# Financial Market - A Network Perspective

Jukka-Pekka Onnela[1], Jari Saramäki[1], Kimmo Kaski[1], and János Kertész[2]

[1] Laboratory Computational Engineering, Helsinki University of Technology, P.O.Box 9203, FIN-02015 HUT. jonnela@lce.hut.fi
[2] Department of Theoretical Physics, Budapest University of Technology and Economics, Budafoki út 8, H-1111 Budapest, Hungary.

We construct a weighted financial network for a subset of NYSE traded stocks, in which the nodes correspond to stocks and edges to interactions between them. We identify clusters of stocks in the network, based on the Forbes business sector classification, and study their intensity and coherence. Our approach indicates to what extent the business sector classifications are visible in market prices, enabling us to gauge the extent of group-behaviour exhibited by stocks belonging to a given business sector.

## 1 Introduction

Complex networks provide a very general framework, based on the concepts of statistical physics, for studying systems with large numbers of interacting agents [1]. The nodes of the network represent the agents and a link connecting two nodes indicates an interaction between them. In the complex networks framework, interactions have typically been considered to be binary in nature, meaning that either two nodes interact (are connected) or they do not (are not connected). Imposing a binary interaction requires setting a threshold value for interaction strength, such that interactions falling below it are discarded. Although this approach is a suitable first approximation, thresholding can lead to a loss of information. Consequently, a natural step forward is to assign weights on the links to reflect the strengths of interactions.

In a financial market the performance of a company is compactly characterised by a single number, the stock price, which results from a large number of interactions between different market participants. Although the exact nature of these interactions is not known, they are certainly reflected in the equal-time return correlations. In this paper we study a financial network in which the nodes correspond to stocks and links to return correlation based interactions between them. Mantegna [2] was the first to construct such networks and the idea was followed and extended by others [3, 4, 5, 6, 7].

# 2 Methods

## 2.1 Constructing the Network

We start by considering a price time series for a set of $N$ stocks and denote the daily closing price of stock $i$ at time $\tau$ (an actual date) by $P_i(\tau)$. Since investors work in terms of relative as opposed to absolute returns, logarithmic returns are commonly used in studies, and thus we denote the daily logarithmic return of stock $i$ by $r_i(\tau) = \ln P_i(\tau) - \ln P_i(\tau - 1)$. We extract a time window of width $T$, measured in days and in this paper set to $T = 1000$ (equal to four years, assuming 250 trading days a year), and obtain a return vector $\boldsymbol{r}_i^t$ for stock $i$, where the superscript $t$ enumerates the time window under consideration. Then equal time correlation coefficients between assets $i$ and $j$ can be written as

$$\rho_{ij}^t = \frac{\langle \boldsymbol{r}_i^t \boldsymbol{r}_j^t \rangle - \langle \boldsymbol{r}_i^t \rangle \langle \boldsymbol{r}_j^t \rangle}{\sqrt{[\langle {\boldsymbol{r}_i^t}^2 \rangle - \langle \boldsymbol{r}_i^t \rangle^2][\langle {\boldsymbol{r}_j^t}^2 \rangle - \langle \boldsymbol{r}_j^t \rangle^2]}}, \tag{1}$$

where $\langle ... \rangle$ indicates a time average over the consecutive trading days included in the return vectors. These correlation coefficients between $N$ assets form a symmetric $N \times N$ correlation matrix $\mathbf{C}^t$ with elements $\rho_{ij}^t$. The different time windows are displaced by $\delta T$, where we have used a step size of one week, i.e. $\delta T = 5$ days.

Next we define interaction strengths, or link weights, based on the correlation coefficients. One of the simplest alternatives is to use the absolute values of the correlation coefficients, in which case the interaction strength reflects the strength of linear coupling between the logarithmic returns of stocks $i$ and $j$ in time window $t$. If we use $w_{ij}^t$ to denote the weight on the link connecting node $i$ and node $j$, with this choice we have $w_{ij}^t = |\rho_{ij}^t|$, or in matrix form $\mathbf{W}^t = |\mathbf{C}^t|$. Because the correlation coefficients $\rho_{ij}^t$ vary between $-1$ and 1, the interaction strengths $w_{ij}^t$ are naturally limited to the [0, 1] interval. In the correlation matrix $\mathbf{C}^t$ we have estimated the correlations between all the assets. Thus, the resulting network will be fully connected consisting of $N$ nodes and $N(N-1)/2$ links, corresponding to the elements in the upper (or lower) triangular part of the the weight matrix.[3]

## 2.2 Characterising Network Clusters

Let us now consider any cluster or subgraph $g$ in the above defined network. To characterise how compact or tight the subgraph is, we use the concept of subgraph *intensity* $I(g)$ introduced in [8]. Put differently, subgraph intensity allows us to characterise the interaction patterns within clusters. If we use $v_g$ to denote the set of nodes and $\ell_g$ the set of links in the subgraph with weights $w_{ij}$, we can express subgraph intensity as the *geometric mean* of its weights:

[3] It is possible, using some heuristic, to insert only a fraction of all the links in the network, but this would result in an additional parameter to be determined.

$$I(g) = \left( \prod_{(ij) \in \ \ell_g} w_{ij} \right)^{1/|\ell_g|} . \tag{2}$$

Due to the nature of the geometric mean, the subgraph intensity $I(g)$ may be low because one of the weights is very low, or it may result from all of the weights being low. In order to distinguish between these two extremes, we use the concept of subgraph *coherence* $Q(g)$ [8]. It assumes values from the interval [0, 1] and is close to unity only if the subgraph weights do not differ much, i.e. are internally coherent. Subgraph coherence is defined as the ratio of the geometric to the arithmetic mean of the weights as

$$Q(g) = I|\ell_g| / \sum_{(ij) \in \ \ell_g} w_{ij}. \tag{3}$$

In order to compare intensity and coherence values, we need to establish a reference. A very natural reference system is obtained by considering the entire market. In other words, we take all of the $N$ nodes and $N(N-1)/2$ links making up the network $G$, and then using the above definitions compute $I(G)$ and $Q(G)$. We can also use *relative cluster intensity* for cluster $g$, given by $I(g)/I(G)$, and *relative cluster coherence*, given by $Q(g)/Q(G)$, if instead of absolute values we wish to examine the cluster intensity or coherence relative to the reference system.

## 3 Results

In this section we consider a subset of 116 NYSE-traded stocks from the S&P 500 index from 1.1.1982 to 31.12.2000. We deal with the closing price, resulting in a total of 4787 price quotes for each stock. To divide the stocks into clusters, we obtained the Forbes business sector labels for each stock [9]. The stocks in our dataset fall into 12 business sectors, such as Energy and Utilities. Given these labels for each stock, we use the concepts of subgraph intensity and coherence to gauge how how similarly stocks belonging to a given business behave as a function of time.

Let us consider a cluster $g$, constructed such that all of its nodes $v_g$ belong to the same business sector, and let $n$ denote the number of nodes in this cluster. Then we add all the $n(n-1)/2$ links corresponding to the interaction strengths between any pair of nodes within $g$. In one extreme, if all the link weights are equal to unity, every node participating in $g$ interacts maximally with its $n-1$ neighbours. In the other extreme, if one or more of the weights are zero, the subgraph intensity for the *fully connected subgraph* $g_n$ tends to zero because the original topological structure no longer exists.

In Figure 1, we show the relative cluster intensity as a function of time for selected business sector clusters. Values above unity indicate that the intensity of the cluster is higher than that of the market. This implies that in most cases stocks belonging to a given business sector are tied together in the sense that intra-cluster interaction strengths are considerably stronger than those of the market on the whole.

It is also worth noting the high value for the absolute cluster intensity for the market roughly between 1986 and 1990. This elevated value is due to the 1987 stock market crash (Black Monday), which caused the market to behave in a unified manner[4]. The crash also compresses the relative cluster intensities, which means that the cluster-specific behaviour is temporarily suppressed by the crash, and after the market recovers the clusters regain their characteristic behaviour.

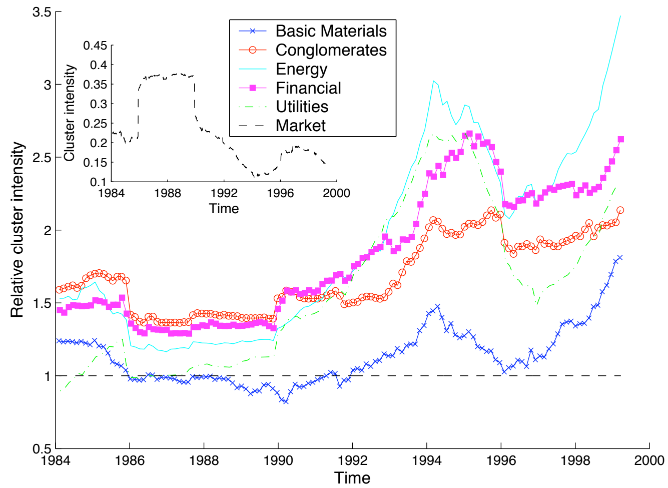


**Fig. 1.** Relative (to the market) cluster intensity as a function of time for select clusters. Inset: The (absolute) cluster intensity for the market used for normalisation.

Business sector clusters are also more coherent than the market, as shown in Figure 2, except for Basic Materials. One explanation is obtained from the industry classifications, which is a finer classification scheme, of stocks comprising the BM cluster. These include Metal Mining, Paper, Gold & Silver and Forestry & Wood Products. Therefore, it is clear that the Basic Materials business sector is extremely diverse. Also, the price of some of these items is determined, at least partially, outside the stock market. Consequently, it is not so surprising that the cluster intensity remains low, at times even falling below the market reference. Similarly, the low coherence values indicate that there are stocks in this cluster with very high correlations (those belonging to the same industry, such as gold mining), but also very low (companies belonging to different industries). In conclusion, our results indicate that, in most cases, stocks belonging to the same business sector have higher intensity and more coherent intra-cluster than inter-cluster interactions.

[4] The length of this elevated period is related to the window width parameter.

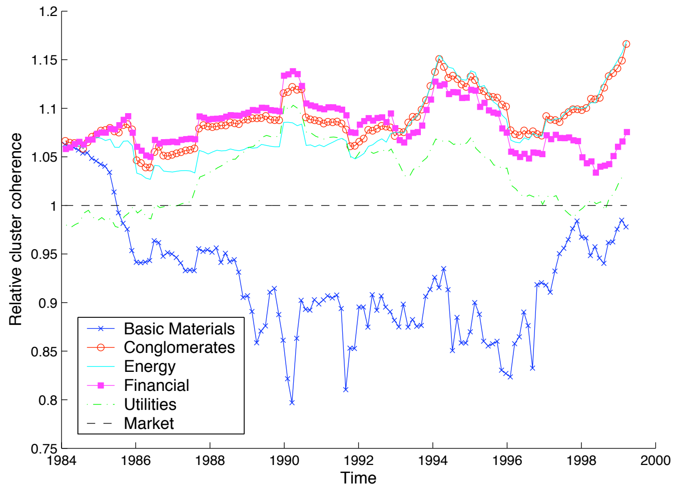


**Fig. 2.** Relative (to the market) cluster coherence as a function of time.

## References


1. Albert R, Barabasi A-L (2002) Statistical mechanics of complex networks. Reviews of Modern Physics 74, 47-97
2. Mantegna R N (1999) Hierarchical structure in financial markets. European Physical Journal B 11, 193-197
3. Vandewalle N, Brisbois F, Tordoir X (2001) Non-random topology of stock markets. Quantitative Finance 1, 372-374
4. Marsili M (2002) Dissecting financial markets: Sectors and states. Quantitative Finance 2, 297-302
5. Caldarelli G, Battiston S, Garlaschelli D, Catanzaro M (2004) In: Ben-Naim E, Frauenfelder H, Toroczkai Z (eds) Complex Networks. Springer
6. Onnela J-P, Chakraborti A, Kaski K, Kertesz J, Kanto A (2003) Dynamics of market correlations: Taxonomy and portfolio analysis. Physical Review E 68, 056110
7. Onnela J-P, Chakraborti A, Kaski K, Kertesz J, Kanto A (2003) Asset trees and asset graphs in financial markets. Physica Scripta T106, 48-54
8. Onnela J-P, Saramäki J, Kertész J, Kaski K Intensity and coherence of motifs in weighted complex networks. cond-mat/0408629
9. The website of Forbes at www.forbes.com