потрясений российский рынок теряет свою “опору”, ведёт себя непредсказуемо.

* 1. Объединения максимальных клик

Далее мы старались определить акции, составляющие эту самую “опору”. Из рис. 10 и рис. 11 можно понять, что основными двигателями российского рынка в обычное время являются крупные нефтяные и газовые компании – добывающая отрасль. Это Роснефть, Лукойл, Газпром, Норильский никель, Газпром нефть. Меньшее влияние оказывает банковская сфера в лице компаний Сбербанк и ВТБ.

С началом же возмущений центральные игроки стали терять свой вес в кликах. В периодах, начиная с 2020 года вклады любых компаний в максимальные клики не превышают 10% (рис. 10). В целом заметно, что со времени предыдущего исследования централизация рынка в нормальное время сместилась практически полностью в добывающий сектор экономики.

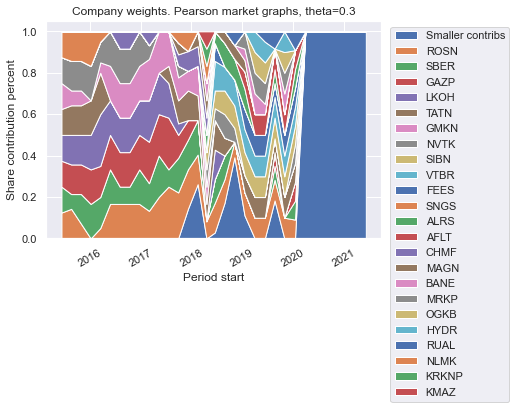


Рисунок 10 – Вклады компаний в макс. клики в различных рыночных графах. Theta Пирсона не менее 0.3. Показаны вклады от 10%

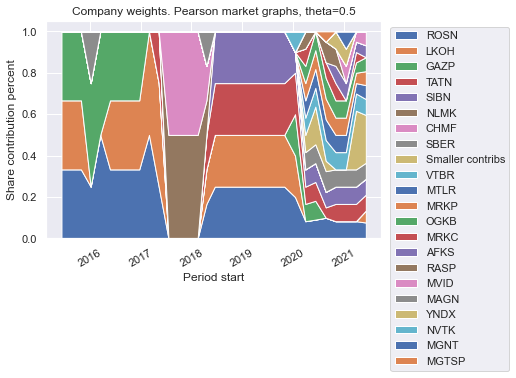


Рисунок 11 – Вклады компаний в макс. клики в различных рыночных графах. Theta Пирсона не менее 0.5. Показаны вклады от 5%.

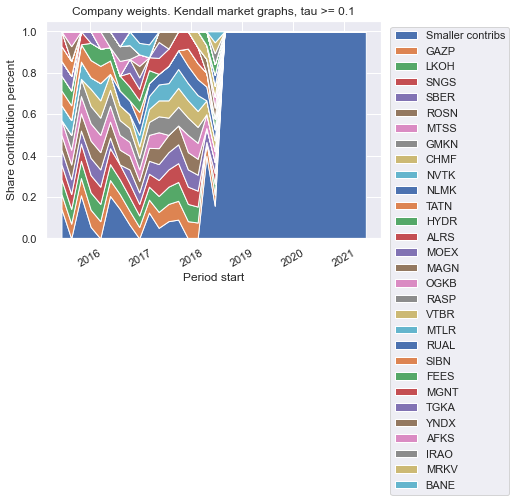


Рисунок 12 – Вклады компаний в макс. клики в различных рыночных графах. Tau Кендалла не менее 0.1. Показаны вклады от 5%

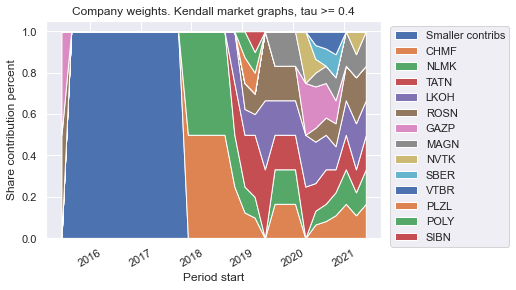


Рисунок 13 – Вклады компаний в макс. клики в различных рыночных графах. Tau Кендалла не менее 0.4. Показаны вклады от 3%

Картины на основании Кенделла менее информативны, но всё же на рис. 13 отчётливо видно участок до 2018 года (пустые рыночные графы, размер максимальной клики равен 1), участок до 2020 (максимальная клика размера 2 представлена добывающими компаниями Северсталь и НЛМК), и участок после – больший размер максимальной клики.

1. Заключение

Было проведено исследование динамики рыночных графов и связей в них. Мы познакомились с понятием рыночного графа, научились их строить различными способами, а также поняли смысл понятия клики в рыночном графе, изучили способы их нахождения и анализа. Исследовали рыночные графы, построенные на настоящих биржевых данных, полученных, начиная с 2015 года и сравнили результаты с исследованием десятилетней давности.

Мы также привнесли некоторые дополнения к исследованию – строили рыночные графы с помощью корреляции Кендалла, а объединения вершин из максимальных клик учитывали с весами, чтобы нагляднее видеть главные элементы в кликах.

1. Список литературы
2. Boginski, V., Butenko, S., Pardalos, P.M.: Mining market data: a network approach // Comput. Oper. Res. 33(11):3171–3184, (2006)
3. Boginski, V., Butenko, S., Pardalos, P.M.: On structural properties of the market graph // Innov. Financ. Econ. Netw. 29–45, (2003)
4. Boginski, V., Butenko, S., Pardalos, P.M.: Statistical analysis of financial networks // Comput. Stat. Data Anal. 48(2):431–443, (2005)
5. Mantegna R.N., Stanley H.E.: An introduction to econophysics: corrleations and complexity in finance // Cambridge Universiy Press, Cambridge, (2000)
6. Kalyagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P.M.: Statistical analysis of graph structures in random variable networks // Springer Brief in Optimization, Springer, (2020)
7. Tumminello, M., Aste, T., Di Matteo, T., Mantegna, R.N.: A tool for filtering information in complex systems // Proc. Natl. Acad. Sci. U.S.A. 102(30):10421–10426, (2005)
8. Kuratowski, K.: Sur le problème des courbes gauches en topologie, Fund. Math. Т. 15: 271–283, (1930)
9. Vizgunov, A.N., Goldengorin, B., Kalyagin, V.A., Koldanov, A.P., Koldanov, P., Pardalos, P.M.: Network approach for the Russian stock market // Comput. Manag. Sci. 11:45–55, (2014)
10. Tinkoff Invest API // Сервис взаимодействия с биржей через брокера Tinkoff. [Электронный ресурс]. – Режим доступа: <https://www.tinkoff.ru/invest/open-api/> (Дата обращения: 27.05.2023)
11. Bautin G.A., Kalyagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P.M.: Simple Measure of Similarity for the Market Graph Construction // Computational Management Science. Vol. 10. P. 105—124, (2013)
12. Boppana, R., Halldórsson, M. M.: Approximating maximum independent sets by excluding subgraphs // BIT Numerical Mathematics, 32(2), 180–196. Springer, (1992)
13. Bron, C., Kerbosch, J.: Algorithm 457: finding all cliques of an undirected graph // Communications of the ACM 16, 9, 575–577, (1973)
14. Tomita, E., Tanaka, A., Takahashi, H.: The worst-case time complexity for generating all maximal cliques and computational experiments // Theoretical Computer Science, Volume 363, Issue 1, Computing and Combinatorics, 10th Annual International Conference on Computing and Combinatorics, 25 October 2006, Pages 28–42, (2006)
15. Cazals, F., Karande, C.: A note on the problem of reporting maximal cliques // Theoretical Computer Science, Volume 407, Issues 1–3, 6 November 2008, Pages 564–568, (2008)
16. Приложения
    1. Исходный код для выгрузки банковских данных

for idx, row in tqdm(shares.iterrows()):

share\_data = {}

to\_s = [now() - timedelta(days=365 \* i) for i in range(0, years\_needed)]

for to in to\_s:

candles = None

while candles is None:

try:

candles = client.market\_data.get\_candles(from\_=to - timedelta(days=365),

to=to,

interval=CandleInterval.CANDLE\_INTERVAL\_DAY,

instrument\_id=row["uid"]).candles

except:

print("Too many requests. Sleeping for 5 seconds.")

time.sleep(5)

for candle in candles:

share\_data[candle.time.date()] = [quot\_to\_price(candle.open),

quot\_to\_price(candle.close)]

share\_df = pd.DataFrame.from\_dict(share\_data, orient="index", columns=["open", "close"]).sort\_index()

share\_df.to\_csv(daily\_data\_folder + daily\_data\_fname(row["ticker"], row["uid"]))

* 1. Исходный код для преобразования данных доходности

*# Count share returns*

*# "fill" - fill with 0*

*# "cut" - cut to maximum start date*

act\_w\_missing\_starts = "fill"

*# If the stock started trading after this date - remove it*

max\_first\_date\_to\_exist = datetime.strptime(date\_to, '%d.%m.%Y') - timedelta(days=365 \* years - 50)

common\_df, date\_start = None, None

for idx, info in shares\_info.iterrows():

cur\_df = pd.read\_csv(daily\_data\_folder + daily\_data\_fname(info["ticker"], info["uid"]), index\_col=0)

cur\_df.index = pd.to\_datetime(cur\_df.index)

*# From the book*

cur\_df = pd.DataFrame(np.log(cur\_df["close"] / cur\_df["open"]), columns=[info["ticker"]])

*# Diff percentage*

*# cur\_df = pd.DataFrame((cur\_df["close"] - cur\_df["open"]) / cur\_df["open"], columns=[info["ticker"]])*

cur\_min\_date = cur\_df.index.min()

print(**f**"{info['name']}\t first candle date is: {cur\_min\_date}")

if max\_first\_date\_to\_exist is not None and cur\_min\_date > max\_first\_date\_to\_exist:

print(**f**"Stock {info['ticker']} started trading too late for research")

continue

if common\_df is None:

common\_df = cur\_df

date\_start = cur\_min\_date

else:

common\_df = common\_df.join(cur\_df, how="outer")

if date\_start < cur\_min\_date:

date\_start = cur\_min\_date

if act\_w\_missing\_starts == "fill":

print(**f**"Min date is: {common\_df.index.min()}")

elif act\_w\_missing\_starts == "cut":

common\_df = common\_df.loc[common\_df.index >= date\_start]

print(**f**"Min date is: {date\_start}")

else:

raise RuntimeError("Unknown missing starts action")

common\_df = common\_df.fillna(0).sort\_index().copy()

common\_df

* 1. Исходный код для формирования временных периодов

*# Choosing periods for market graphs. Let's take it monthly, yearly and overall. First inclusive, second exclusive*

periods = []

*# 500 traded-days peroids, 50-days shift*

period = 500

shift = 50

cur\_idx = 0

while common\_df[cur\_idx: cur\_idx + period].shape[0] == period:

periods.append((common\_df.index[cur\_idx], common\_df.index[cur\_idx + period]))

cur\_idx += shift

periods = periods[:32]

periods\_print = pd.DataFrame({"Period start": [period[0] for period in periods],

"Period end": [period[1] for period in periods]},

index=np.arange(1, len(periods) + 1))

periods\_print.head(16)

* 1. Исходный код для построения рыночных графов и поиска макс. клик

*# Build threshold graphs + find max cliques*

thresholds\_range = [("pearson", 0.1\*i) for i in range(10)] + [("kendall", 0.05\*i) for i in range(13)]

data = []

pbar = tqdm(total=len(periods) \* len(thresholds\_range))

for period in periods:

*# Calculating Kendall correlation*

sub\_df = common\_df.loc[(period[0] <= common\_df.index) & (common\_df.index < period[1])]

vertices = sub\_df.columns

edges\_p = np.abs(sub\_df.corr(method="pearson"))

edges\_k = sub\_df.corr(method="kendall")

np.fill\_diagonal(edges\_p.values, 0)

np.fill\_diagonal(edges\_k.values, 0)

data.append((period, {}))

for method, threshold in thresholds\_range:

if method == "pearson":

edges = edges\_p

elif method == "kendall":

edges = edges\_k

else:

raise RuntimeError("Unknown correlation method")

cur\_data\_name = **f**"{method}\_{threshold**:.2f**}"

data[-1][1][cur\_data\_name] = {}

*# Building a threshold market graph*

G = nx.Graph()

G.add\_nodes\_from(vertices)

for i in range(edges.shape[0]):

for j in range(i+1, edges.shape[0]):

if edges.values[i, j] > threshold:

G.add\_edge(vertices[i], vertices[j])

data[-1][1][cur\_data\_name]["graph"] = G

*# Finding maximum cliques*

max\_cliques\_arr = []

max\_cliques\_size = 0

for clique in nx.clique.find\_cliques(G):

if len(clique) > max\_cliques\_size:

max\_cliques\_size = len(clique)

max\_cliques\_arr = [sorted(clique)]

elif len(clique) == max\_cliques\_size:

max\_cliques\_arr.append(sorted(clique))

data[-1][1][cur\_data\_name]["max\_clique\_size"] = max\_cliques\_size

data[-1][1][cur\_data\_name]["max\_cliques\_number"] = len(max\_cliques\_arr)

*# Union of max cliques w/ contribution*

shares\_appearance\_perc = {}

one\_stock\_contrib = max\_cliques\_size \* len(max\_cliques\_arr)

for max\_cl in max\_cliques\_arr:

for share in max\_cl:

if share in shares\_appearance\_perc:

shares\_appearance\_perc[share] += 1 / one\_stock\_contrib

else:

shares\_appearance\_perc[share] = 1 / one\_stock\_contrib

data[-1][1][cur\_data\_name]["shares\_appearance\_perc"] = shares\_appearance\_perc

pbar.update(1)

* 1. Исходный код для построения диаграмм вкладов в макс. клики

**def** plot\_area\_chart(data, data\_name, contrib\_threshold=0.1):

common\_df = None

for period, info in data:

cur\_df = pd.DataFrame.from\_dict(info[data\_name]["shares\_appearance\_perc"], orient="index", columns=[period[0]])

if common\_df is None:

common\_df = cur\_df

else:

common\_df = common\_df.join(cur\_df, how="outer")

common\_df = common\_df.fillna(0).sort\_index()

*# Sum small values and drop empty columns*

other\_sum = common\_df[common\_df < contrib\_threshold].sum().rename("Smaller contribs")

common\_df[common\_df < contrib\_threshold] = 0

common\_df = common\_df.append(other\_sum)

common\_df = common\_df.loc[(common\_df > 0).any(axis=1)].copy()

*# Sort by summary area*

common\_df["sum"] = common\_df.sum(axis=1)

common\_df.sort\_values(by="sum", ascending=False, inplace=True)

common\_df.drop(columns="sum", inplace=True)

ax = plt.stackplot(common\_df.columns,

[row for \_, row in common\_df.iterrows()],

labels=common\_df.index)

plt.legend(bbox\_to\_anchor=(1.4, 1))

plt.xlabel("Period start")

plt.xticks(rotation=30)

plt.ylabel("Share contribution percent")

return common\_df