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

1. Introduction
2. Similarity based networks
   2.1. Correlation based graphs: Introduction
   2.2. Correlation Based Graphs: Methods
      2.2.1. Threshold or Relevance Network
      2.2.2. Hierarchical Methods
      2.2.3. Validation of Correlation Based Networks with Bootstrap Techniques
   2.3. Correlation Based Networks Applied To Finance
      2.3.1. Stock Returns
      2.3.2. Stock Market Indices
      2.3.3. Market Volatility
      2.3.4. Interest Rates
      2.3.5. Hedge Funds
      2.3.6.. Financial Agent Networks
3. Direct interaction networks
   3.1. Direct Interaction Networks Applied To Finance
      3.1.1. Board of Directors
      3.1.2. Ownership Networks
      3.3.3. Bank Networks
      3.3.4. Credit Networks
4. Conclusions
Glossary
Bibliography
Biographical Sketch

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 $C$ 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 $N(N-1)/2 \sim N^2$ 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 $\rho_{ij}$ is the linear (or Pearson’s) cross correlation between element $i$ and $j$. This defined as,

$$\rho_{ij} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{\langle r_i^2 \rangle - \langle r_i \rangle^2} \sqrt{\langle r_j^2 \rangle - \langle r_j \rangle^2}}$$ (1)
where \( r_i \) and \( r_j \) are the investigated variables and the symbol \(<\ldots>\) is a statistical average defined as

\[
\langle r_i \rangle = \frac{1}{T} \sum_{k=1,T} r_i(k) \quad k = 1, \ldots T
\]

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

\[
d_{ij}^E = \sqrt{\sum_{k=1,T} (r_i(k) - r_j(k))^2}
\]

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) \( d_{ij} \leq d_{ik} + d_{kj} \). 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,

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

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

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

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


Biographical Sketch

Fabrizio Lillo is professor at the University of Palermo (Italy) and at the Santa Fe Institute. He has been awarded 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.