NETWORKS IN FINANCE

Fabrizio Lillo
Dip. di Fisica e Tecnologie Relative, viale delle Scienze I-90128, Palermo, Italy
Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501,USA

Keywords: Financial networks**,** Economic networks, Correlation based networks, Systemic risk, Hierarchies, Bootstrap, Market models

Contents

Summary

This chapter reviews a vast literature on the application of network theory to financial markets. The investigated networks are divided in two main categories, *similarity based networks* and *direct interaction networks*. In the first type of networks a link between two nodes represents a similarity in the behavior or activity of the agents (traders, investors, firms, stocks, banks, etc.) represented by the nodes. In the second case the link represents a quantity exchanged or, more generally, a direct dependence (ownership, credit, etc.) between the agents. For similarity based networks this chapter reviews several recent methods to extract a network from a similarity matrix with a special emphasis to correlation based networks, i.e. networks where the similarity measure is the linear cross correlation. These methods include the threshold networks, the

minimum spanning trees, and the planar maximally filtered graphs. The financial investigated systems include: price returns and volatilities of a portfolio of stocks or market indices, interest rates, hedge funds, and financial agents. For the direct interaction networks, this chapter describes the empirical results obtained by investigating large databases describing board of directors, ownership networks, interbank and payment bank networks, and credit networks. It is shown here how these network are built from large databases and the empirical topological properties of these networks.

1. Introduction

The application of networks in Finance is relatively recent but has exploded in the last decades. It is impossible to account for all the approaches that have been pursued (for approaches different form the one followed here, see, for example, Ref. [1, 2]). This chapter mainly focuses on (i) methodological aspects, i.e. which types of networks can be constructed for financial systems, (ii) empirical results on networks obtained by investigating large databases of financial data ranging from individual transactions in a financial market to strategic decisions at a bank level, and (iii) the use of networks to validate simple models with real data. In other words, this chapter surveys a subsample of the empirical works that have been performed on the analysis of financial networks extracted from data and the comparison of the properties of the obtained networks with financial models. An important classification of networks in general, and that will be considered here, is the one that divides networks in similarity based networks and direct interaction networks. To be specific in the case of financial case consider a network whose nodes are financial agents (investors, banks, hedge funds, etc.). What is the meaning of the links? In similarity based networks a link between two nodes exists if the two nodes (agents) have a strong similarity in their characteristics, strategy, behavior, etc.. In this case one needs to assign a criterion to establish whether the similarity between two agents is relevant and is associated to a link. Moreover the agents may also not interact directly, but if they are similar enough they are connected. By converse in direct interaction networks a link between two nodes signals the presence of an interaction between the entities represented by the two nodes connected by the link. In the financial case the interaction can be a transaction between two agents, a ownership relation of one node with respect to the other, a credit relation, etc. both types of networks will be considered here and several instances of both will be shown here. More specifically, Section 2 considers methods for similarity based networks and their application in Finance, whereas Section 3 discusses several examples of direct interaction networks in financial systems. Section 4 concludes. The choice of the topics reflects the knowledge and the interest of the author and it has no ambition to cover all topics.

2. Similarity Based Networks

In recent years there has been a growing interest in applying concepts and tools of similarity based graphs to financial data. This section will review the definition of similarity based graphs and the author will summarize the application of such graphs to financial markets. A similarity based graph is a graph where the links between nodes convey information on the similarity between the entities represented by the nodes. To be specific, consider a system composed of N elements. Each element is represented by T variables. These may be variables describing different properties of the elements or they can represent values of the variables at different times. In this latter case the elements are represented by time series. It is possible to define in many different ways a similarity $N \times N$ matrix $\mathbf{C}$ associated to the system, where the generic element c_{ij} is the similarity between element i and j. Clearly, one can also consider dissimilarity measures, such as for example distance measures. It is natural to associate a weighted and completely connected network to a similarity matrix. Each of the N elements is represented by a node and the link connecting node i and j is associated to a weight related to c_{ij}. Unfortunately this equivalent representation of the similarity matrix is typically not very useful or enlightening. Apart from some exceptional cases, the corresponding network is completely connected and thus the topology is in some sense trivial. Of course important information is contained in the weights, i.e. the similarities, associated with the links. Very often similarity measures are statistical variables. A typical example discussed below is the correlation coefficient between two time series. This means that, for example, a similarity might be different from zero also when the two elements are unrelated because of statistical fluctuations. Moreover the number of different similarity indicators to be estimated is typically large, making the problem of statistical fluctuations more relevant. This is a very common problem in multivariate analysis and it is sometime called curse of dimensionality. The problem is that the similarity (or correlation) matrix has $N\left(N-1\right)/2 \sim N^2$ distinct elements that must be estimated. The number of data points for this estimation is NT. Therefore, unless $T \gg N$, the statistical reliability of the similarity matrix is small. This fact implies that great part of the similarity matrix is affected by statistical noise and it is difficult to discriminate the noise from the signal. This fact has also practical implications (see also below for an example in portfolio optimization). The challenge is then to devise methods to filter the similarity matrix by retaining only the "part" which is statistically more significant. What the word part means discriminates among different filtering procedures. In many cases of interest here the filtering of the similarity matrix destroys the complete connectedness of the network. If the filtering procedure is able to remove the noise from the similarity matrix, the topology of the network representing the filtered similarity matrix can give important insight on the true similarity structure of the system. The different ways in which one filters the similarity matrix lead to different graphs which are then investigated by using tools of network theory.

2.1. Correlation based graphs: Introduction

A specific, but important, example of similarity indicators is the case of correlation matrices which is the matrix whose element ρ_{ij} is the linear (or Pearson's) cross correlation between element i and j. This defined as,

$$\rho_{ij} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{\left(\langle r_i^2 \rangle - \langle r_i \rangle^2\right)\left(\langle r_j^2 \rangle - \langle r_j \rangle^2\right)}} \tag{1}$$

where r_i and r_j are the investigated variables and the symbol $< \ldots >$ is a statistical average defined as

$$\langle r_i \rangle = \frac{1}{T} \sum_{k=1,T} r_i(k) \qquad k = 1,\ldots T$$

and $r_i(k)$ is the $k-$th variable of element i. The correlation coefficient has values between −1 and +1, corresponding, respectively, to perfectly anticorrelated and perfectly correlated (i.e. identical) variables. A correlation coefficient equal to zero means that the two variables are uncorrelated. It is important to stress that two uncorrelated variables are not necessarily independent (while the opposite is true), because non linear correlation may be significant. Apart the Pearson's correlation defined above, there are many other definitions of correlation between elements, such as for example the Kendall's tau and the Spearman rank coefficient [3].

A measures of dissimilarity are distances d_{ij} between the vectors $\vec{r}_i$ and $\vec{r}_j$ defined in the T dimensional space by the components $r_i(k)$ and $r_j(k)$. The most common distance is the

Euclidean distance

$$d_{ij}^{\mathrm{E}} = \sqrt{\sum_{k=1,T} \left(r_i(k) - r_i(k) \right)^2}$$

but there are many other possible distance definitions (e.g. the Manhattan distance, the p-norm distance, etc.). A distance d_{ij} fulfills the three axioms of a metric – (i) $d_{ij} = 0$ if and only if $i = j$; (ii) $d_{ij} = d_{ji}$ and (iii) $d_{ij} \le d_{ik} + d_{kj}$. Any distance is a dissimilarity measure, i.e. the larger the distance, the smaller the similarity. It is interesting to note that the linear correlation coefficient of Eq. (1) can be associated to a metric distance [4, 5]. In fact, if one standardizes the variables by subtracting the mean and dividing by the standard deviation,

$$\tilde{r}_i(k) = \frac{r_i(k) - \langle r_i \rangle}{\sqrt{\langle r_i \rangle^2 - \langle r_i^2 \rangle}}$$

the cross correlation $\tilde{\rho}_i$ between $\tilde{r}_i$ and $\tilde{r}_j$ is related to the Euclidean distance $\tilde{d}_{ij}^{\mathrm{E}}$ between the corresponding vectors by the relation

$$\tilde{d}_{ij}^{\mathrm{E}} = \sqrt{2\left(1 - \tilde{\rho}_{ij}\right)}$$

This shows that cross correlation is related to a distance measure. Given the monotonicity of this relation, most of the results that the author shows below holds both for correlation coefficient and for the associated distance $\tilde{d}_{ij}^{\mathrm{E}}$.

2.2. Correlation Based Graphs: Methods

Here the author describes several methods to filter correlation matrices. Even if explicitly linear correlation is considered, the methods described below work also for generic similarity measures.

-
-
-

Bibliography

[1] Nagurney A. and Siokos S. (1997). *Financial Networks: Statics and Dynamics*, Springer. [A classical reference on the application of networks to finance by using an economist's approach]

[2] Jackson M. O. (2008). *Social and Economic Networks*, Princeton University Press. [A new book on social and economic networks, focused on game theory on networks]

[3] Press W.H., Teukolsky S.A., Veterling W.T., and Flannery B.P. (1992). *Numerical Recipes in Fortran: the art of scientific computing*, 2nd Edition, Cambridge University Press, Cambridge. [A book on numerical methods and their computational implementation]

[4] Gower J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325-338. [This paper introduces the relation between linear correlation and a Euclidean distance]

[5] Mantegna R.N. and Stanley H.E. (2000). *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge Univ. Press, Cambridge UK. [One of the first books on Econophysics]

[6] Butte A. J., Tamayo P., Slonim D., Golub T.R., and Kohane I.S. (2000). Discovering functional relationship between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proceedings of the National Academy of Sciences USA* **97**, 12182-12186. [An application of threshold network approach to a biological problem of functional relationship detection]

[7] Eguiluz V.M., Chialvo D.R., Cecchi G.A., Baliki M., and Apkarian A.V. (2005). Scale-Free Brain Functional Networks. *Physical Review Letters* **94**, 018102. [Threshold networks applied to fMRI data]

[8] Mardia K.V., Kent J.T., and Bibby J.M. (1979). *Multivariate Analysis*, Academic Press, San Diego, CA. [A classical reference in multivariate analysis]

[9] Anderberg M.R. (1973). *Cluster Analysis for Applications*, Academic Press, New York. [A classical reference in cluster analysis]

[10] Giada L. and Marsili M (2001). Data clustering and noise undressing of correlation matrices. *Physical Review E* **63**, 1101 (2001). [A non-conventional approach to data clustering and correlation analysis]

[11] Tumminello M., Lillo F., and Mantegna R. N. (2008). Correlation, hierarchies, and networks in financial markets. Preprint at arXiv:0809.4615; in press in *Journal of Economic Behavior and Organization*, (2010). [A recent review of correlation based approach to hierarchies]

[12] West D.B. (2001). *An Introduction to Graph Theory*, Prentice-Hall, Englewood Cliffs, NJ. [A classical reference in graph theory]

[13] Rammal R., Toulouse G., Virasoro M.A. (1986). Ultrametricity for physicists. *Review of Modern Physics* **58**, 765-788. [A review article on the mathematics of ultrametricity and their applications to statistical physics problem]

[14] Tumminello M., Coronnello C., Lillo F., Miccichè S., and Mantegna R.N. (2007). Spanning Trees and Bootstrap Reliability Estimation in Correlation-Based Networks. *International Journal of Bifurcation and Chaos* **17** 2319-2329. [This article introduces the average linkage minimum spanning tree and discusses the application of bootstrap to correlation based clustering]

[15] Aste T., Di Matteo T., and Hyde S.T. (2005). Complex networks on hyperbolic surfaces.*Physica A* **346**, 20-26. [A first paper on correlation based planar graphs]

[16] Tumminello M., Aste T., Di Matteo T. and Mantegna R. N. (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences USA* **102**, 10421-10426. [This paper introduces the planar maximally filtered graph]

[17] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics* **7**, 1-26. [The seminal paper on bootstrap]

[18] Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap, *Evolution* **39**, 783-791. [Paper discussing the application of bootstrap to phylogenetic analysis]

[19] Tumminello M., Lillo F., Mantegna R.N. (2007). Hierarchically nested factor model from multivariate data *EPL* **78**, 30006. [This article introduces a method for associating a time series factor model to a dendrogram]

[20] Onnela J.-P., Chakraborti A., Kaski K., Kertesz J., and Kanto A. (2003). Asset Trees and Asset Graphs in Financial Markets. *Physica Scripta* **T106**, 48-54. [This article uses threshold networks to investigate correlation between asset returns]

[21] Onnela J.-P., Kaski K., and Kertesz J. (2004). Clustering and information in correlation based financial networks *European Physical Journal B* **38**, 353-362. [A comparison between threshold networks and minimum spanning trees for financial asset returns]

[22] Mantegna R.N. (1999). Hierarchical structure in financial markets. *European Physical Journal B* **11** 193-197. [One of the first application of the minimum spanning tree to correlation between asset returns]

[23] Coronnello C., Tumminello M., Lillo F., Micciche S., Mantegna R.N. (2005). Sector identification in a set of stock return time series traded at the London Stock Exchange. *Acta Physica Polonica B* **36** 2653-2679. [An in-depth comparison of different methods for the identification of economic sectors from price return correlation matrices]

[24] Gopikrishnan P., Rosenow B., Plerou V., Stanley H.E. (2001). Quantifying and interpreting collective behavior in financial markets. *Physical Review E* **64**, 035106. [A spectral method for identifying economic sectors from price return correlation matrices]

[25] Penrose, M.D. (1996). The random minimal spanning tree in high dimensions. *Annals of Applied Probability* **24**, 1903-1925. [A theoretical paper computing the topology of a random minimum spanning tree]

[26] Bonanno G., Caldarelli G., Lillo F., and Mantegna R.N. (2003). Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E* **68**, 046130. [A comparison of the topological properties of minimum spanning trees obtained from financial data with trees obtained from simple market models]

[27] Campbell Y.J., Lo A.W., and Mackinlay A.C. (1997). The *Econometrics of Financial Markets*, Princeton University Press, Princeton. [A classical textbook of financial econometrics]

[28] Vandewalle N., Brisbois F., and Tordoir X. (2001). Non-random topology of stock markets. *Quantitative Finance* **1**, 372-374. [One of the first paper investigating the topological properties of financial minimum spanning trees]

[29] Bonanno G., Lillo F., and Mantegna R.N. (2001). High-frequency cross-correlation in a set of stocks *Quantitative Finance* **1**, 96-104. [An empirical study of the return cross correlation at different time scales with an emphasis on minimum spanning trees and dendrograms]

[30] Epps T.W. (1979). Comovements in stock prices in the very short run. *Journal of American Statistical Association* **74** 291-296. [A first paper showing the decrease of return cross correlation at very short time scales]

[31] Onnela J.-P., Chakraborti A., Kaski K., and Kertesz J. (2002). Dynamic asset trees and portfolio analysis. *European Physical Journal B* **30**, 285-288. [Threshold networks for portfolio optimization]

[32] Tumminello M., Di Matteo T., Aste T., and Mantegna R.N. (2007). Correlation based networks of equity returns sampled at different time horizons. *European Physical Journal. B* **55**, 209-217. [An empirical study of the change of correlation based networks of financial cross correlations]

[33] Lo A. and MacKinlay A.C. (1990). When are contrarian profits due to stock market overreaction? *Review of Financial Studies* **3**, 175-208. [A paper showing the existence of lagged return cross-correlation between stocks with different capitalization]

[34] Kullmann L., Kertesz J., and Kaski K. (2002). Time-dependent cross-correlations between different stock returns: A directed network of influence. *Physical Review E* **66**, 026125. [An empirical study of high frequency lagged return cross correlation and directed networks]

[35] Elton E.J. and Gruber M. J. (1995). *Modern Portfolio Theory and Investment Analysis*, John Wiley and Sons, New York. [A textbook on portfolio optimization]

[36] Tola V., Lillo F., Gallegati M., Mantegna R.N. (2008). Cluster analysis for portfolio optimization. *Journal of Economic Dynamics & Control* **32** 235-258. [Application of hierarchical clustering to portfolio optimization]

[37] Bonanno G., Vandewalle N., and Mantegna R.N. (2000). Taxonomy of stock market indices. *Physical Review E* **62**, R7615-R7618. [Application of minimum spanning tree to the analysis of worldwide financial market indices]

[38] Miccichè S., Bonanno G., Lillo F., and Mantegna R.N. (2003). Volatility in financial markets: stochastic models and empirical results. *Physica A* **314**, 756-761. [Minimum spanning tree and dendrograms for the analysis of cross correlations of volatility]

[39] Di Matteo T., Aste T., Hyde S.T., and Ramsden S. (2004). Interest rates hierarchical structure. *Physica A* **335** 21-33. [Correlation based clustering and networks applied to interest rates]

[40] Miceli M.A. and Susinno G. (2003). Using trees to grow money. *Risk*, November, S11-S12. [Correlation based clustering and networks applied to hedge funds]

[41] Miceli M.A. and Susinno G. (2004). Ultrametricity in fund of funds diversification. *Physica A* **344**, 95-99. [Correlation based clustering and networks applied to hedge funds]

[42] Lillo F., Moro E., Vaglica G., and Mantegna R.N.(2008). Specialization and herding behavior of trading firms in a financial market. *New Journal of Physics* **10** 043019 (2008). [Analysis of cross correlation matrices of member firm inventory variation and community identification]

[43] Zovko I.I., and Farmer J.D. (2007). Correlations and clustering in the trading of members of the London Stock Exchange. Preprint at arxiv.org:0709.3261v1. [Analysis of cross correlation matrices of member firm inventory variation]

[44] Tumminello M., Piilo J., Lillo F., and Mantegna R.N. (2010). (in preparation). [Analysis of the similarity of trading behavior of individual investors in the Finnish stock market]

[45] Davis, G.F., Yoo, M., and Baker, W. E. (2003). The small world of the corporate elite,1982-2001. *Strategic Organization* **1**, 301-326. [Analysis of the board of directors networks]

[46] Battiston, S. and Catanzaro, M. (2004). Statistical properties of corporate board and director networks. *European Physical Journal B* **38**, 345-352. . [Analysis of the board of directors networks]

[47] Caldarelli G. (2007). *Scale-Free Networks complex webs in nature and technology*, Oxford University Press. [Book on scale free networks]

[48] Garlaschelli, D., Battiston, S., Castri, M., Servedio, V.D.P., and Caldarelli, G. (2005). The scale free topology of market investments. *Physica A* **350** 491-499. [Network empirical analysis of investment relation between firms and individuals]

[49] Pareto V. (1897). *Cours d'Economie Politique*, Lausanne and Paris. [Classical work on the distribution of wealth]

[50] Glattfelder J.B., Battiston S. (2009). Backbone of complex networks of corporations: The flow of control. *Physical Review E* **80**, 036104. [Modeling and empirical analysis of the network of control of worldwide companies]

[51] Boss M., Elsinger H., Summer M., and Thurner S. (2004). Network topology of the interbank market, *Quantitative Finance* **4** 677-684. [Empirical analysis of the Austrian interbank market]

[52] Girvan M. and Newman M.E.J. (2002). Community structure in social and biological networks. *Proceedings of National Academy of Science USA* **99** 7821-7826. [Paper introducing a popular method for community identification in networks]

[53] De Bandt O. and Hartmann P. (2000). Systemic Risk: a survey. CEPR Discussion Paper No. 2634. [Classical survey on systemic risk]

[54] Sorämaki K., Bech M.L., Arnold J., Glass R.J., and W.E. Beyeler (2007). The topology of the interbank payment flows. *Physica A* **379**, 317-333. [Empirical paper investigating the topology of the US interbank payment market]

[55] De Masi G., Iori G., and Caldarelli G. (2006). Fitness model for the interbank money market. *Physical Review E* **74**, 066112. [Application of the fitness model to the interbank market network]

[56] Iori G., Renó R., De Masi G., Caldarelli G. (2007). Trading strategies in the Italian interbank market. *Physica A* **376**, 467-479. [Empirical analysis of the Italian interbank market network]

[57] Iori G., De Masi G., Precup O.V., Gabbid G., Caldarelli G. (2008). A network analysis of the Italian overnight money market. *Journal of Economic Dynamics & Control* **32**, 259-278. [Empirical analysis of the Italian interbank market network]

[58] Caldarelli G., Capocci A., De Los Rios P., and Mu˜noz M.A (2002). Scale-free networks from varying vertex intrinsic fitness. *Physical Review Letters* **89**, 258702. [Model introducing the vertex fitness for building scale-free networks]

[59] Iori G., Jafarey S., Padilla F. (2006). Systemic risk on the interbank market. *Journal of Economic Behaviour and Organization* **61**, 525-542. [Model of systemic risk of the interbank market with methods of statistical physics]

[60] G. De Masi and M. Gallegati, (2007) Debt-credit Economic Networks of Banks and Firms: the Italian Case, in *Econophysics of Markets and Business Networks*, A. Chatterjee and B.K. Chakrabarti (editors), pp 159-171 [Paper on the bank-firm credit network in the Italian economy]

[61] De Masi G., Fujiwara Y., Gallegati M., Greenwald B., and Stiglitz J.E. (2009). An Analysis of the Japanese Credit Network. Preprint at arXiv:0901.2384. [An empirical analysis of the Japanese bank-firm credit network]

[62] Fujiwara Y., Aoyama H., Ikeda Y., Iyetomi H., and SoumaW. (2009). Structure and Temporal Change of the Credit Network between Banks and Large Firms in Japan. *Economics* **3**, 2009-7. [An empirical analysis of the Japanese bank-firm credit network]

Biographical Sketch

Fabrizio Lillo is professor at the University of Palermo (Italy) and at the Santa Fe Institute. He has been awarded of the Young Scientist Award for Socio- and Econophysics of the German Physical Society in 2007. His research is focused on the application of methods and tools of statistical physics to economic, financial, and biological systems. Recently he has been interested in the microstructure of financial markets and in the empirical study of economic and financial systems where the data allow investigating the behavior of individual agents with the aim of building empirically based agent-based models. He obtained his PhD in Physics at the University of Palermo. He has been researcher of the National Institute for the Physics of Matter (Italy) and postdoctoral fellow at the Santa Fe Institute. He is author of more than 60 scientific papers.