Market Graph Model

Valery Kalyagin

Laboratory of Algorithms and Technologies for Network Analysis
National Research University Higher School of Economics,
Nizhny Novgorod, Russia
Team of the market network analysis

vkalyagin@hse.ru

Seminar-workshop CSTLSN

Nizhny Novgorod, Nov 04-07

Overview

Network Model for a Stock Market.

- Each stock corresponds to a vertex in the network
- Link between two stocks (vertices) is represented by a weighted edge, where the weight reflects degree of similarity between stocks.
- Market network is a complete weighted graph.
- Mining market data: filter the information in complete weighted graph in order to extract the most valuable information.
- Common measure of similarity: Pearson correlation

Network structures

- **Maximum Spanning Tree (MST)**. MST in market network analysis was introduced in [Mantegna R.N. Europ. Physic. J. 1999]. More than 650 papers have used this tool for market analysis.
- **Planar Maximally Filtered Graph (PMFG)**. PMFG in market network analysis was introduced in [Tumminello M. at al. PNAS, 2005]. Kruskal type algorithm. More than 50 papers have used this tool for market analysis.
- **Market graph (MG). Maximum cliques (MC). Maximum Independent Sets (MIS).** A market graph is obtained from the market network by removing all edges with weights less than a specified threshold (threshold method) Market graph was introduced in [Boginski V., Butenko S., Pardalos P. J. Comp. Stat. Data Analysis, 2005]. More than 150 papers have used MG for market network analysis.

Market network analysis. Some results.

- MST generates an hierarchical structure on the market [Mantegna and others 1999, 2005, 2009 ...]
- PMFG was used to extract cluster structure of the market [Won-Min Song, T. Di Matteo, Tomaso Aste 2014]
- US market graph has a power law property [Boginski V. Butenko S. Pardalos P. 2005].
- Dynamics of the US market graph reflects some globalization process [Boginski V. Butenko S. Pardalos P. 2006].
- Power law property is observed for different markets [LATNA 2013].
- Russian market is dominated by a stable group of closely related companies (maximum clique) which accumulate 97% of traiding volume [LATNA, 2013].
- Maximum independent sets of the market graph are suitable for portfolio optimization [LATNA 2013].

Example. US Stock Market.

We choose a set of 10 stocks on US market

1. A (Agilent Technologies Inc),
2. AA (Alcoa Inc),
3. AAP (Advance Auto Parts Inc),
4. AAPL (Apple Inc),
5. AAWW (Atlas Air Worldwide Holdings Inc),
6. ABAX (Abaxis Inc),
7. ABD (ACCO Brands Corp),
8. ABG (Asbury Automotive Group Inc),
9. ACWI (iShares MSCI ACWI Index Fund),
10. ADX (Adams Express Company).

Example. US Stock Market.

Matrix of pairwise correlations (similarity measure) between daily returns of a set of 10 stocks traded in the US stock market for one year period, starting from Nov 2010.

	1	2	3	4	5	6	7	8	9	10
1	1.00	0.72	0.46	0.48	0.62	0.53	0.62	0.62	0.77	0.79
2	0.72	1.00	0.43	0.59	0.63	0.57	0.66	0.62	0.85	0.86
3	0.46	0.43	1.00	0.34	0.34	0.27	0.40	0.40	0.46	0.48
4	0.48	0.59	0.34	1.00	0.45	0.44	0.46	0.49	0.64	0.66
5	0.62	0.63	0.34	0.45	1.00	0.56	0.59	0.53	0.71	0.71
6	0.53	0.57	0.27	0.44	0.56	1.00	0.49	0.47	0.64	0.62
7	0.62	0.66	0.40	0.46	0.59	0.49	1.00	0.60	0.71	0.71
8	0.62	0.62	0.40	0.49	0.53	0.47	0.60	1.00	0.68	0.67
9	0.77	0.85	0.46	0.64	0.71	0.64	0.71	0.68	1.00	0.95
10	0.79	0.86	0.48	0.66	0.71	0.62	0.71	0.67	0.95	1.00

Table: US stock market, set of 10 stocks.

Example. MST (minimum spanning tree).

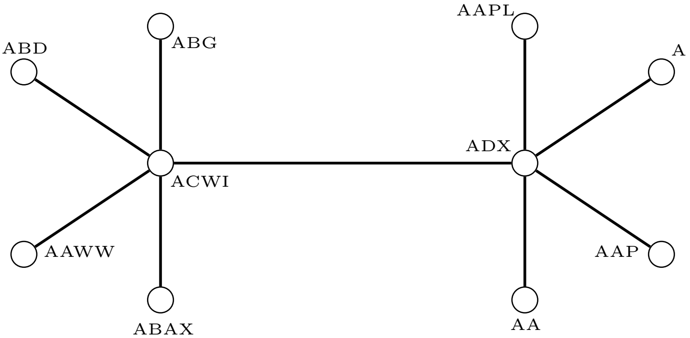


Figure: MST

Example. PMFG (planar maximally filtered graph).

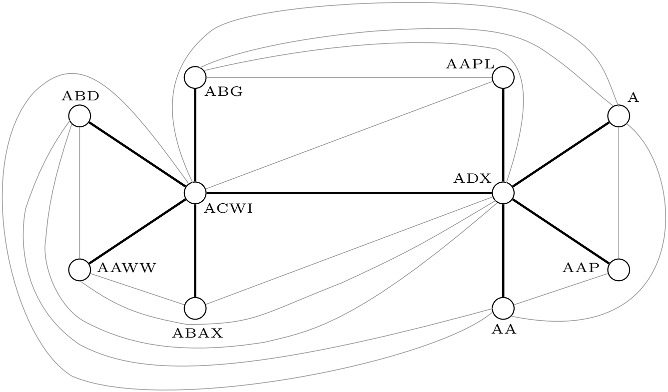
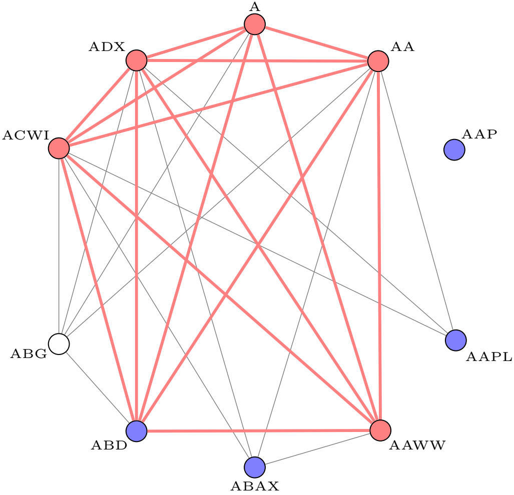


Figure: PMFG

Example. Market graph (MG), Maximum clique (MC) and maximum independent set (MIS).



Market network. What is it?

- Random graphs. Erdos-Renyi model.
- Random matrix theory (RMT). Wishart-Laguerre ensamble. Global regime.

Uncertainty problem

Problem:
How reliable are the results of market network analysis?
How to measure uncertainty of the obtained results?

New topic in market network analysis. A large new field of investigation.

Probabilistic model. Reference network.

- Let N is a number of stocks, n is a number of days of observations.
- Stock k return for day t is defined as

$$R_k(t) = \ln \frac{P_k(t)}{P_k(t-1)}, \tag{1}$$

where $P_k(t)$ is the price of stock k on day t.

- We assume that for fixed k, $R_k(t)$, $t = 1, \dots, n$, are independent random variables with the same distribution as R_k (i.i.d.) (Markovitz type model)
- Let $\rho_{i,j}$ be Pearson correlation between random variables R_i and R_j.
- We introduce *reference network* as the complete weighted graph with N nodes and weight matrix $||\rho_{ij}||$.
- For the reference network one can consider corresponding reference network structures, e.g. reference MST, reference PMFG, reference MG, MC, MIS and others.

Network structure identification problem.

Let $r_k(t), k = 1, \ldots, N, t = 1, \ldots, n$, be the observed values of returns.

Problem: identify the reference network structures from observations.
Possible Solution: let $\widehat{\rho_{i,j}}$ be an estimations of the Pearson correlations $\rho_{i,j}$ from observations. We introduce *sample network* as the complete weighted graph with N nodes and weight matrix $||\widehat{\rho_{i,j}}||$. For the sample network one can consider corresponding sample structures, e.g. sample MST, sample PMFG, sample MG, MC, MIS and others. We use the sample structures to identify the reference structures.

Multiple decision theory

Identification of a given structure (MST, PMFG, MG, MC, MIS and others) is equivalent to the selection of one particular subgraph from the finite family of possible ones. Any statistical procedure of identification is therefore a multiple decision statistical procedure.

One has L hypothesis $H_1, H_2, \ldots, H_L$ corresponding to the family of possible subgraphs associated with a given structure. Multiple decision statistical procedure $\delta(x)$ is a map from the sample space of observations
$R^{N\times n} = \{r_i(t) : \ i = 1, 2, \ldots, N; \ t = 1, 2, \ldots, n\}$
to the decision space $D = \{(d_1, d_2, \ldots, d_L)\}$, where d_j is the decision of acceptance of the hypothesis H_j, $j = 1, 2, \ldots, L$.

Example. Market graph. the problem of identification of the reference market graph for the threshold ρ_0 can be formulated as a multiple decision problem.
This is the problem of the selection of one from the set of hypotheses:

$$
\begin{aligned}
&H_1 : \rho_{i,j} < \rho_0, \forall (i,j),\ i<j,\\
&H_2 : \rho_{12} \geq \rho_0, \rho_{i,j} < \rho_0, \forall (i,j) \neq (1,2),\ i<j,\\
&H_3 : \rho_{12} \geq \rho_0, \rho_{13} \geq \rho_0, \rho_{i,j} \leq \rho_0, \forall (i,j) \neq (1,2), (i,j) \neq (1,3),\\
&\ldots\\
&H_L : \rho_{i,j} \geq \rho_0, \forall (i,j),\ i<j,
\end{aligned}
$$

where $L = 2^M$with $M = N(N-1)/2$. All together these hypotheses describe all possible reference market graphs.

Conditional risk

Quality of the multiple decision statistical procedure $\delta(x)$ according to Wald is measured by it's conditional risk (expected value of the loss function). There are many way to do it. For example for a given structure S one can define a conditional risk by

$$R(S,\delta) = \sum_{1 \le i < j \le N} [a_{ij} P^a_{i,j}(S,\delta) + b_{ij} P^b_{i,j}(S,\delta)],$$

where $a_{i,j}$ is the loss from erroneous inclusion of the edge (i,j) in the structure S, $P^a_{i,j}(S,\delta)$ is the probability that decision procedure δ takes this decision, $b_{i,j}$ is the loss from erroneous non inclusion of the edge (i,j) in the structure S, $P^b_{i,j}(S,\delta)$ is the probability that decision procedure δ takes this decision.

Two terms in the formula can be considered as type I and type II statistical errors [Lehmann Romano 2005].

Multiple decision theory. Optimality.

Lehmann theory of optimal multiple decision statistical procedures

1. Generating hypothesis
2. Compatibility conditions
3. Additivity of the loss function

Example. Market graph: generating hypothesis

$$h_{i,j}: \ \rho_{i,j} < \rho_0 \ \text{ against } \ k_{i,j}: \ \rho_{i,j} \geq \rho_0$$

Compatibility of generating hypothesis can be verified
Additivity of the loss function - additivity of error in market graph identification (Type I and Type II errors)
Result: Optimal multiple decision statistical procedures for the market graph construction based on the optimal two decisions tests.

Statistical uncertainty of different structures

Conditional risk can be used as a measure of statistical uncertainty.
Numerical experiments. Assumptions.

- Measure of similarity: Pearson correlation
- Distribution of random vector $R = (R_1, R_2, \dots, R_N)$: multivariate normal distribution. Correlation matrix is taken from real markets.
- Network structures: MST, PMFG, MG
- Multiple decision statistical procedure for network structure identification: sample MST, PMFG, MG.
- Conditional risk: total fraction of error

Numerical experiments.

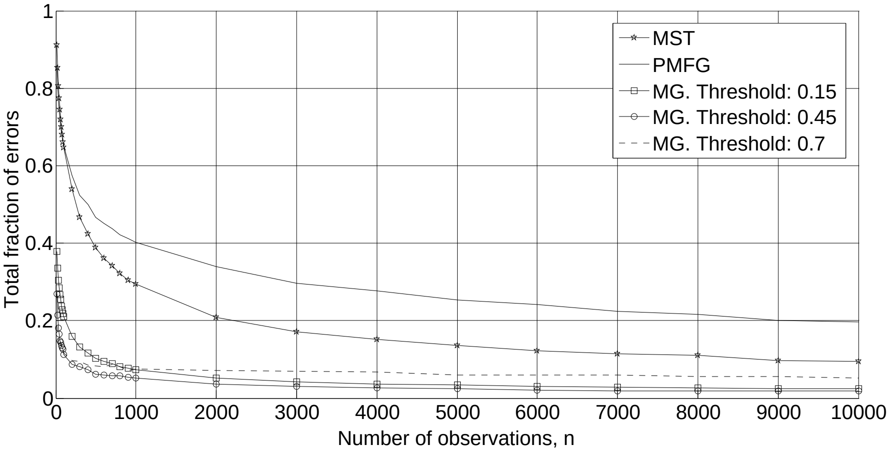


Figure: Model of Russian market. Conditional risk as function of number of observations for MST, PMFG, MG.

Robustness

Multivariate distribution of stocks returns is not known. Can we construct a robust statistical procedures for network structures identification?

Robust multiple decision statistical procedure - conditional risk does not depend on distributions from a large class

Popular class of distributions: elliptically contoured distributions (includes multivariate Gaussian and Student distributions)

How to construct a robust procedures? Use a different measure of similarity.

Sign correlation.

The sign correlation is defined as:

$$s_{ij} = E\left[sign\left((R_i - \mu_i)(R_j - \mu_j)\right)\right] \tag{2}$$

where $\mu_i = E(R_i)$, $\mu_j = E(R_j)$ and

$$sign(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$$

Sign correlation is connected to the probability p_{ij} of coincidence of the signs of $(R_i - \mu_i)$ and $(R_j - \mu_j)$. One has

$$s_{ij} = 2p_{ij} - 1$$

Sign correlation.

For the sign correlation, the natural estimator is the sample sign correlation (also known as Fechner correlation):

$$\widehat{s_{ij}} = \frac{1}{n}\sum_{t=1}^{n} sign(R_i(t) - \overline{R_i})sign(R_j(t) - \overline{R_j}) = \frac{e_{ij} - d_{ij}}{e_{ij} + d_{ij}} \tag{3}$$

Here e_{ij} denotes the number of pairs $\langle (R_i(t) - \overline{R_i}), (R_j(t) - \overline{R_j}) \rangle$ that have the same sign and d_{ij} denotes the number of such pairs that have different signs.

Robustness

Main result: conditional risk for sign correlation network does not depend on distribution from the class of elliptically contoured distributions. This is not true for Pearson correlation network.

Robustness

Illustration. Mixture distribution consisting of multivariate normal distribution and multivariate Student distribution.

$$R \sim \begin{cases} N(\mu, \Sigma), & \text{with probability } \gamma \\ t_3(\mu, \Sigma), & \text{with probability } 1 - \gamma \end{cases}$$

Note that the parameters of distributions are identical, so the resulting mean values will be μ, and covariance matrix Σ.
For $\gamma = 0$ one has multivariate Student distribution.
For $\gamma = 1$ one has multivariate Normal (Gaussian) distribution.

Robustness

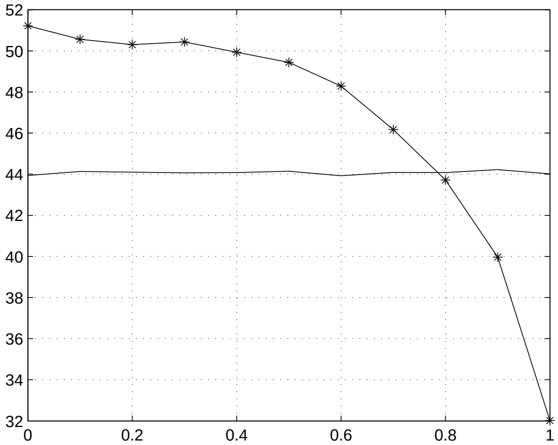


Figure: Maximum spanning tree. Number of observations 400. Conditional risk as function of γ for Pearson (star line) and sign (continuous line) correlations.

Conclusions

In the framework of proposed approach one can

- measure the quality of identification of network structures (conditional risk)
- design optimal statistical procedures for identification (multiple decision theory)
- measure uncertainty of the results of market network analysis
- construct robust statistical procedures of identification of network structures

Team.

Laboratory of Algorithms and Technologies for Network Analysis (LATNA),
National research University Higher School of Economics, Nizhny Novgorod, Russia
Scientific Head of the Laboratory: professor P.M.Pardalos, University of Florida, USA

Team: Grigory Bautin, Alexander Koldanov, Peter Koldanov, Anton Kocheturov, Panos Pardalos, , Viktor Zamaraev, master and bachelor students

LATNA publications on market network analysis.

- Koldanov A.P., Koldanov P.A., Kalyagin V.A., Pardalos P.M. Statistical Procedures for the Market Graph Construction, Computational Statistics & Data Analysis, v.68, pp.17-29 (2013).
- 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, Volume 10, Issue 2 (2013), Page 105-124.
- Bautin G.A., Kalyagin V.A., Koldanov A.P. Comparative analysis of two similarity measures for the market graph construction, Models, Algorithms and Technologies for Networks Analysis, Springer Proceedings on Mathematics and Statistics, v.59 (2013), pp.29-41.
- Koldanov A. P., Koldanov P.A: Optimal Multiple Decision Statistical Procedure for Inverse Covariance Matrix. Springer Optimization and Its Applications. Vol. 87 (2014), pp 205-216.

LATNA publications on market network analysis.

- Kalyagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P.M., Zamaraev V.A.: Measures of uncertainty in market network analysis, Physica A: Statistical Mechanics and its Applications, 2014. Vol. 413, pp. 59-70.
- Vizgunov A.N., Goldengorin B.I., Kalyagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P. M., Network approach for the Russian stock market, Computational Management Science, Springer-Verlag, Volume 11, Issue 1 (2014), Page 45-55.
- Kalyagin V.A., Koldanov P.A., Zamaraev V.A. Network structures uncertainty for different markets, Springer Optimization and Its Applications, v.100, pp.181-198 (2014)
- Koldanov P.A. Bautin G.A. Multiple decision problem for stock selection in market network, Lecture Notes in Computer Science, Vol. 8426, pp 98-110. Springer-Verlag Berlin Heidelberg (2014)

LATNA publications on market network analysis.

- Kalygin V.A. Koldanov A.P. and Pardalos P.M. A General Approach to Network Analysis of Statistical Data Sets, Lecture Notes in Computer Science, Vol. 8426, pp 88-97. Springer-Verlag Berlin Heidelberg (2014)
- Kalyagin V. Koldanov A. Koldanov P. Zamaraev V. Market Graph and Markovitz Model, Optimization in Science and Engineering, 2014, pp 293-306
- Bautin G.A., Koldanov A.P., Pardalos P.M. Robustness of sign correlation in market network analysis, Springer Optimization and Its Applications, v.100, pp.25-34 (2014)
- Kocheturov A.A. Batsyn M.V. Pardalos P.M. Dynamics of cluster structures in a financial market network, Physica A: Statistical Mechanics and its Applications, 2014. Vol. 413, pp. 523-533

THANK YOU FOR YOUR ATTENTION!