RESEARCH PAPER QUANTITATIVE FINANCE VOLUME 1 (2001) 372–374
quant.iop.org INSTITUTE OF PHYSICS PUBLISHING

Non-random topology of stock markets

N Vandewalle[1], F Brisbois[1] and X Tordoir[2]

[1] GRASP, Institut de Physique B5, Université de Liège, B-4000 Liège, Belgium
[2] IPNE, B15, Université de Liège, B-4000 Liège, Belgium

Received 26 September 2000

Abstract

We have analysed the cross correlations of daily fluctuations for $N = 6358$ US stock prices during the year 1999. From those $N(N-1)/2$ correlation coefficients, the minimum spanning tree (MST) has been built. We have investigated the topology exhibited by the MST. Even though the average coordination number of stocks is $\langle n \rangle \approx 2$, the variance σ of the topological distribution $f(n)$ diverges! More precisely, we have found that $f(n) \sim n^{-2.2}$ holds over two decades. We have studied the topological correlations for neighbouring nodes: an extremely broad set of local configurations exists, confirming the divergence of σ.

In a series of works [1, 2], Mantegna and co-workers have studied the minimum spanning tree (MST) of financial data, mainly the cross correlations between N different stocks. Building such a MST is a common method in spin glasses for putting into evidence the primary structure (the skeleton) of a complex system with a non-trivial dynamics [3]. The MST captures the cooperative behaviours behind the stock market.

Herein, we report some analysis of the topology exhibited by the MST[3]. Due to initial public offerings (IPOs) and mergers, it is difficult to perform some MST-analysis over long periods. The number of companies is indeed always evolving. We focused our work mainly on N different US companies during the year 1999. Those stocks are all traded on the Nasdaq, NYSE and AMEX places. In fact, our database contains about 9000 different US stocks. Some of them are illiquid, i.e. not traded every day. Those illiquid stocks were not considered. We have selected $N = 6358$ different companies in our study.

Cross correlation ρ_{ij} for the pair ij of stocks is computed as

$$\rho_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2)(\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}} \qquad (1)$$

where Y_i is the daily fluctuation of the log of the stock i price.

[3] A Belgian company, Market Topology SPRL, provides complementary information and updated data about this work. See their website http://www.market-topology.com.

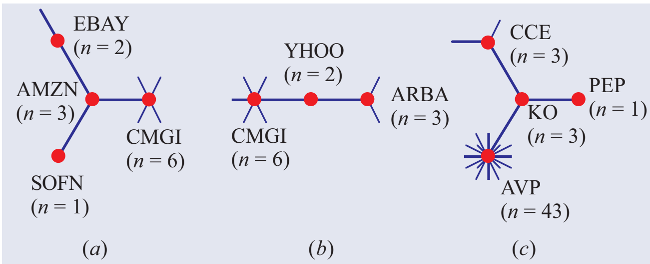

Figure 1. Three typical configurations of the nearest neighbourhood of three companies on the MST: (*a*) AMZN (Amazon.com Inc.) connected to EBAY (eBay Inc.), CMGI (CMGI Inc.) and SOFN (SoftNet Systems Inc.); (*b*) YHOO (Yahoo! Inc.) is connected to ARBA (Ariba Inc.) and CMGI; (*c*) KO (Coca-Cola Company) is connected to PEP (PepsiCo, Inc.), CCE (Coca-Cola Enterprises, Inc.) and AVP (Avon Products Inc.). The topological indices n are given.

The MST has been built following Kruskal's algorithm [5] in order to find the $N - 1$ most important correlated pairs of stocks among the $N(N-1)/2 = 20\,208\,903$ possible pairs! The Kruskal algorithm assures the unicity of the structure and avoids loops. A tree is thus formed. Figure 1 presents three parts of the MST: AMZN (Amazon.com Inc.), YHOO (Yahoo! Inc.), KO (Coca-Cola Company) and their nearest neighbours. AMZN is connected to EBAY (eBay Inc.), CMGI (CMGI Inc.) and SOFN (SoftNet Systems Inc.). YHOO is connected to ARBA (Ariba Inc.) and also connected to CMGI

1469-7688/01/030372+03$30.00 PII: S1469-7688(01)17459-2

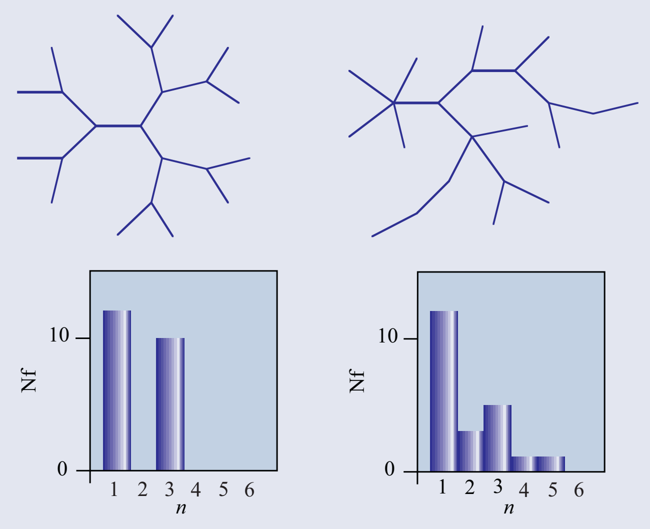

Figure 2. Two different trees and their respective topological distributions $f(n)$. Each tree is made of 22 sites and 21 links. (*a*) Hierarchical tree with a constant branching ratio fixed to 3. (*b*) Random tree built by gluing successive links on randomly chosen extremitites.

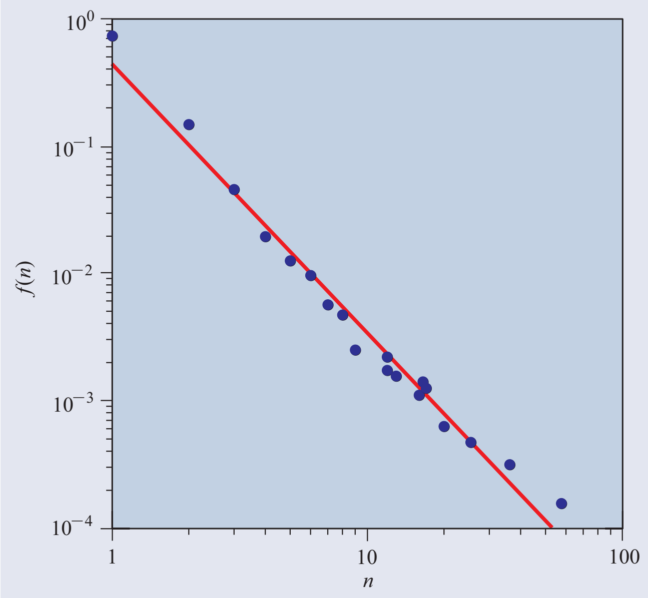

Figure 3. Log–log plot of the topological distribution $f(n)$ of the MST. The continuous line is a fit with the power law (7).

such that YHOO is the next nearest neighbour of AMZN. KO is connected to PEP (PepsiCo, Inc.), CCE (Coca-Cola Enterprises, Inc.) and AVP (Avon Products Inc.). In addition to KO, AVP is connected to 42 other stocks!

One can observe a clustering of companies being in the same sector such as energy, technology, internet, transportation, food and cosmetics, etc. The clustering in various sectors, each being a branch of the MST, has been previously observed in [1]. In addition to the sectorization, we have observed a global clustering of stocks traded on different market places. Most Nasdaq stocks are located around the QQQ ticker (Nasdaq Composite ticker), and largest stocks of NYSE are located around DIA (Dow Jones ticker).

One can distinguish three types of topological configurations for the N companies: (i) important nodes, (ii) links and (iii) dangling ends. The important nodes (like AVP) should correspond to companies controlling or mediating the daily fluctuations of their neighbourhood. The links (like YHOO) are mediating the information (fluctuation) along the branches. The dangling ends (like PEP) are supposed to be the less influent companies, or to be those less influenced by the market fluctuations. The latter companies do not necessary follow the indices.

From the mathematical point of view, we can associate a 'coordination number' $1 \leqslant n \leqslant (N-1)$ at each site of the tree. The coordination number n is defined as the number of nearest neighbours of the considered site. The statistics of the tree is thus determined by the statistical distribution $f(n)$ of coordination numbers. In order to illustrate the parameter n, figure 2 shows two different trees, each being made of 22 sites and 21 links. The first structure is a hierarchical (Cayley) tree with a constant branching ratio except for dangling ends. The branching ratio has been set to 3. The f distribution is also illustrated: only $n = 1$ and $n = 3$ sites are observed. The second tree is a random tree, i.e. a tree built by gluing successive links on randomly chosen extremitites. This stochastic aggregation process for generating the tree produces a distribution of n values ranging from 1 to 5 in this example. One expects that the MST of the US stockmarket looks similar to a random tree. This hypothesis will be tested below.

The mean coordination number is given by

$$\langle n \rangle = \sum_{n=1}^{\infty} n f(n). \tag{2}$$

If one multiplies both sides of this equation by N, the summation becomes equivalent to counting twice all the $N-1$ connections along the tree. As a consequence, one has

$$\langle n \rangle = 2\frac{N-1}{N} \tag{3}$$

which is approximately equal to 2 when the number of sites N becomes large, i.e. in our case. The above mathematical property is quite general and holds for all types of trees. In the following, we will consider that $\langle n \rangle \approx 2$.

The first moment $\langle n \rangle$ of f cannot give any information about the topology of trees since it does not reflect the shape of the tree. Both trees in figure 2 have a mean topological number $\langle n \rangle = \frac{21}{11}$! However, the second moment σ of the f distribution can be quite useful. Let us assume a topological correlation in between neighbouring nodes. The Aboav law, a well-known law in the physics of random cellular structures [4], stipulates that topological correlations can be captured by the following relationship

$$m_n = A + \frac{B}{n} \tag{4}$$

where m_n is the mean coordination number of the nearest neighbours of a node characterized by a coordination number n. A and B are positive constants to be determined. The form of the Aboav law implies that important nodes with large n are

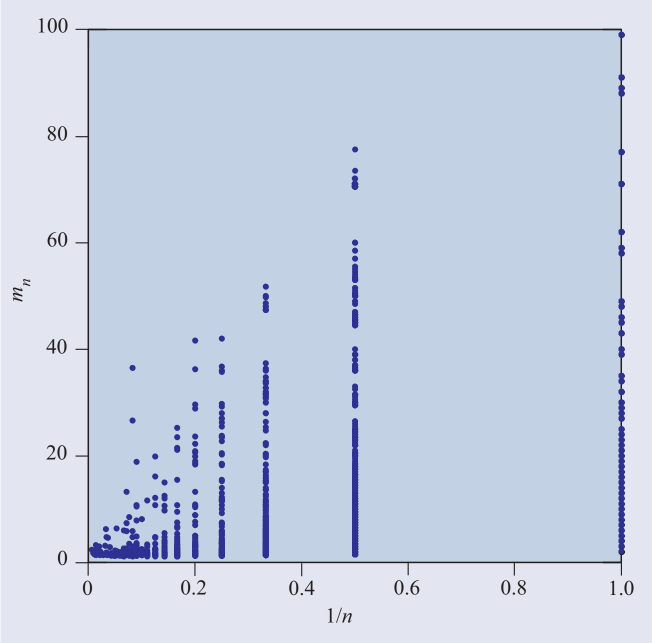

Figure 4. The $(m_n, \frac{1}{n})$ diagram for the $N = 6358$ companies. No Aboav (linear) behaviour emerges.

more favourably connected to dangling ends. Indeed, when the coordination number becomes large and reaches $n \to N - 1$, which correspond to a central node connected to $N-1$ dangling ends, one should obtain $m_n = 1$. This leads to the constraint[4] $A = 1$. For a hierarchical tree with a constant branching ratio n, one should have the condition $m_n = n$, i.e. no correlation ($B = 0$). That case is of course unrealistic for the present study. For a random tree, one expects that the statistical average of the product nm_n is similar to $\langle n^2 \rangle$, one has

$$\langle n \rangle + B = \langle nm_n \rangle \approx \langle n^2 \rangle = \sigma^2 + 4 \tag{5}$$

and thus $B = \sigma^2 + 2$ and the Aboav law becomes

$$m_n = 1 + \frac{\sigma^2 + 2}{n} \tag{6}$$

for any kind of tree with a large number N of nodes. The above equation is only valid in the hypothesis of a random topology.

We have defined some mathematical tools ($f(n)$, σ and the Aboav law) for describing the topology of trees in full generality. Let us investigate the specific topology of financial MST. The log–log plot of the topological distribution is shown in figure 3. The distribution $f(n)$ is broad and looks like a power law

$$f(n) \sim n^{-\alpha} \tag{7}$$

holding over two decades with an exponent $\alpha = 2.2 \pm 0.1$. This result means that the variance of the $f(n)$ distribution diverges even when the mean $\langle n \rangle$ remains finite (see below)! In other words, nodes with high coordination numbers are not so rare. For random trees, one expects an exponential decay of $f(n)$ and quite different statistics. The market seems to be self-organized in a coordination invariant structure, similarly to self-organized critical structures [6]! One should also remark that our results have an important graphical consequence: large n values for nodes are not so rare such that it is extremely hard to draw the entire MST.

The power law (7) and the divergence of σ are not numerical artefacts. Indeed, we have checked the Aboav-like law for which the variance of the topological distribution is relevant (see above). Figure 4 presents the $(m_n, \frac{1}{n})$ diagram in which the Aboav law should be a line with a slope $\sigma^2 + 2$. However, the N values of the measured m_n on the MST are sparsely distributed in the diagram such that no relevant linear fit can be performed. Equation (6) is not valid expressing that statistical averages cannot be taken as for usual random topological structures. We argue that the extremely broad dispersion of local configurations in $(m_n, \frac{1}{n})$ expresses the divergence of σ.

In summary, we put into evidence the emergence of some order out of the apparent disorder of the stock markets! Indeed, we have found signatures of non-trivial correlations in the topology of stock markets. We did the same analysis for subsets of financial data and we obtained the same behaviour (a power law for $f(n)$ and a divergence of σ). For shifted periods, the main results do not change: the structure evolves slowly and locally. Indeed, month after month, a node keeps the majority of its neighbours[5]. The non-randomness of the stock market topology is thus a robust property.

Acknowledgments

NV thanks the FNRS for financial support. Valuable discussions with R N Mantegna are gratefully acknowledged. Any correspondance should be sent to Professor Nicolas Vandewalle.

References

[1] Mantegna R N 1999 Hierarchical structure in financial markets *Eur. Phys. J.* B **11** 193–7
[2] Mantegna R N and Stanley H E 1999 *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge: Cambridge University Press)
[3] Mézard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
[4] Rivier N 1985 Statistical crystallography, structure of random cellular networks *Phil. Mag.* B **52** 795–819
[5] Papadimitriou C H and Steiglitz K 1982 *Combinatorial Optimization* (Englewood Cliffs, NJ: Prentice-Hall)
[6] Bak P, Tang C and Wiesenfeld K 1987 Self-organized criticality: an explanation of $1/f$ noise *Phys. Rev. Lett.* **59** 381–4

[4] This constraint has a counterpart $A = 5$ in the case of foams [4].

[5] This is illustrated on the Market Topology SPRL website, see footnote 3.