Available online at www.sciencedirect.com
ScienceDirect
Physica A 370 (2006) 156–161
PHYSICA A
www.elsevier.com/locate/physa

Dynamical networks from correlations

T. Aste, T. Di Matteo*

Department of Applied Mathematics, Research School of Physical Sciences and Engineering, The Australian National University, 0200 Canberra, Australia

Available online 15 May 2006

Abstract

The extraction of relevant and meaningful information from large streams of data has become one of the major challenges for scientists working in the field of complex systems. In particular, one of the main goals is to get information about the underlying system of interactions that leads to complex collective dynamics. In this paper, we discuss how a set of relevant interactions can be extracted from the analysis of the cross-correlation matrix. We show that an active and adaptive correlation filtering procedure can be associated to the dynamics of a network which is a sort of 'hyper-molecule' warped on a *D*-dimensional unitary sphere.

Keywords: Complex systems; Time series analysis; Networks; Financial data correlations; Econophysics

1. Introduction

The analysis of the correlations is one of the main instruments in the study of the collective behavior of a system comprised by many elements. In general, the goal is to extract a *structure of relevant interactions* which give information about the collective properties of the system under investigation. However, the extraction of information from the observed correlations is not straightforward. Schematically, one can point out two major challenges. The *first* challenge consists in the fact that, for finite time series, random fluctuations can generate spurious correlations, making uncorrelated events to appear correlated and vice versa [1]. Therefore, the correlation structure needs to be 'undressed' from the 'dressing' noise contribution [2–4]. The *second* major challenge is associated to the fact that there is a correlation coefficient for any couple of elements in the system, whereas one aims to extract only a 'backbone skeleton' containing a few relevant interactions. Therefore, the correlation structure needs to be 'filtered', eliminating the less-relevant and redundant information [5–8].

In this paper, we present and discuss some approaches to extract the underlying structure of relevant interactions from cross-correlations. In particular, in Section 2, we discuss a direct (but unsatisfactory) method to associate a system of interactions to the correlation matrix. In Section 3, the difficulties associated with the previous method are overcome by adopting a geometrical interpretation of the correlation structure.

*Corresponding author. Tel.: +61 2 61250166; fax: +61 2 61250732.
E-mail address: tiziana.dimatteo@anu.edu.au (T. Di Matteo).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2006.04.019

Such a geometrical approach is used in Section 4 where a system of interactions is constructed directly from the set of correlations. Conclusions and discussions are given in Section 5.

2. Interactions from correlations

Let us consider a cross-correlation matrix **C** (calculated—for instance—from a system of N financial time series over a given time period) and let us show how to associate to this correlation a system of N elements which interact with exchange coefficients $J_{i,j}$:

$$U_1(\{r\}) = \sum_{i,j=1}^{N} r_i J_{i,j} r_j, \tag{1}$$

where $r_i \in \mathbb{R}$ (with $i = 1, \ldots, N$) are the values of the 'instantaneous states' of the system of N interacting 'agents'. The quantity U_1 is a 'cost' or an 'energy' associated with each state $\{r\} = (r_1, r_2, \ldots, r_i, \ldots, r_N)$ of the system. The most probable distribution for the values $\{r\}$ corresponding to a given average cost, can be obtained by maximizing the system's entropy ($S = -\sum_{\{r\}} P(\{r\}) \ln P(\{r\})$) while constraining the average cost, yielding to the following probability distribution: $P(\{r\}) = p_0 \exp[-\beta U_1(\{r\})$ (where p_0 is a normalization constant and β is a Lagrange multiplayer fixed by the average cost). From this expression for the probability distribution, it can be calculated that an exchange matrix with coefficients

$$J_{i,j} = (\mathbf{C}^{-1})_{i,j} = \sum_{v=1}^{N} \frac{e_i^{(v)} e_j^{(v)}}{\lambda^{(v)}} \tag{2}$$

(with $\lambda^{(v)}$ the vth eigenvalue of **C** and $e_i^{(v)}$ is the ith component of the vth eigenvector of **C**) yields to cross-correlation coefficients $C_{i,j}$.

From this result it appears that a set of interactions (i.e. $J_{i,j}$) can be always and uniquely extracted from any given (strictly positively defined) correlation matrix. Unfortunately, the resulting system of interactions is not, in general, reflecting the important interactions in the underlying system. Indeed, by looking at Eq. (2), one can note that the exchange coefficients $J_{i,j}$ are weighted with the reciprocal eigenvalues of the original correlation matrix: $1/\lambda^{(v)}$. This implies that the small eigenvalues dominate the structure of the interactions. On the other hand, small eigenvalues are strongly affected by the noise [1]. Therefore, in many practical cases, $J_{i,j}$ is dominated by the noise and the most part of the relevant information is hidden by the 'dressing' noise [2,3].

As a meaningful simple example let us discuss here the case of correlations between the fluctuations in time of 16 Eurodollar interest rates with different maturity dates. These correlations have been the subject of several investigations with various techniques reported in previous papers by the authors [6,7,9]. They therefore are good candidates for a comparative analysis.

In Fig. 1a is reported the minimum spanning tree (MST) constructed by connecting together the most correlated couples while constraining the resulting graph to be a tree (no closed loops). As one can see, the result is a rather simple structure of interactions with the small maturity dates (from 3 months to 2 years) that 'line' up in increasing order, followed by only a few re-arrangements and branching at large maturity dates (between 2 and 4 years).

By construction, the correlation coefficients can be always written as the following sum: $C_{i,j} = \sum_{v=1}^{N} \lambda^{(v)} e_i^{(v)} e_j^{(v)}$, which brings in evidence the contribution from each eigenvalue λ^v. The large eigenvalues are the ones that contribute most. Whereas, the small eigenvalues are the most affected by random, information-less, fluctuations. Indeed, it has been calculated that a set of N independent, uncorrected, random series of length T, produces correlation matrices with small off-diagonal entries and eigenvalues distributed in a narrow range around 1 with extremes: $\lambda_{min}^{max} = 1 + N/T \pm 2\sqrt{N/T}$ [1]. It has been argued that in financial data, most of the observed eigenvalue's spectrum is associated with noise and only a few large eigenvalues are instead associated with collective properties of the system [2,3]. Therefore, it appears that, to retrieve relevant information, it is necessary to 'undress' the observed correlations from the noise part [4] by removing the contribution from the small eigenvalues. For instance, for the Eurodollars interest rates, we find that there are only two eigenvalues which are larger than 0.5. In Fig. 1b it is shown the MST

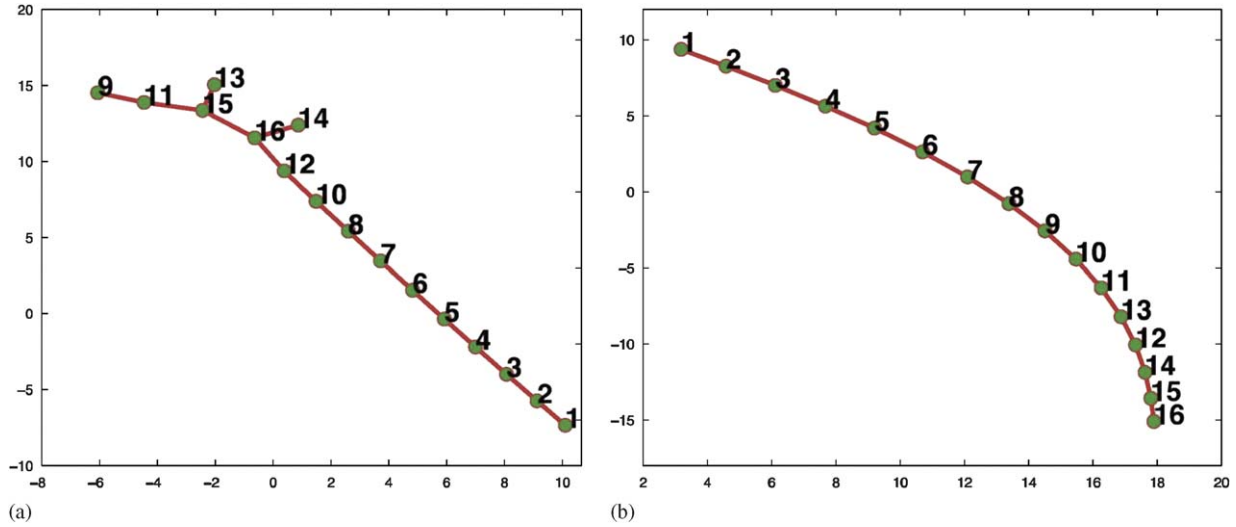

Fig. 1. (a) Minimum spanning tree (MST) constructed by connecting the most correlated couples (largest $C_{i,j}$). (b) MST from the contribution corresponding to the two largest eigenvalues.

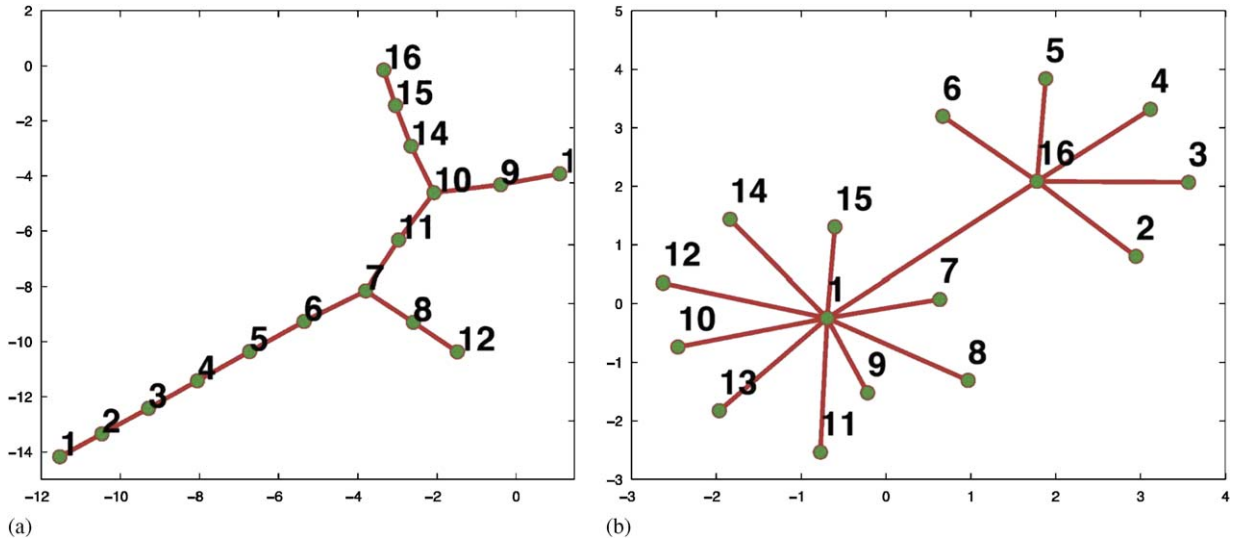

Fig. 2. (a) Minimum spanning tree constructed by joining together the most negative interactions $J_{i,j}$. (b) MST from the contribution to **J** associated with the two largest eigenvalues.

constructed from the correlation matrix recalculated by keeping the contributions from the first and the second eigenvalue only ($C^{(2)}_{i,j} = \sum_{v=1}^{2} \lambda^{(v)} e_i^{(v)} e_j^{(v)} / \sqrt{\sum_{v=1}^{2} \lambda^{(v)} [e_i^{(v)}]^2 \sum_{v=1}^{2} \lambda^{(v)} [e_j^{(v)}]^2}$). As one can see this component reveals a simpler structure: all the interest rates are now aligned by maturity dates.

Let us now analyze how the correlation structure is 'reflected' in the structure of the interactions. It is easy to see that large *negative* values of the exchange coefficient $J_{i,j}$ must be associated with elements that tend to fluctuate synchronous (a 'ferromagnetic' interaction), whereas large positive values of the $J_{i,j}$ must be associated with a-synchronous anti-correlated variables, and small or zero coefficients should correspond to non-interacting elements. A schematic view of the resulting interactions can be visualized by constructing the MST associated with the *most negative* interactions. Such a tree, shown in Fig. 2a, can be compared with the

one in Fig. 1a. As one can see the results are quite different. Similar analysis performed on more complex set of data such as the 100 US stocks described in Refs. [5,8] show that indeed the two analysis produce very different outputs.

Let now see whether such differences might be a consequence of the dressing noise and let us understand if there is a way to construct a system of interactions among the N agents which correctly reflects the structure of the undressed correlations. A straightforward approach cannot be adopted. Indeed, the undressed correlation matrix has several zero eigenvalues and it cannot be inverted. From the principal component analysis we know that the large eigenvalues are the ones that most contribute to $\mathbf{C}$. In a similar way, also the interaction is made of a sum of terms, each corresponding to the contribution from each eigenvalue: $J_{i,j} = \sum_{v=1}^{N} e_i^{(v)} e_j^{(v)} / \lambda^{(v)}$. It seems therefore reasonable to construct an undressed interaction matrix by keeping only the contribution to $\mathbf{J}$ from the largest $\lambda^{(v)}$. The resulting MST for $v = 1$ and 2 is shown in Fig. 2b. As one can see, the resulting hierarchical structure is rather different from the ones for $\mathbf{J}$ and $\mathbf{C}$. The question is whether or not these structures of interactions are meaningful. A careful analysis on these and other cross-correlation data (such as the 34 US interest rates or 100 US stocks studied in Refs. [5–8]), seems to indicate that such structures do not reflect any known property of the market.

It appears therefore that a simple interpretation of the observed data in terms of a system of interacting elements described by Eq. (1) is not adequate and it fails in catching the essential features associated with these systems. On the other hand, the idea of being able to model and analyze the complex dynamics of such systems in terms of simple interactions between elements is intriguing. Therefore, in the following we discuss an alternative approach which can provide such a description.

3. Interactions in D-dimensions

Let us here adopt a geometrical perspective to understand how, in general, a structure of interactions can be extracted from a cross-correlation matrix $\mathbf{C}$.

From a geometrical perspective, the cross-correlation among N time series $r_i(t)$, with $i = 1, \ldots, N$ and $t = t_1, \ldots, t_T$, are the scalar products between N unit vectors in a T-dimensional Euclidean space: $C_{i,j} = \vec{v}_i \vec{v}_j$, where $\vec{v}_i$ are standardized series with zero mean and unitary variance (i.e. unitary vectors in $\mathbb{R}^T$).

We want to start from a given correlation matrix $\mathbf{C}$ and search for a corresponding representation as a system of N unitary vectors $\vec{n}_i$ in a D-dimensional space. Of course, the original time series (opportunely standardized), is a solution in $D = T$, but there are an infinite number of other possible solutions that we want to explore. The first problem to address is to find a suitable space dimension D where one can place the N vectors $\vec{n}_i$ in such a way that they stay at relative distances $d_{i,j}$. In terms of information content, it is clear that if the rank of the correlation matrix is equal to N, then one needs at least N dimensions to construct such a set of vectors. Actually, there is a known construction that allow us to realize such a system *exactly* in N dimensions. This is done from the following vectors:

$$\vec{n}_i = \left(\sqrt{\lambda^{(1)}} e_i^{(1)}, \sqrt{\lambda^{(2)}} e_i^{(2)}, \ldots, \sqrt{\lambda^{(N)}} e_i^{(N)} \right), \tag{3}$$

which correctly gives: $\vec{n}_i \vec{n}_j = \sum_{v=1}^{N} \lambda^{(v)} e_i^{(v)} e_j^{(v)} = C_{i,j}$. The system's collective dynamics is in this way reduced to the study of the relative trajectories of N points on a N-dimensional unit sphere. The general goal is then to extract from such a structure some relevant patterns which are robust and persistent.

However, this requires the analysis of the dynamics of N points in a space of dimensionality larger or equal than N. In order to simplify the problem the dimensionality of space must be further reduced. This is possible if we release some of the constraints by performing a *correlation filtering*.

4. Correlation filtering

The topological structure associated to such system of correlations is a complete graph where each node is connected with every other node. The corresponding geometrical structure is an irregular simplex (triangle in 2D, tetrahedron in 3D, etc.) with vertices constrained on the surface of a unitary sphere and with edge lengths equal to $d_{i,j} = \sqrt{(\vec{v}_i - \vec{v}_j)^2} = \sqrt{2(1 - C_{i,j})}$. We know that such a structure can be realized in N dimensions.

However, for many practical purposes, correlations still retain an excessive amount of information and several tools have been developed to filter out the redundancies and keep the essential features that characterize the systems properties [6–8].

A straightforward possibility is to keep only the contribution from the $D \leqslant N$ largest eigenvalues. This leads to the following set of unit vectors: $\vec{n}_i^{(D)} = \vec{u}_i^{(D)}/|\vec{u}_i^{(D)}|$ with $\vec{u}_i^{(D)} = (\sqrt{\lambda^{(1)}}e_i^{(1)}, \sqrt{\lambda^{(2)}}e_i^{(2)}, \ldots, \sqrt{\lambda^{(D)}}e_i^{(D)})$. Projecting in this way the effective dimension from N to D. Another possibility consists in the production of a surrogate correlation matrix where only a sub-set of relevant coefficients are kept. From a geometrical perspective, such a correlation filtering consists in removing the constraint associated with the relative distance between the corresponding elements. Doing so one reduces the number of constraints and consequently one should be able also to reduce the space-dimensionality.

A correlation filtering generated by retaining only the essential links is unsatisfactory from a geometrical point of view. Indeed, in such a construction all the elements which are not connected in the filtered framework can eventually stay at any relative distance. The result is that in some cases, couples of elements that are weakly correlated and not connected in the filtered correlation network might result at very close distances acting in this representation as highly correlated bonds. Therefore, the proper representation, associated with a good correlation filtering, must meet two main criteria: (1) all the elements which are connected in the filtered system of correlations must stay at a distance which is as close as possible to $\sqrt{2(1-C_{i,j})}$; (2) all the unconnected couples must not stay too close.

Finding the right position: Let us address the general problem of placing at the right relative distance a set of N elements in a multidimensional space. This problem can be studied in terms of a dynamical model where an interaction between connected nodes is introduced by using 'springs' with rest lengths equal to the distances $d_{i,j}$. This corresponds to the following energy:

$$U_s = \sum_{i,j>i} K_{i,j} A_{i,j} \left(\sqrt{(\vec{n}_i - \vec{n}_j)^2} - d_{i,j} \right)^2, \tag{4}$$

where $A_{i,j}$ is the adjacency matrix ($A_{i,j} = 1$ if (i,j) are connected, $A_{i,j} = 0$ otherwise). Whereas $K_{i,j} > 0$ are the spring constants.

The energy U_s is minimized when all the connected points stay at the 'right' distances $d_{i,j}$. But, unconnected points can stay at arbitrary relative positions. In order to keep separate unconnected points, on can think in terms of packing of rigid bodies: connected couples ($A_{i,j} = 1$) should stay at the right distances $d_{i,j}$, whereas unconnected couples ($A_{i,j} = 0$) must stay at distances larger or equal to $\sigma_{i,j} = \max_k \{d_{i,k}, d_{j,k}\}$. This can be enforced by introducing the additional Lennard–Jones term in the energy:

$$U_L = \sum_{i,j>i} (1 - A_{i,j}) h_{i,j} \left[\left(\frac{\sigma_{i,j}^2}{(\vec{n}_i - \vec{n}_j)^2} \right)^6 - \left(\frac{\sigma_{i,j}^2}{(\vec{n}_i - \vec{n}_j)^2} \right)^3 \right], \tag{5}$$

with $h_{i,j} > 0$ the coupling constants. In this way, the unconnected nodes are treated as atoms with a long-range attraction and a short-range repulsion which becomes very strong at distances around or below $\sigma_{i,j}$. The whole system is now an hyper-molecule in $\mathbb{R}^D$ folded on the surface of a unit sphere. The space dimension D is, so far, arbitrary. Indeed a relaxation dynamics can be implemented in any dimension. However, larger is the space-dimensionality, higher are the degrees of freedom and therefore higher are the possibilities to find a configuration which minimizes the energy. On the other hand, the computational weight increases linearly with D and systems under-constrained lead to multiple solutions. The optimal protocol should be a balance between geometrical frustration and the computational attainability.

We can further generalize the procedure by introducing elementary topological moves (such as edge-switching *T1* or addiction and removal of edges [10,11]). Such moves modify the adjacency matrix and accordingly the energy. A Monte Carlo simulation, Glauber dynamics [12] should drive the system towards a global equilibrium. The introduction of topological moves transforms these dynamical geometrical constructions into active correlation filtering procedures. A straightforward advantage of such filtering is that any resulting filtered matrix $C_{i,j}^{filt} = \vec{n}_i \vec{n}_j$ is always a proper correlation matrix with rank equal to D

(indeed, by definition, any matrix constructed as a scalar product of a set of vectors is positively defined and vice versa).

5. Discussion and conclusions

In this paper, we have discussed some possible methodologies for modeling, representing and filtering the structure of correlations among N interacting elements. In Section 3, we have shown that a simple quadratic modeling of interactions (Eq. (1)) does not catch the essential features of such systems leading to a network of interactions which does not seem to reflect any known property of the underlying system. On the other hand, it is known that the correlation structure is associated with meaningful properties of the system [5–8]. Therefore, in Section 4 we have developed a geometrical method that will produce the correct correlation structure by keeping the relevant interactions. The great advantage of this approach is that it does not involve the inversion of the correlation matrix and therefore, differently from the first method, it is not prone to the effect of small eigenvalues and, consequently, noise. However, to exactly reproduce the original correlation matrix, such a method requires a space dimensionality larger or equal than N. A correlation filtering (Section 4) can then be used to reduce the space-dimensionality while keeping the relevant structure of interactions. In this respect the geometrical representation of the first approach (Section 3) corresponds to a geometrical embedding on a circle (one degree of freedom per element) which eventually implies that only the structure associated with the two largest eigenvalues ($D = 2$) might be recovered. We have shown in Section 4 that the reduction of dimensionality can be done in two ways. The first method consists in projecting the set of N points, which represents the system in N dimensions, into a subspace of lower dimensionality D. This method, is the geometrical analogous of the principal components analysis. The second method is more elaborated and consists in the creation of a set of interactions which maps the system of correlations into a sort of hyper-molecule warped on a D-dimensional unitary sphere. A dynamics of elementary moves that attaches, detaches and switches bonds in such an hyper-molecule will correspond to an active and adaptive correlation filtering procedure. The study of the dynamics of such hyper-molecule is left to future work.

Acknowledgments

We wish to acknowledge the partial support by ARC Discovery Projects: DP03440044 (2003) and DP0558183 (2005), COST P10 ''Physics of Risk'' project and M.I.U.R.-F.I.S.R. Project ''Ultra-high frequency dynamics of financial markets''.

References

[1] A.M. Sengupta, P.P. Mitra, Distributions of singular values for some random matrices, Phys. Rev. E 60 (1999) 3389–3392.
[2] L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, Noise dressing of financial correlation matrices, Phys. Rev. Lett. 83 (1999) 1467–1470.
[3] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, Universal and nonuniversal properties of cross correlations in financial time series, Phys. Rev. Lett. 83 (1999) 1471–1474.
[4] L. Giada, M. Marsili, Data clustering and noise undressing of correlation matrices, Phys. Rev. E 63 (2001) 061101.
[5] R.N. Mantegna, Hierarchical structure in financial markets, Eur. Phys. J. B 11 (1999) 193–197.
[6] T.D. Matteo, T. Aste, R.N. Mantegna, An interest rate cluster analysis, Physica A 330–339 (2004) 181–188.
[7] T.D. Matteo, T. Aste, S.T. Hyde, S. Ramsden, Interest rates hierarchical structure, Physica A 335 (2005) 21–33.
[8] M. Tumminello, T. Aste, T.D. Matteo, R.N. Mantegna, A tool for filtering information in complex systems, Proc. Natl. Acad. Sci. 102 (2005) 10421–10426.
[9] T.D. Matteo, T. Aste, How does the eurodollars interest rate behave?, J. Theoret. Appl. Finance 5 (2002) 122–127.
[10] T. Aste, D. Boosé, N. Rivier, From one cell to the whole froth: a dynamical map, Phys. Rev. E 53 (1996) 6181–6191.
[11] T. Aste, The shell map, in: J.F. Sadoc, N. Rivier (Eds.), Foams and Emulsions, 1999, pp. 497–510.
[12] T. Aste, D. Sherrington, Glass transition in self organizing cellular patterns, J. Phys. A: Math. Gen. 32 (1999) 7049–7056.