

Анализ динамической сети акций индекса S&P 500

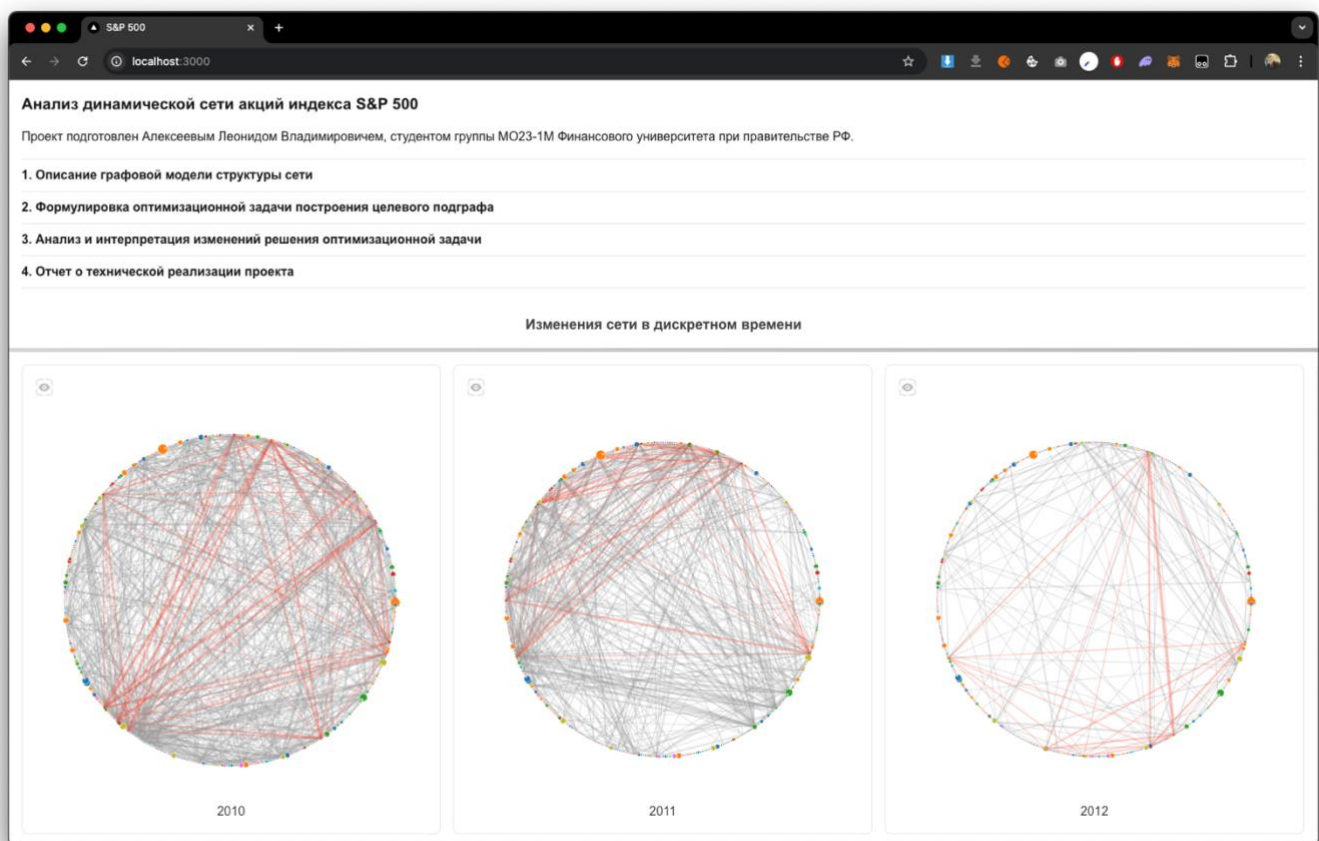
Проект подготовлен Алексеевым Леонидом Владимировичем, студентом группы MO23-1M Финансового университета при правительстве РФ.

Демонстрация проекта

Ссылка на проект:

<https://sp500-d3-graph.vercel.app/>

Скриншот сайта:



1. Описание графовой модели структуры сети

Общая характеристика:

Динамическая структура: Самое главное, что нужно отметить - это нестабильность структуры графа. Количество ребер, средняя степень узлов и стандартное отклонение этих степеней сильно меняются от года к году. Это говорит о том, что взаимосвязи между акциями не являются статичными, а зависят от макроэкономических условий, настроений рынка и других факторов.

Граф не является ни постоянно плотным, ни постоянно разреженным, а переходит между этими состояниями.

Разреженность: В целом, граф является разреженным. Это означает, что большинство акций не имеют сильных корреляций с большинством других акций. Плотность графа (количество ребер) значительно меньше, чем максимально возможное количество ребер в полном графе. Это говорит о том, что рынок акций не является однородным и состоит из групп акций, которые более тесно связаны между собой, чем с остальными.

Преобладание одной связной компоненты: Несмотря на разреженность, граф в основном состоит из одной большой связной компоненты. Это означает, что большинство акций косвенно связаны друг с другом через цепочки корреляций. Хотя прямые корреляции между многими парами акций могут быть слабыми или отсутствовать, существует сеть косвенных связей, которая объединяет большинство акций в единую систему.

Неравномерное распределение связей: Распределение степеней узлов является неравномерным. Это означает, что есть акции, которые имеют очень мало связей (степень близка к 0), и есть акции, которые имеют много связей (степень значительно выше средней). Это может указывать на наличие "хабов" или "ключевых" акций, которые играют важную роль в распространении рыночных движений.

Цикличность и фазы: Наблюдается цикличность в изменениях структуры графа. Периоды высокой корреляции (большое количество ребер, высокая средняя степень) чередуются с периодами низкой корреляции (малое количество ребер, низкая средняя степень). Это может быть связано с различными фазами рыночного цикла (рост, спад, консолидация) или с изменениями в макроэкономической ситуации.

Влияние внешних факторов: Динамика графа, вероятно, отражает влияние внешних факторов, таких как экономические кризисы, политические события, изменения в процентных ставках, технологические прорывы и т.д. Например, 2013 и 2017 годы с высокой корреляцией могут быть связаны с периодами стабильного роста, а 2012, 2015 и 2018 годы с низкой корреляцией - с периодами неопределенности или коррекции.

Эволюция рынка: Постепенное увеличение количества узлов отражает эволюцию рынка и добавление новых компаний в индекс S&P 500. Это также говорит о том, что модель должна быть способна адаптироваться к изменениям в составе рынка.

Общие тенденции:

Рост количества узлов: Наблюдается постепенное увеличение количества узлов (акций) в графе с 439 в 2010 году до 502 в 2024 году. Это отражает естественное развитие рынка и добавление новых компаний в индекс S&P 500.

Волатильность количества ребер: Количество ребер (сильных корреляций) сильно варьируется от года к году. Это говорит о том, что структура корреляций между акциями не является постоянной и меняется в зависимости от рыночных условий.

Низкая плотность графа: Плотность графа, которая равна количеству ребер, относительно невелика по сравнению с количеством узлов. Это означает, что граф является разреженным, и большинство акций не имеют сильных корреляций друг с другом.

Низкое количество компонент связности: Значение количества компонент связности близко к нулю, что говорит о том, что граф в основном состоит из одной большой связной компоненты. Это означает, что большинство акций косвенно связаны друг с другом через цепочки корреляций.

Наличие узлов с нулевой степенью: Во все годы есть узлы со степенью 0, что означает, что есть акции, которые не имеют сильных корреляций ни с одной другой акцией в данный год.

Широкий диапазон степеней узлов: Наблюдается большой разброс в степенях узлов (от 0 до максимального значения), что указывает на наличие как изолированных акций, так и акций, сильно коррелирующих с многими другими.

Анализ по годам:

Характеристики графа 2010:

Количество узлов: 439

Количество ребер: 718

Плотность графа: 718

Количество компонент связности: 0.007468197751219563

Степени узлов - мин: 0

Степени узлов - макс: 43

Степени узлов - среднее: 3.2710706150341684

Степени узлов - стандартное отклонение: 6.804635573601475

Относительно высокая плотность графа (718 ребер), средняя степень узлов (3.27) и стандартное отклонение (6.80) говорят о том, что в этом году было больше сильных корреляций, но они были распределены неравномерно.

Характеристики графа 2011:

Количество узлов: 445

Количество ребер: 478

Плотность графа: 478

Количество компонент связности: 0.00483854641158012

Степени узлов - мин: 0

Степени узлов - макс: 31

Степени узлов - среднее: 2.148314606741573

Степени узлов - стандартное отклонение: 5.232050232034227

Снижение количества ребер (478) и средней степени (2.15) по сравнению с 2010 годом. Это может указывать на снижение общей корреляции между акциями.

Характеристики графа 2012:

Количество узлов: 453

Количество ребер: 132

Плотность графа: 132

Количество компонент связности: 0.001289339506534607

Степени узлов - мин: 0

Степени узлов - макс: 14

Степени узлов - среднее: 0.5827814569536424

Степени узлов - стандартное отклонение: 2.0264542988042242

Резкое снижение количества ребер (132) и средней степени (0.58) говорит о том, что в этом году было очень мало сильных корреляций между акциями. Рынок, возможно, был более фрагментированным.

Характеристики графа 2013:

Количество узлов: 462

Количество ребер: 2518

Плотность графа: 2518

Количество компонент связности: 0.02364519067339024

Степени узлов - мин: 0

Степени узлов - макс: 80

Степени узлов - среднее: 10.9004329004329

Степени узлов - стандартное отклонение: 19.366396412534225

Резкий рост количества ребер (2518) и средней степени (10.9) по сравнению с 2012 годом. Это указывает на сильное увеличение корреляций между акциями. Возможно, рынок был более синхронизированным.

Характеристики графа 2014:

Количество узлов: 469

Количество ребер: 502

Плотность графа: 502

Количество компонент связности: 0.00457419860404935

Степени узлов - мин: 0

Степени узлов - макс: 29

Степени узлов - среднее: 2.140724946695096

Степени узлов - стандартное отклонение: 4.981742768528988

Снижение количества ребер (502) и средней степени (2.14) по сравнению с 2013 годом. Корреляции между акциями снова ослабевают.

Характеристики графа 2015:

Количество узлов: 474

Количество ребер: 127

Плотность графа: 127

Количество компонент связности: 0.001132906932141551

Степени узлов - мин: 0
Степени узлов - макс: 10
Степени узлов - среднее: 0.5358649789029536
Степени узлов - стандартное отклонение: 1.3213887685741268

Дальнейшее снижение количества ребер (127) и средней степени (0.54), аналогично 2012 году.

Характеристики графа 2016:
Количество узлов: 479
Количество ребер: 599
Плотность графа: 599
Количество компонент связности: 0.005232309291498153
Степени узлов - мин: 0
Степени узлов - макс: 23
Степени узлов - среднее: 2.501043841336117
Степени узлов - стандартное отклонение: 4.941445733528242

Небольшое увеличение количества ребер (599) и средней степени (2.50) по сравнению с 2015 годом.

Характеристики графа 2017:
Количество узлов: 481
Количество ребер: 2070
Плотность графа: 2070
Количество компонент связности: 0.017931392931392932
Степени узлов - мин: 0
Степени узлов - макс: 75
Степени узлов - среднее: 8.607068607068607
Степени узлов - стандартное отклонение: 16.043478423667977

Сильный рост количества ребер (2070) и средней степени (8.61), похожий на 2013 год.

Характеристики графа 2018:
Количество узлов: 484
Количество ребер: 204
Плотность графа: 204
Количество компонент связности: 0.0017452902828396899
Степени узлов - мин: 0
Степени узлов - макс: 73
Степени узлов - среднее: 0.8429752066115702
Степени узлов - стандартное отклонение: 3.535699659501603

Резкое снижение количества ребер (204) и средней степени (0.84), снова видим фрагментированный рынок.

Характеристики графа 2019:
Количество узлов: 490

Количество ребер: 936
Плотность графа: 936
Количество компонент связности: 0.007812695630399399
Степени узлов - мин: 0
Степени узлов - макс: 39
Степени узлов - среднее: 3.820408163265306
Степени узлов - стандартное отклонение: 7.785492764338896

Увеличение количества ребер (936) и средней степени (3.82), корреляции снова усиливаются.

Характеристики графа 2020:
Количество узлов: 494
Количество ребер: 1401
Плотность графа: 1401
Количество компонент связности: 0.011505202388089119
Степени узлов - мин: 0
Степени узлов - макс: 63
Степени узлов - среднее: 5.672064777327935
Степени узлов - стандартное отклонение: 7.731403061150662

Дальнейший рост количества ребер (1401) и средней степени (5.67), возможно, из-за пандемии и ее влияния на рынок.

Характеристики графа 2021:
Количество узлов: 494
Количество ребер: 755
Плотность графа: 755
Количество компонент связности: 0.00620016260029071
Степени узлов - мин: 0
Степени узлов - макс: 37
Степени узлов - среднее: 3.0566801619433197
Степени узлов - стандартное отклонение: 7.189751122802579

Снижение количества ребер (755) и средней степени (3.06) по сравнению с 2020 годом.

Характеристики графа 2022:
Количество узлов: 496
Количество ребер: 281
Плотность графа: 281
Количество компонент связности: 0.0022890192245030956
Степени узлов - мин: 0
Степени узлов - макс: 15
Степени узлов - среднее: 1.1330645161290323
Степени узлов - стандартное отклонение: 2.61588687162159

Граф 2022 года характеризуется низкой связностью, с небольшим количеством корреляций между акциями, что указывает на относительно изолированные

движения цен на рынке. Средняя степень узла также низкая, что говорит о том, что большинство акций не сильно связаны друг с другом.

Характеристики графа 2023:

Количество узлов: 498

Количество ребер: 160

Плотность графа: 160

Количество компонент связности: 0.0012928979499486882

Степени узлов - мин: 0

Степени узлов - макс: 10

Степени узлов - среднее: 0.642570281124498

Степени узлов - стандартное отклонение: 1.5358738460256935

В 2023 году граф становится еще более разреженным, с крайне низким количеством связей между акциями и очень низкой средней степенью узла, что указывает на то, что рынок в целом был слабо скоординирован и акции двигались достаточно независимо друг от друга.

Характеристики графа 2024:

Количество узлов: 502

Количество ребер: 767

Плотность графа: 767

Количество компонент связности: 0.006099355074711136

Степени узлов - мин: 0

Степени узлов - макс: 39

Степени узлов - среднее: 3.055776892430279

Степени узлов - стандартное отклонение: 6.944668433749338

Граф 2024 года демонстрирует значительное увеличение связности, с заметным ростом количества корреляций и средней степени узла, что говорит о том, что рынок стал более взаимосвязанным, и движения цен акций стали более скоординированными по сравнению с предыдущими годами.

Заключение:

Графовая модель корреляций акций S&P 500 представляет собой динамическую и сложную систему, которая отражает взаимосвязи между акциями и их изменения во времени. Она не является статичной, а постоянно адаптируется к рыночным условиям. Эта модель характеризуется разреженностью, наличием одной большой связной компоненты, неравномерным распределением связей и цикличностью в изменениях. Анализ этой модели может дать ценную информацию о структуре рынка, его уязвимостях и потенциальных возможностях.

2. Формулировка оптимизационной задачи построения целевого подграфа

Задача:

Найти максимальную клику (подграф, в котором каждая пара вершин связана ребром) в каждом графе траектории динамического графа акций SP500, где графы представляют собой корреляционные сети акций за каждый год с 2010 по 2024.

Формализация:

Динамический граф: $G = \{G_{2010}, G_{2011}, \dots, G_{2024}\}$, где G_t - граф корреляций акций за год t .

Вершины: V_t - множество вершин графа G_t , представляющие акции.

Ребра: E_t - множество ребер графа G_t , где ребро (u, v) существует, если корреляция между акциями u и v превышает определенный порог (например, 0.7).

Целевая функция: Для каждого графа G_t найти клику $C_t \subseteq V_t$ максимального размера (количества вершин).

Ограничения: Подграф C_t должен быть кликой, то есть для любых двух вершин $u, v \in C_t$ должно существовать ребро $(u, v) \in E_t$.

Математическая запись:

Для каждого графа $G_t \in G$, максимизировать $|C_t|$ при условии, что $\forall u, v \in C_t, (u, v) \in E_t$.

Обоснование прикладного аспекта:

Выявление групп активов с высокой корреляцией: Максимальная клика в графе корреляций акций представляет собой группу активов, которые имеют высокую степень взаимосвязи. Это означает, что их цены движутся в одном направлении, что может быть полезно для различных финансовых задач.

Управление рисками: Зная группы высокоррелированных активов, инвесторы могут более эффективно управлять рисками своего портфеля. Инвестирование в несколько активов из одной клики может не обеспечить желаемой диверсификации, так как они будут реагировать на рыночные изменения схожим образом.

Построение портфелей: Понимание структуры корреляций позволяет строить более сбалансированные портфели. Инвесторы могут избегать перекоса в сторону активов, которые сильно зависят друг от друга.

Торговые стратегии: Выявление клик может помочь в разработке торговых стратегий, основанных на парной торговле (торговле разницей в ценах между коррелированными активами).

Анализ динамики рынка: Изучение изменения максимальных клик во времени (на траектории динамического графа) позволяет отслеживать, как меняются взаимосвязи между акциями с течением времени. Это может дать представление о структурных изменениях на рынке.

Идентификация секторов: Клика может указывать на то, что несколько акций принадлежат к одному сектору экономики, что может быть полезно для анализа рынка.

Обоснование выбора максимальной клики:

Максимальная взаимосвязь: Максимальная клика представляет собой группу активов с максимальной степенью взаимосвязи, что делает ее интересной для анализа.

Простота интерпретации: Клика – это относительно простая структура, которую легко интерпретировать.

Широкая применимость: Задача поиска максимальной клики является классической задачей теории графов и имеет множество применений, в том числе и в финансах.

3. Анализ и интерпретация изменений решения оптимизационной задачи

Общие наблюдения:

Динамика размеров клик: Размер максимальной клики значительно меняется от года к году, что говорит о том, что структура корреляций между акциями не является постоянной и подвержена значительным изменениям.

Периоды высокой и низкой корреляции: Наблюдаются периоды с большими кликами (например, 2013, 2017), что указывает на высокую корреляцию между многими активами, и периоды с малыми кликами (например, 2015, 2018, 2023), что говорит о низкой корреляции.

Плотность графов: Все клики являются полными графами, что означает, что все активы внутри клики сильно коррелируют друг с другом.

Степени вершин: Степени всех вершин в клике равны количеству вершин минус 1, что подтверждает, что это полные графы.

Анализ изменений по годам:

Характеристики целевого подграфа 2010:

Количество узлов: 11

Количество ребер: 55

Плотность графа: 55

Количество компонент связности: 1.0

Степени узлов - мин: 10

Степени узлов - макс: 10

Степени узлов - среднее: 10

Степени узлов - стандартное отклонение: 0.0

Максимальная клика состоит из 11 акций, образуя плотный подграф с высокой степенью связности (каждая акция связана со всеми остальными).

Характеристики целевого подграфа 2011:

Количество узлов: 9

Количество ребер: 36

Плотность графа: 36
Количество компонент связности: 1.0
Степени узлов - мин: 8
Степени узлов - макс: 8
Степени узлов - среднее: 8
Степени узлов - стандартное отклонение: 0.0

Размер клики уменьшился до 9 акций, но по-прежнему сохраняется высокая плотность и связность внутри группы.

Характеристики целевого подграфа 2012:

Количество узлов: 7
Количество ребер: 21
Плотность графа: 21
Количество компонент связности: 1.0
Степени узлов - мин: 6
Степени узлов - макс: 6
Степени узлов - среднее: 6
Степени узлов - стандартное отклонение: 0.0

Продолжается уменьшение размера клики до 7 акций, что говорит о снижении количества сильно коррелированных активов.

Характеристики целевого подграфа 2013:

Количество узлов: 25
Количество ребер: 300
Плотность графа: 300
Количество компонент связности: 1.0
Степени узлов - мин: 24
Степени узлов - макс: 24
Степени узлов - среднее: 24
Степени узлов - стандартное отклонение: 0.0

Происходит резкое увеличение размера клики до 25 акций, что свидетельствует о периоде высокой корреляции между многими активами.

Характеристики целевого подграфа 2014:

Количество узлов: 8
Количество ребер: 28
Плотность графа: 28
Количество компонент связности: 1.0
Степени узлов - мин: 7
Степени узлов - макс: 7
Степени узлов - среднее: 7
Степени узлов - стандартное отклонение: 0.0

Размер клики снова уменьшается до 8 акций, возвращаясь к более умеренному уровню корреляции.

Характеристики целевого подграфа 2015:

Количество узлов: 4

Количество ребер: 6

Плотность графа: 6

Количество компонент связности: 1.0

Степени узлов - мин: 3

Степени узлов - макс: 3

Степени узлов - среднее: 3

Степени узлов - стандартное отклонение: 0.0

Наблюдается значительное сокращение клики до 4 акций, что может указывать на усиление дифференциации в поведении активов.

Характеристики целевого подграфа 2016:

Количество узлов: 11

Количество ребер: 55

Плотность графа: 55

Количество компонент связности: 1.0

Степени узлов - мин: 10

Степени узлов - макс: 10

Степени узлов - среднее: 10

Степени узлов - стандартное отклонение: 0.0

Размер клики увеличивается до 11 акций, что говорит о возвращении к более высокой корреляции между некоторыми активами.

Характеристики целевого подграфа 2017:

Количество узлов: 19

Количество ребер: 171

Плотность графа: 171

Количество компонент связности: 1.0

Степени узлов - мин: 18

Степени узлов - макс: 18

Степени узлов - среднее: 18

Степени узлов - стандартное отклонение: 0.0

Происходит значительное увеличение размера клики до 19 акций, что свидетельствует о периоде высокой взаимосвязанности активов.

Характеристики целевого подграфа 2018:

Количество узлов: 4

Количество ребер: 6

Плотность графа: 6

Количество компонент связности: 1.0

Степени узлов - мин: 3

Степени узлов - макс: 3

Степени узлов - среднее: 3

Степени узлов - стандартное отклонение: 0.0

Размер клики снова резко уменьшается до 4 акций, что указывает на период сниженной корреляции между активами.

Характеристики целевого подграфа 2019:

Количество узлов: 10

Количество ребер: 45

Плотность графа: 45

Количество компонент связности: 1.0

Степени узлов - мин: 9

Степени узлов - макс: 9

Степени узлов - среднее: 9

Степени узлов - стандартное отклонение: 0.0

Размер клики увеличивается до 10 акций, что говорит о некотором усилении корреляции между активами.

Характеристики целевого подграфа 2020:

Количество узлов: 9

Количество ребер: 36

Плотность графа: 36

Количество компонент связности: 1.0

Степени узлов - мин: 8

Степени узлов - макс: 8

Степени узлов - среднее: 8

Степени узлов - стандартное отклонение: 0.0

Размер клики остается относительно стабильным, уменьшаясь до 9 акций, что может указывать на сохранение схожих тенденций.

Характеристики целевого подграфа 2021:

Количество узлов: 13

Количество ребер: 78

Плотность графа: 78

Количество компонент связности: 1.0

Степени узлов - мин: 12

Степени узлов - макс: 12

Степени узлов - среднее: 12

Степени узлов - стандартное отклонение: 0.0

Размер клики увеличивается до 13 акций, что свидетельствует о росте взаимосвязанности активов.

Характеристики целевого подграфа 2022:

Количество узлов: 5

Количество ребер: 10

Плотность графа: 10

Количество компонент связности: 1.0
Степени узлов - мин: 4
Степени узлов - макс: 4
Степени узлов - среднее: 4
Степени узлов - стандартное отклонение: 0.0

Размер клики уменьшается до 5 акций, что может говорить о снижении корреляции между активами.

Характеристики целевого подграфа 2023:

Количество узлов: 4
Количество ребер: 6
Плотность графа: 6
Количество компонент связности: 1.0
Степени узлов - мин: 3
Степени узлов - макс: 3
Степени узлов - среднее: 3
Степени узлов - стандартное отклонение: 0.0

Размер клики остается небольшим, 4 акции, что указывает на сохранение низкой корреляции между активами.

Характеристики целевого подграфа 2024:

Количество узлов: 8
Количество ребер: 28
Плотность графа: 28
Количество компонент связности: 1.0
Степени узлов - мин: 7
Степени узлов - макс: 7
Степени узлов - среднее: 7
Степени узлов - стандартное отклонение: 0.0

Размер клики увеличивается до 8 акций, что может свидетельствовать о некотором усилении корреляции между активами.

4. Отчет о технической реализации проекта

Введение:

Данный проект посвящен анализу динамики корреляций между акциями, входящими в индекс S&P 500, с использованием графовых моделей. Целью проекта является визуализация и анализ изменений в структуре корреляций акций на протяжении времени (с 2010 по 2024 год) посредством выделения максимальных клик в графах корреляций за каждый год.

Технологии и инструменты:

Данные:

Датасет: Использован датасет исторических данных по акциям S&P 500, доступный на Kaggle [andrewmvd/sp-500-stocks](https://www.kaggle.com/andrewmvd/sp-500-stocks).

Формат данных: CSV-файлы с ценами акций за каждый год.

Библиотеки Python:

`pandas`: Для обработки и анализа данных, загрузки данных из CSV-файлов.

`numpy`: Для математических вычислений, в частности, для расчета корреляционных матриц.

`networkx`: Для построения и анализа графов, включая поиск максимальных клик.

Фронтенд:

`React`: Для создания пользовательского интерфейса.

`Next.js`: Фреймворк React для серверного рендеринга и маршрутизации.

`d3.js`: Библиотека для визуализации данных, в частности, для отображения графов.

Развертывание:

`Vercel`: Платформа для развертывания веб-приложений.

Этапы реализации:

1. Сбор и подготовка данных:

Загрузка исторических данных по акциям S&P 500 из датасета Kaggle.

Обработка данных:

Фильтрация данных для каждого года с 2010 по 2024.

Расчет корреляционных матриц между акциями за каждый год.

Создание графов корреляций: ребро между двумя акциями добавляется, если их корреляция превышает заданный порог (например, 0.7).

2. Построение и анализ графов:

Использование `networkx` для создания графов корреляций для каждого года.

Реализация алгоритма поиска максимальной клики в каждом графе.

Расчет характеристик максимальных клик: количество узлов, количество ребер, плотность графа, количество компонент связности, степени узлов.

3. Фронтенд разработка:

Создание React-приложения с использованием Next.js.

Разработка компонентов для визуализации графов с использованием `d3.js`.

Реализация интерактивности: возможность выбора года для просмотра соответствующего графа и его характеристик.

4. Развертывание:

Деплой приложения на платформу `Vercel`.

Оптимизационная задача:

Задача: Найти максимальную клику в каждом графе корреляций акций за каждый год.

Целевая функция: Максимизировать количество вершин в клике.

Алгоритм: Использован встроенный в `networkx` алгоритм поиска максимальной клики.

Визуализация:

Визуализация графов корреляций с использованием `d3.js`.

Минимизация пересечений ребер при визуализации.

Отображение максимальной клики в каждом графе.

Отображение характеристик максимальной клики (количество узлов, ребер, плотность и т.д.).

Результаты:

Получены максимальные клики для каждого года с 2010 по 2024.

Проведен качественный и количественный анализ изменений размеров клик и их характеристик.

Сделана интерпретация результатов с учетом экономических циклов и кризисных периодов.

Заключение:

Данный проект успешно демонстрирует применение графовых моделей и алгоритмов оптимизации для анализа динамики корреляций на рынке акций S&P 500. Использование библиотеки `networkx` позволило эффективно построить и проанализировать графы корреляций, а алгоритм поиска максимальной клики выделил группы наиболее взаимосвязанных активов. Визуализация результатов с помощью `d3.js` в React-приложении, развернутом на `Vercel`, обеспечивает наглядное представление динамики изменений в структуре рынка. Проведенный анализ позволил выявить периоды высокой и низкой корреляции, а также связать эти изменения с экономическими циклами и кризисными явлениями, что подчеркивает практическую ценность разработанного подхода для финансового анализа и управления рисками.