

УДК: 519.862.2

Б.И. Гольденгорин, А.А. Кочетуrow

**АНАЛИЗ ФИНАНСОВЫХ РЫНКОВ СРЕДСТВАМИ МОДЕЛИ
О Р-МЕДИАНАХ**

НИУ Высшая школа экономики

Рассматривается новый метод анализа финансовых рынков. На основе цен акций, торгуемых на финансовых рынках, строится матрица расстояний между доходностями акций. Матрица расстояний служит основой для построения различных сетевых структур таких, как графы рынка, минимальные остовные деревья, клики, независимые множества и др. В работе рассматривается другая структура, которая получается при решении задачи о Р-Медиане. Данная структура позволяет найти стабильную кластеризацию на рынке. Кластеры образуются в результате решения дискретной задачи оптимизации, а также группами акций, суммарное расстояния в которых до центров минимально. Сравнение с различными структурами позволяет говорить о том, насколько сильны или слабы те или иные кластеры. Отмечены особенности полученных структур для различных фондовых рынков.

Ключевые слова: финансовые рынки, кластеризация, минимальное остовое дерево, задача о Р-Медиане.

Введение

Целью данной работы является нахождение устойчивых характеристик и закономерностей в структуре финансовых рынков. Для рассмотрения мы выбрали три рынка: рынок финансовых инструментов США, как самый крупный на планете; рынок России, в качестве представителя развивающихся экономик; и рынок Швеции, в качестве представителя стабильных экономик с большим показателем ВВП на душу населения.

В работе Мантегна (1999) предложено использовать сетевую структуру Минимального Остовного Дерева для анализа индекса Доу–Джонса. За основу берутся цены закрытия акций, входящих в индекс в 1999 году. Далее ищутся корреляций между доходностями – акций и на их основе рассчитываются расстояния между акциями. Если представить акции, как вершины, между которыми есть ребра с весами, равными расстояниям, то получается полный взвешенный граф. На этой структуре далее ищется Минимальное Остовное Дерево, которое используется для конечного анализа. Эта работа послужила основой для многих других исследователей, которые применяют похожие или отличные структуры для анализа финансовых рынков (Kullmann et al. 2000; Onella et al. 2002, 2003 a,b; Cukur, 2007; Kenett, 2010 a,b; Boginski et al. 2005,2006).

Последнее время большое внимание уделяется анализу рынков конкретных стран и/или выявлению закономерностей на этих рынках в периоды кризиса, роста или спада экономики (Jung et al., 2006; Huang et al., 2009; Tabak et al. 2010).

В данной работе проводятся исследования по сложившейся схеме, когда сначала необходимо найти меру близости между всеми объектами, которой может быть матрица корреляций или расстояний между акциями. Далее на основе этой меры строится (иногда используя некоторый порог) граф связей между акциями, и на нем уже ищется определенная структура (Минимальное Остовное Дерево, клики и прочее). Потом проводится анализ полученных структур и сравнение с другими подходами. Делаются выводы.

Данные. Для анализа используются цены закрытия в каждый торговый день по всем акциям, торгуемым на определенном рынке. Мы рассматривали 1000 торговых дней, что включает в себя более четырех календарных лет: с 3 сентября 2007 года по 15 сентября 2011 года. Так как многие акции часто не торгуются, то необходимо восполнять пропуски в данных, чтобы в дальнейшем посчитать корреляции. Если в конкретный торговый день данных по цене закрытия нет, то используется цена закрытия предыдущего дня. Если пропусков больше 20%, то такая акция из рассмотрения удаляется. Из-за такой фильтрации на россий-

ском рынке осталась всего 151 акция, на рынке Швеции – 266 акций, на рынке США – 3378 финансовых инструментов, среди которых большой процент рыночных фондов, часто отражающих базовые финансовые индексы (DJIA, S&P500, S&P100 и прочие) и являющиеся взвешенной суммой их составляющих акций.

Другая необходимая информация – это принадлежность акции к конкретной компании и то, в какой отрасли и в каком секторе эта компания работает. Как и многие другие авторы, мы полагаем, что вся необходимая информация о поведении акций содержится в данных, описанных ранее.

Нахождение матрицы корреляций. Мы предполагаем, что из восстановленных временных рядов можно найти парные корреляции между доходностями акций, и найденная матрица корреляций наиболее полно описывает рынок и взаимодействие между акциями.

По формуле

$$\rho_{ij} = \frac{E\{(R_i - E\{R_i\})(R_j - E\{R_j\})\}}{\sqrt{\text{var}(R_i)\text{var}(R_j)}} \quad (1)$$

мы находим все парные коэффициенты корреляции между доходностями акций i и j , где R_i – временной ряд, описывающий доходность акции i , который можно получить из цен закрытия той же акции по формуле $R_i(t) = \ln \frac{P_i(t)}{P_i(t-1)}$, где $P_i(t)$ – цена закрытия акции i в день t .

Теперь найденную матрицу корреляций мы будем использовать в качестве меры близости акций.

Задача о P-Медиане (P-Median Problem, PMP). Следующим шагом анализа является выбор структуры, которая будет вычисляться на полученной матрице корреляций. Идея для выбора структуры такова: необходимо разбить весь рынок на определенное количество компонент, или кластеров, внутри которых акции сильнее всего коррелируют с центральной акцией, или медианой. При фиксированном количестве медиан или кластеров получается формулировка задачи о P-Медиане.

Дадим математическую формулировку данной задачи в терминах финансовых рынков: Пусть X – множество из n акций, торгующихся на рынке, а $\rho = [\rho_{ij}]_{n \times n}$ – матрица коэффициентов корреляций между акциями. Под $d(i, S)$ будем понимать расстояние от акции i до множества акций $S \subseteq X$, равное максимальной корреляции между акцией i и акциями из S , то есть $d(i, S) = \max_{j \in S} (\rho_{ij})$. Теперь потребуем найти множество S из p акций такое, что оно максимизирует суммарное расстояние от всех акций до этого множества. Иными словами, решим задачу:

$$\max_{S \subseteq X, |S|=p} \left(\sum_{i=1}^n d(i, S) \right) = \max_{S \subseteq X, |S|=p} \left(\sum_{i=1}^n \max_{j \in S} (\rho_{ij}) \right). \quad (2)$$

Данная задача является частным случаем задачи о P-Медиане. Классические методы её решения работают с требованием минимальности объектной функции и неотрицательности матрицы издержек (в терминологии финансовых рынков – матрицы корреляций). Детальный обзор методов решения и областей применения задачи можно найти в следующих работах: Reese, 2006; Mladenovic [et al.], 2007. Данных требований можно легко добиться, если из матрицы корреляций получить вспомогательную матрицу $C = E - \rho$, где E – единичная матрица размерности n . Задача (2) сводится к следующей задаче минимизации:

$$\min_{S \subseteq X, |S|=p} \left(\sum_{i=1}^n \min_{j \in S} (c_{ij}) \right). \quad (3)$$

Задача о Р-Медиане принадлежит к классу полимиально - неразрешимых задач. Что означает, что при современных вычислительных мощностях невозможно точно решить данную задачу для рынка США, где приходится использовать эвристические методы, которые в силу специфики данных позволяют найти очень точные решения.

Решив задачу (3), мы получаем множество S медиан и соответствующую им кластеризацию на $P = |S|$ компонент. Например, для рынка России из 151 акции для $p = 10$ получаем соответствующую кластеризацию.

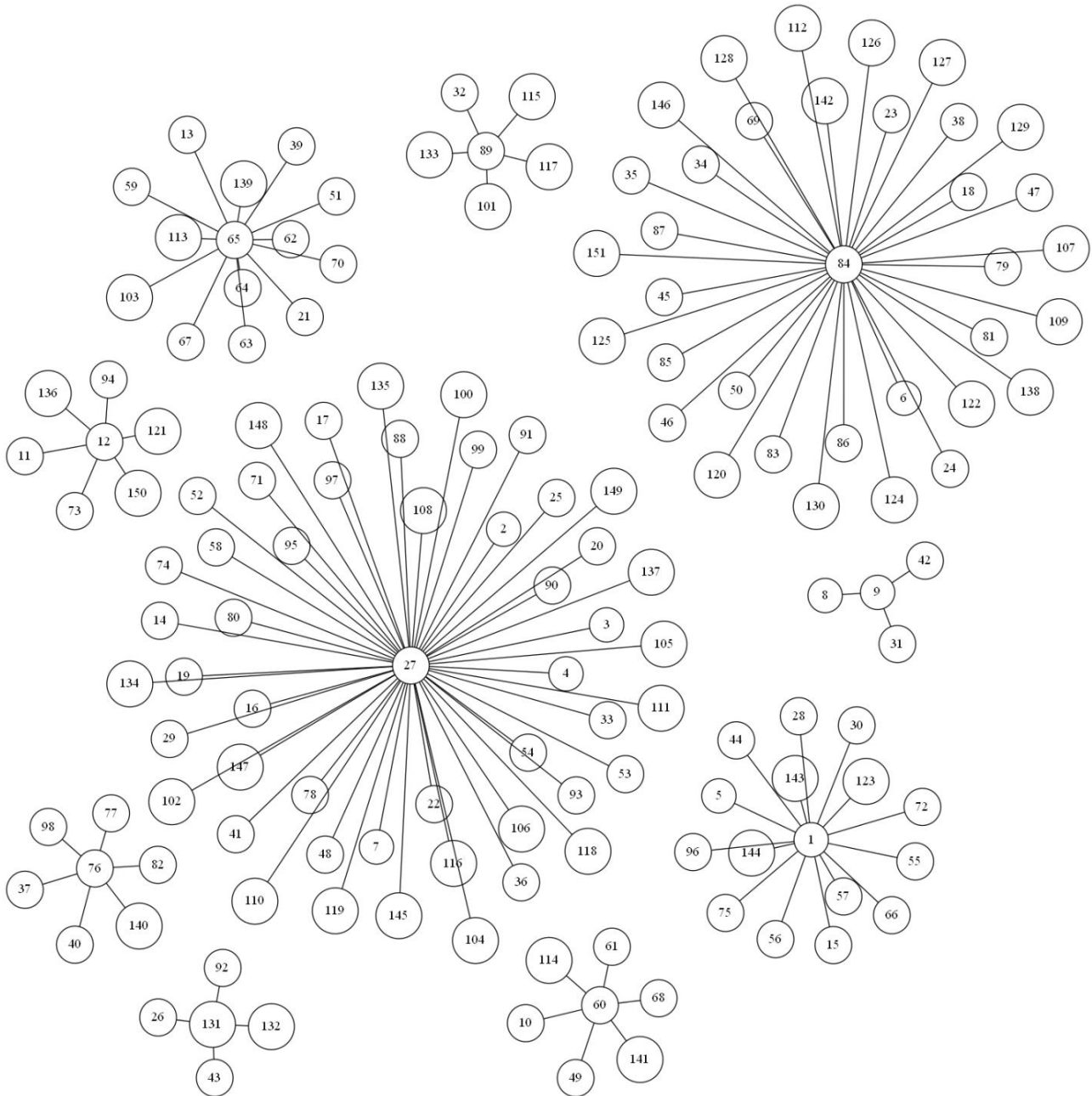


Рис. 1. Кластеризация для российского рынка при $p = 10$

Анализ кластеризации. Для начала заметим, что для рынков акций, в частности, и для любой матрицы корреляций в целом справедливо следующее утверждение:

$$f(p) = \max_{S \subseteq X, |S|=p} \left(\sum_{i=1}^n \max_{j \in S} (\rho_{ij}) \right) - \text{монотонно-неубывающая функция } \forall p = 1..n.$$

Действительно, допустим известно $f(k)$. Данному значению соответствует разбиение рынка на k кластеров. Теперь возьмем любой кластер, состоящий не из одной медианы (далее будем называть такие кластеры *нетривиальными*), выделим любую из акций (можно взять и саму медиану) в тривиальный отдельный кластер и сделаем её медианой этого кластера. Внутри полученного кластера суммарная корреляция равна 1, так как медиана полностью коррелирует с самой собой. А суммарная корреляция в кластере, из которого мы убрали акцию, уменьшилась на величину одного коэффициента корреляции (пусть это будет число a), не превосходящего единицу, так как $-1 \leq \rho_{ij} \leq 1, \forall i, j = 1..n$. Полная суммарная корреляция по всем кластерам изменилась на величину $1 - a \geq 0$, т.е. не уменьшилась. Так как такое разбиение на $k + 1$ кластеров является допустимым, то значение суммы корреляции, ему соответствующее, явно не больше максимума по всем возможным разбиениям, т.е. $f(k + 1)$.

Так же заметим, что максимум функции $f(p)$, достигающийся при $p = n$, равен n : все кластеры тривиальны, их n штук и корреляции внутри равны 1. Если положить за основу качества кластеризации значения функции $f(p)$, то при $p = n$ получим самую качественную кластеризацию, что не несет в себе никакой информации. Локальных экстремумов у данной функции тоже нет, поэтому необходим другой интуитивно-понятный метод нахождения оптимального количества кластеров: найдем такое $p = k$, при котором ещё нет тривиальных кластеров, а при $p = k + 1$ уже есть. Будем называть такое число кластеров «числом стабильной кластеризации», а соответствующее разбиение на компоненты – «стабильной кластеризацией».

На всех трех рынках наблюдаются следующие закономерности: во-первых, если $p = k$ – стабильное число кластеризации, то при всех $p > k$ есть тривиальные кластеры. Во-вторых, при стабильной кластеризации кластеры достаточно однородны: есть несколько самых крупных центральных кластеров, несколько среднего размера и остальные – небольшие, но относительно равные по размеру. При делении на большее количество компонент некоторые из средних или небольших по размеру (но никогда не самые крупные) начинают распадаться на отдельные тривиальные кластеры. Поэтому мы считаем, что стабильные числа кластеризации отражают действительную кластеризацию рынка. Для всех трех рынков числа стабильной кластеризации представлены в табл. 1.

Таблица 1

Страна	Число стабильной кластеризации
Россия	15
США	31
Швеция	12

Сравнение структуры кластеризации с другими структурами. На той же самой матрице коэффициентов корреляций Boginski [et al.] предложили искать максимальные по размеру клики, т.е. полные подграфы с максимальным числом вершин. Для этого необходимо сначала сделать усечение данных по порогу $\theta \in [-1; 1]$, который преобразует матрицу корреляций в матрицу $A = [a_{ij}]_{n \times n}$ из нулей и единиц, где $a_{ij} = \begin{cases} 1, \rho_{ij} \geq \theta, \\ 0, \rho_{ij} < \theta \end{cases}$. Таким образом, мы получаем граф связей между акциями, а далее действуем по следующей процедуре: мы находим клику наибольшего размера и удаляем её вершины из графа, потом выделяем следующую максимальную клику и снова удаляем, и так действуем до тех пор, пока не останутся клики максимального размера, равного 2. Все выделенные ранее клики мы запоминаем и сравниваем их со стабильной кластеризацией. Данную процедуру мы делаем для всех поро-

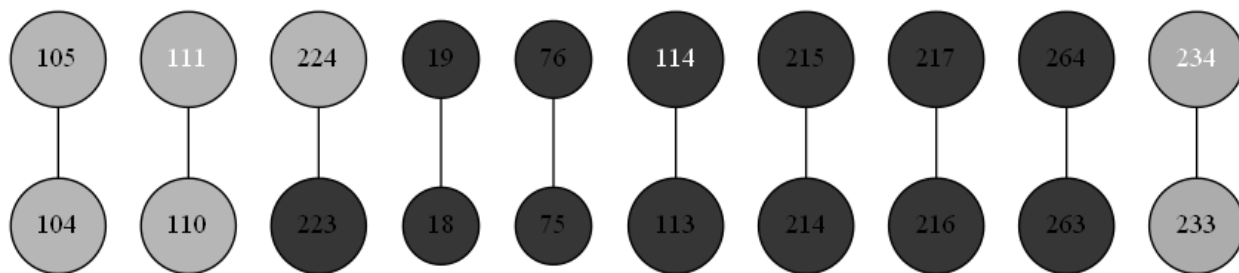


Рис. 3. МОЛ для рынка Швеции



Рис. 4. МОЛ для рынка США

На этих рисунках одинаковым цветом показаны акции, принадлежащие к одинаковым кластерам стабильной кластеризации. Если номер акции белого цвета, то эта акция – медиана соответствующего кластера. Для рынка России, где число стабильной кластеризации 15, а компонент в МОЛ – 7, видно, что каждая компонента соответствует отдельному кластеру, причем медиана этого кластера обязательно включена в эту компоненту и является её «центром» (из географических соображений), что говорит в пользу выбора критерия стабильной кластеризации. Так же некоторые из средних по размеру кластеров не включены в МОЛ, что означает, что эти кластеры не устойчивы. Для рынка Швеции (10 компонент МОЛ, 12 кластеров) мы наблюдаем похожую ситуацию: каждая компонента МОЛ соответствует одному кластеру, но один кластер может соответствовать нескольким компонентам. Также в силу того, что эти компоненты очень малы можно сделать вывод, что рынок Швеции более независим и однороден по сравнению с рынком России. Для рынка США (35 компонент МОЛ, 31 кластер) ситуация более сложная, но всё равно можно провести аналогию, что более крупным компонентам связанности соответствуют более крупные медианы.

Выводы

В работе введен критерий, указывающий на оптимальное количество кластеров на рынке. Сравнения с другими сетевыми структурами, показали, что стабильная кластеризация хорошо описывает рынок. Также с помощью сравнения различных сетевых структур можно делать вывод о том, насколько сильны или слабы те или иные кластеры и говорить, насколько независим рынок в целом.

Библиографический список

1. **Mantegna, R. N.** Hierarchical structure in financial markets // *European Physical Journal B*, 1999. № 11. P. 193–197.
2. **Kullmann, J L.** Mantegna, «Identification of clusters of companies in stock indices via Potts superparamagnetic transactions / L.J. Kullmann, R. N. Kertesz // *Physica A*. 2000. № 287. P. 412–419.
3. **Onnela, J.-P.** Asset Trees and Asset Graphs in Financial Markets / J.-P. Onnela [et al.] // *Physica Scripta*. 2003. № T106. P. 48–54.
4. **Onnela, J.-P.** Dynamic asset trees and Black Monday / J.-P. Onnela [et al.] // *Physica A*. 2003. № 324. P. 247–252.
5. **Onnela, J.-P.** Dynamic asset trees and portfolio analysis / J.-P. Onnela [et al.] *The European Physical Journal B*. 2002. № 30. P. 285–288.
6. **Cukur, S.** Cross correlations in an emerging market financial data / S. Cukur, M. Eryigit, R. Eryigit, *Physica A*. 2007. № 376. P. 555–564.
7. **Kenett, D. Y.** Dominating Clasp of the Financial Sector Revealed by Partial Correlation Analysis of the Stock Market / D. Y. Kenett [et al.] // *PLoS ONE*, 2010. T. 5. № 12.
8. **Kenett, D. Y.** Dynamics of Stock Market Correlations / D. Y. Kenett [et al.] // *AUCO Czech Economic Review*. 2010. № 4. P. 330–340,
9. **Boginski, V.** Mining market data: A network approach / V. Boginski, S. Butenko, P. M. Pardalos // *Computers & Operations Research*. 2006. № 33. P. 3171–3184.
10. **Boginski, V.** Statistical analysis of financial networks / V. Boginski, S. Butenko, P. M. Pardalos *Computational Statistics & Data Analysis*. 2005. № 48. P. 431–443.
11. **Jung, W.-S.** Characteristics of the Korean stock market correlations / W.-S. Jung [et al.] // *Physica A*. 2006. № 361. P. 263–271.
12. **Huang, W.-Q.** A network analysis of the Chinese stock market / W.-Q. Huang, X.-T. Zhuang, S. Yao *Physica A*. 2009. № 388. P. 2956–2964.
13. **Tabak, B. M.** Topological properties of stock market networks: The case of Brazil / B. M. Tabak, T. R. Serra и D. O. Cajueiro // *Physica A*. 2010. № 389. P. 3240–3249.

14. **Reese, J.** Solution methods for the p-Median problem: an annotated bibliography // Networks. 2006. Т. 3. № 48. P. 125–142.
15. Mladenovic, N. The p-Median problem: a survey of metaheuristic approaches / N. Mladenovic [et al.] // European Journal of Operational Research. 2007. № 179. P. 927–939.

*Дата поступления
в редакцию 12.11.2013*

B.I. Goldtngorin, A.A. Kocheturov

**THE ANALYSIS OF THE FINANSIAL MARKETS BY MEANS
OF THE R–MEDIAN MODEL**

Nizhny Novgorod state technical university n.a. R.E. Alexeev

Purpose: In this paper we consider a new approach of analyzing financial markets.

Design/methodology/approach: We build a distance matrix between stock returns based on prices of these stocks. The matrix is an initial point to build different network structures such as Market Graphs, Minimum Spanning Trees, Cliques, Independent Sets, etc.

Findings: We consider another structure obtained by means of the P–Median Problem.

Research limitation/ implications: The structure allows finding a stable clustering of a market. The clusters are obtained by the solution of discrete optimization problem and are formed by a groups of stocks with minimal total sum of distances from elements of groups to their centers.

Originality/value: Comparison between different structures allows estimation the strength of a certain cluster. Peculiarity of constructed structures for different stock markets are noted.

Key words: topographical analysis, traffic collision, TRC, traffic accident, TRA, GIS, Geographic Information System, an interactive map.