Dynamical correlations in financial systems

F. Pozzi[a], T. Aste[a], G. Rotundo[b] and T. Di Matteo[a]

[a]Department of Applied Mathematics, The Australian National University, 0200 Canberra, ACT, Australia.

[b]Department of Mathematics for Economic, Financial and Insurance Decisions, University of Rome La Sapienza, Via del Castro Laurenziano, 9 Rome, 00161, Italy.

ABSTRACT

One of the main goals in the field of complex systems is the selection and extraction of relevant and meaningful information about the properties of the underlying system from large datasets. In the last years different methods have been proposed for filtering financial data by extracting a structure of interactions from cross-correlation matrices where only few entries are selected by means of criteria borrowed from network theory. We discuss and compare the stability and robustness of two methods: the Minimum Spanning Tree and the Planar Maximally Filtered Graph. We construct such graphs dynamically by considering running windows of the whole dataset. We study their stability and their edges's persistence and we come to the conclusion that the Planar Maximally Filtered Graph offers a richer and more significant structure with respect to the Minimum Spanning Tree, showing also a stronger stability in the long run.

Keywords: Econophysics; Complex Systems; Networks; Minimum Spanning Tree; Planar Maximally Filtered Graph; Financial Data Correlations.

1. INTRODUCTION

In the last few years, many filtering methods have been developed by econophysicists in order to extract relevant information from huge amount of financial data. Two of such methods are the Minimum Spanning Tree (denoted from now on as MST)[1] used by Mantegna for financial data in ref.[2] and the Planar Maximally Filtered Graph (denoted from now on as $PMFG$) introduced by Tumminello et al. in ref.[3]

In this paper we analyze, compare and discuss the robustness, the stability and the structural fluctuations of MST and $PMFG$ considered as graphs dynamically evolving over time.

This paper is organized as follows. In section 2 we illustrate the data set and we introduce the correlation matrices and the associated complete graphs from financial time series. We show that there are some similarities between dynamical systems of correlations built on moving dynamical windows of different lengths Δt and the static system built on the entire data set which can be seen as its long run stable structure. In section 3 the dynamical MST and $PMFG$ are introduced and described, and some properties of such subgraphs are discussed and compared with the dynamical complete graphs from which they are extracted. In section 4 we discuss differences in the averages and in standard deviations computed over subgraphs and complete graphs. A comparison with systems of interest rates has been made and it shows significant differences which are directly associated to the specific peculiarities of the markets. In order to assess the robustness of such graphs, in section 5 the frequencies of appearance of edges in the dynamical MSTs and $PMFG$s are computed for each given Δt. In section 6 we introduce a new graph built as the union of all edges that can be reached with a $T1$ elementary movement from a given $PMFG$.[4,5] Intersections between dynamical subgraphs and their corresponding static subgraphs are computed and relevant differences related to long-run time-persistence of edges are shown. In section 7 we draw some conclusions and propose some suggestions for future research.

Send correspondence to frp110@rsphysse.anu.edu.au

Complex Systems II, edited by Derek Abbott, Tomaso Aste, Murray Batchelor, Robert Dewar,
Tiziana Di Matteo, Tony Guttmann, Proc. of SPIE Vol. 6802, 68021E, (2008)
0277-786X/08/$18 · doi: 10.1117/12.758822

2. DYNAMICAL CORRELATIONS

2.1 Data description

We have analyzed daily time series of the $n = 300$ most capitalized $NYSE$ companies from 2001 to 2003, for a total of $T = 748$ days. Return time series are computed as logarithmic differences of daily prices, and daily prices are computed as averages of daily quotations. Closing quotations are excluded from the computation. In the following we denote with Y the 748×300 matrix of returns.

Stocks are classified into 12 economic sectors and 77 economic subsectors.

2.2 Distance Matrices from correlations

Let us consider all time data subsets of dimensions $\Delta t \times 300$, where Δt corresponds to a moving window, from time (t) to time $(t + \Delta t - 1)$, where $t = 1,\ 2\ ,\ \ldots\ ,\ (T - \Delta t + 1)$ and $\Delta t = 21,\ 42,\ 63,\ 84,\ 126,\ 251$ days, corresponding approximately to $\Delta t = 1,\ 2,\ 3,\ 4,\ 6,\ 12$ months. For each t and Δt, the resulting matrix is denoted as $Y_{\tau,s}$ with $\tau = t, (t+1), ..., (t + \Delta t - 1)$ and $s = 1,\ 2,\ \ldots\ ,\ 300$.

The number of these matrices, for each choice of Δt, is shown in Table 1.

Table 1. Number of dynamical correlation matrices associated to the choice of the moving window Δt.

Δt months	**Δt days**	**cases n°**
1	21	728
2	42	707
3	63	686
4	84	665
6	126	623
12	251	498
36	748	1

For each of such matrices, we computed the correlation matrix $C(t, \Delta t)$, which is a 300×300 matrix with coefficients given by the formula

$$c_{i,j}(t, \Delta t) = \frac{\langle Y_{\tau,i} Y_{\tau,j} \rangle_\tau - \langle Y_{\tau,i} \rangle_\tau \langle Y_{\tau,j} \rangle_\tau}{\sqrt{\left(\left\langle Y^2_{\tau,i} \right\rangle_\tau - \langle Y_{\tau,i} \rangle^2_\tau \right) \left(\left\langle Y^2_{\tau,j} \right\rangle_\tau - \langle Y_{\tau,j} \rangle^2_\tau \right)}} \quad , \tag{1}$$

where $\langle f_\tau \rangle_\tau = \frac{1}{\Delta t} \sum_{\tau=1}^{\Delta t} f_\tau$ is the time average of a given series f_τ. From the correlation coefficients $c_{i,j}$, we can write a well-known measure of distance between stocks i and j: $d_{i,j} = \sqrt{2(1 - c_{i,j})}$. Such distance is the euclidean metric distance computed between standardized returns $Z_{\tau,i}$ of stocks i and j where

$$Z_{\tau,i} = \frac{Y_{\tau,i} - \langle Y_{\tau,i} \rangle_\tau}{\sqrt{\left(\left\langle Y^2_{\tau,i} \right\rangle_\tau - \langle Y_{\tau,i} \rangle^2_\tau \right)}} \quad . \tag{2}$$

The distance $d_{i,j}$ is a function, $d : Y_{\tau,i} \times Y_{\tau,j} \to R$, such that $d_{i,j} \in [0, 2]$, with $d_{i,j} = 0$ if $c_{i,j} = 1$, $d_{i,j} = \sqrt{2}$ if $c_{i,j} = 0$ and $d_{i,j} = 2$ if $c_{i,j} = -1$. All standard properties of a metric distance are satisfied.

The matrices $D(t, \Delta t) = \sqrt{2(1 - C(t, \Delta t))}$ can be interpreted as dynamical distance matrices of weighted complete graphs K_{300} where all 300 stocks are interconnected.

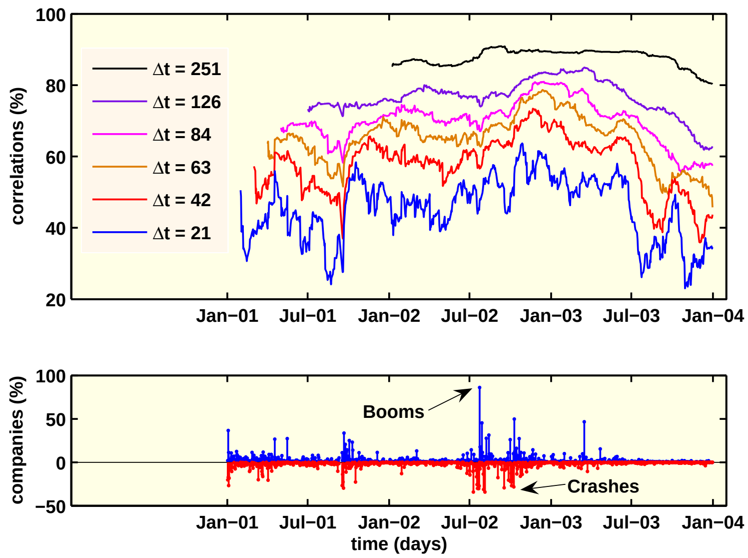

Figure 1. Correlations between the dynamical distance matrices $D(t, \Delta t)$, computed at $\Delta t = 21,\ 42,\ 63,\ 84,\ 126,\ 251$ days, corresponding approximately to 1, 2, 3, 4, 6, 12 months, and the static distance matrix D^* obtained by using all $T = 748$ days available from data set of the 300 most capitalized $NYSE$ stocks' time series, corresponding to years $2001 - 2003$. The higher curve is obtained for $\Delta t = 251$ days, the lower for $\Delta t = 21$ days. At the bottom, percentages of companies whose standardized return, at each time t, exceeds two standard deviations or falls below minus two standard deviations.

2.3 Static Graph and Dynamical Graphs

As a first step we computed the static distance matrix D^* on the entire data set Y and, for each t and Δt, we computed the correlations between such matrix and the dynamical distance matrices $D(t, \Delta t)$. Such correlations are

$$E(t, \Delta t) = \frac{\left\langle d_{i,j}(t, \Delta t) d^*_{i,j} \right\rangle_{i,j} - \left\langle d_{i,j}(t, \Delta t) \right\rangle_{i,j} \left\langle d^*_{i,j} \right\rangle_{i,j}}{\sqrt{\left(\left\langle d^2_{i,j}(t, \Delta t) \right\rangle_{i,j} - \left\langle d_{i,j}(t, \Delta t) \right\rangle^2_{i,j} \right) \left(\left\langle d^{*2}_{i,j} \right\rangle_{i,j} - \left\langle d^*_{i,j} \right\rangle^2_{i,j} \right)}} \tag{3}$$

with $d_{i,j}$ and $d^*_{i,j}$ respectively the elements of the distance matrices $D(t, \Delta t)$ and D^* and where $\langle f_{i,j} \rangle_{i,j} = \frac{2}{n(n-1)} \sum_{i<j} f_{i,j}$ denotes the average of $f_{i,j}$ over all edges.

Correlations $E(t, \Delta t)$ for each t and Δt are shown in $Figure$ 1. At the bottom of the figure, we show the percentage of companies whose standardized return, at each time t, exceeds two standard deviations (proxy measure for booms, positive values) or falls below minus two standard deviations (proxy measure for crashes, negative values).

As Δt increases, we observe that the dynamical distance matrices get at every step closer to the static distance matrix, built on the entire data set. Relevant fluctuations observed at low levels of Δt, turn out to be strongly damped at higher levels. We observe that, after periods of particular turbulence, the dynamical system of correlations becomes closer to the static distance matrix. As we can see from Table 2, when $\Delta t = 21$ days, the range of correlations is from a minimum of 23.02% to a maximum of 63.67%, the average being 45.39% and standard deviation 8.86%. When $\Delta t = 84$ days, the range of correlations is from a minimum of 55.47% to a maximum of 80.94%, the average being 69.95% and standard deviation 6.08%. When $\Delta t = 251$ days, the range of correlations is from a minimum of 80.28% to a maximum of 90.93%, the average being 87.73% and standard deviation 2.34%. Thus, we see that the range becomes progressively narrower and the average higher.

Table 2. Summary for correlations between static distance matrix D^* and the dynamical distance matrices $D(t, \Delta t)$.

Δt	**Min**	**Mean**	**Max**	**Std**
21	0.2302	0.4539	0.6367	0.0886
42	0.3588	0.5778	0.7342	0.0803
63	0.4573	0.6510	0.7865	0.0683
84	0.5547	0.6995	0.8094	0.0608
126	0.6188	0.7664	0.8493	0.0492
251	0.8028	0.8773	0.9093	0.0234

It's worth mentioning that dynamical distance matrices built on only one third of the entire data set (corresponding to $\Delta t = 251$ days) are very close to the static distance matrix built on the entire data set. This is showing a fast convergence with Δt of the dynamical distances towards the static distances.

3. DYNAMICAL MINIMUM SPANNING TREES AND PLANAR MAXIMALLY FILTERED GRAPHS

The graphs associated to matrices $D(t, \Delta t)$ are complete graphs K_{300}, which have $n(n-1)/2 = (300)(299)/2 = 44850$ edges connecting all pairs of nodes. Different methods exist in literature in order to filter such a huge amount of data, otherwise hardly readable and usable. One approach consists in extracting a sub-graph which retains the most valuable information and eliminates most of the redundancies, producing identifiable hierarchies and communities.

A widely used method is the Minimum Spanning Tree (MST), used for the first time in finance literature by Mantegna.[2] The MST is a tree, a graph with no cycles, in which all nodes are connected, and edges are selected in order to minimize the sum of distances. The total number of edges is $n-1$, where n is the number of nodes. Several algorithms to construct the MST have been developed by the community of computer scientists and are widely known since 1926 (Otakar Boruvka's Algorithm). The most commonly used are Prim and Kruskal algorithms that find the MST in polynomial time. The efficiency of algorithms for finding the MST has been continuously enhanced over years (see, for instance, Eisner[6]). An almost linear running time algorithm has been recently developed by Chazelle.[7] Since we have computed almost $4,000$ MSTs out of 300 nodes's graphs, the efficiency of the algorithm had to be considered. We have used Prim's algorithm implemented in Matlab and we have found it efficient enough for our purposes.

A filtering method which uses a similar principle, but allows more interactions and a more complex and rich structure, is the Planar Maximally Filtered Graph ($PMFG$), proposed for the first time by Tumminello et al. in ref.[3] Such method constructs a connected planar graph[8] where edges are selected in order to minimize the sum of distances. In this case, the total number of edges is $3(n-2)$, approximately the triple number of edges than the MST. It has been proved by Tumminello et al. in ref.[3] that the MST is always a subgraph of the $PMFG$. For each dynamical distance matrix $D(t, \Delta t)$ we computed the corresponding dynamical MSTs and $PMFG$s. We computed also the Static MST and $PMFG$, over the entire period, in order to be able to compare their properties with those of the dynamical sub-graphs.

Averages and standard deviations have been computed for each t and Δt for edges belonging to the complete graphs $D(t, \Delta t)$, to the dynamical $MST(t, \Delta t)$ and to the dynamical $PMFG(t, \Delta t)$. Moreover, for each t and Δt we computed the averages and standard deviations for edges belonging to the Static MST and $PMFG$. The average distances in the dynamically moving distance matrices for all the graphs, computed at $\Delta t = 21$ days and at $\Delta t = 251$ days are shown in Figure 2.

We observe that the average of complete graphs's distances can be considered as a superior limit: dynamical MSTs and $PMFG$s must have average distances lower or equal than the corresponding complete graphs. Conversely, distance averages of edges belonging to the Static MST and $PMFG$ can be higher: but if this happens it indicates the total lack of significance and robustness for the relative subgraphs's selection.

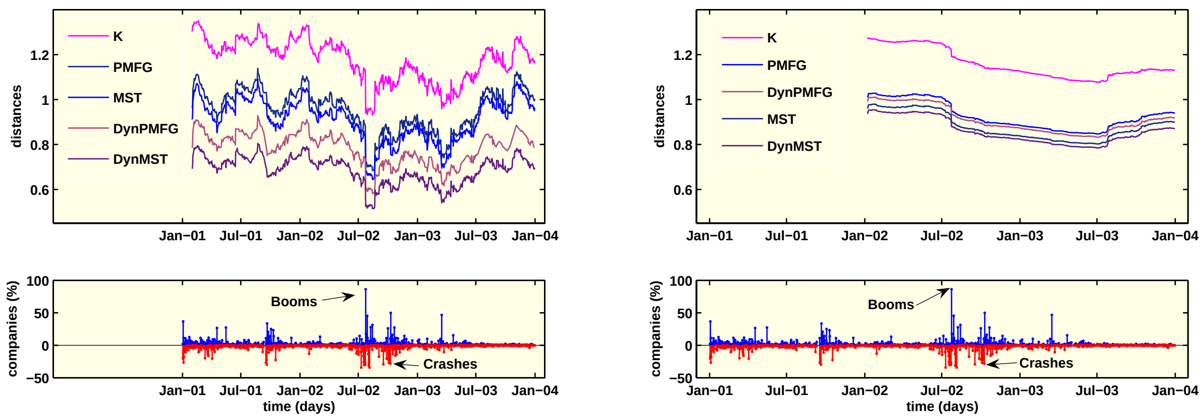

Figure 2. From top to bottom, on the left: the average distances in the dynamically moving distance matrices $D(t, \Delta t)$ computed at $\Delta t = 21$ days, for edges belonging to: the complete graph, the Static $PMFG$ computed on all 748 days, the Static MST computed on all 748 days, the dynamical $PMFG$s computed at $\Delta t = 21$ days, the dynamical MSTs computed at $\Delta t = 21$ days. On the right: $\Delta t = 251$ days; averages of the dynamical $PMFG$s are above those of the Static MST. At the bottom, percentages of companies whose standardized return, at each time t, exceeds two standard deviations or falls below minus two standard deviations.

When $\Delta t = 21$ days, the figure can be divided in three regions: at the top average distances of complete graphs's edges; in the middle average distances of the edges belonging to the Static $PMFG$ and MST; at the bottom average distances of dynamical $PMFG$s and MSTs. All curves show the same patterns and trends: subgraphs MSTs and $PMFG$s reproduce well the properties of their corresponding complete graphs.

When $\Delta t = 251$ days, the figure can be divided in two regions only: at the top average distances of the complete graphs; at the bottom, well beneath the first curve, average distances of edges belonging to the Static $PMFG$; then the dynamical $PMFG$; and further down the Static MST followed by the dynamical MST. Note that in this case, the Static MST is below the dynamical $PMFG$.

Dynamical $PMFG$s exhibit behaviors and performances thoroughly similar to dynamical MSTs, with only slightly higher average distances.

It is of some interest to note that a remarkable sudden fall, consequent to turbulences due to the $July/October$ 2002 stock market downturn, is clearly visible and is protracted for the entire period Δt (21 days in the first case, 251 days in the second) and after that it is suddenly and completely re-absorbed. It is noteworthy that a single anomalous data point in $July$ 2002 influences the average distances for all Δt following periods.

We observe that the average distances of the dynamical graphs MST and $PMFG$ are closer to average distances of edges belonging to the corresponding static graphs than to the complete graph, with tightening gaps as Δt increases. This means that the selection of edges performed by our graph's filtering is significantly robust.

4. THE MEAN-σ PLANE

For each set of edges of the dynamical graphs and for each set of edges belonging to the static graphs, we calculated both the average distance and the standard deviation σ in the matrices $D(t, \Delta t)$. We then consider the mean-σ plane finding that, when $\Delta t = 21$ days, variances of subgraphs's edges are almost always lower than or equal to variances of complete graphs's edges while, conversely, when $\Delta t = 251$ days, variances of subgraphs edges are always higher.

For each time t, at a given Δt, each dynamical graph or each static edge selection is represented by one point in the mean-σ plane. When $\Delta t = 21$ days, these points in the mean-σ plane are distributed uniformly on elliptical clouds, with each subgraph quite clustered. Differently, when $\Delta t = 251$ days, the elliptical clouds become smooth straight lines, now very well distinct from each other, and their slopes are negative. Increasing

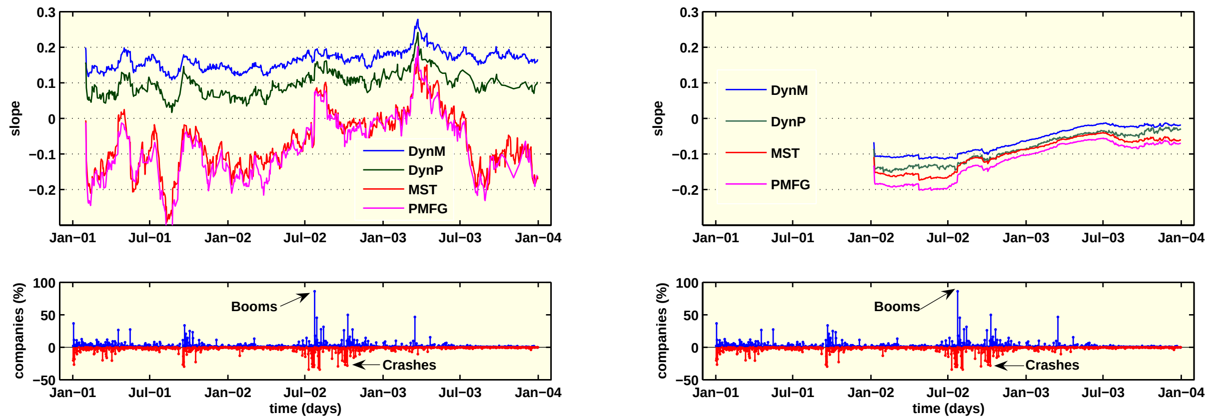

Figure 3. From top to bottom, slope in the mean-σ plane of the lines joining one mean-σ point of the dynamical complete graphs's edges to one mean-σ point of edges belonging to the dynamical MST and $PMFG$, and to the Static MST and $PMFG$. On the left: $\Delta t = 21$; on the right: $\Delta t = 251$. At the bottom, percentages of companies whose standardized return, at each time t, exceeds two standard deviations or falls below minus two standard deviations.

Δt, if the average distance increases then the variance decreases and viceversa. We also computed the slope of the line connecting, at each time t, one mean-σ point of all edges (complete graphs's) to one mean-σ point of the dynamical MST or $PMFG$ or the Static MST or $PMFG$. We find that, when $\Delta t = 21$ days (*Figure* 3, left side), the slopes between the complete graph and the dynamical MST or $PMFG$ are always positive. Conversely the slopes between the complete graph and the Static MST or $PMFG$ are almost always negative, except for periods of particularly intense turbulence. On the other hand, when $\Delta t = 251$ days (*Figure* 3, right side), the same slopes are always negative for all cases and for all periods. These findings imply that the system of 300 stocks is generally poorly correlated, with relatively small variances. But, during and after periods of intense turbulence, the time series get suddenly correlated and the variances increase.

We have made the same computations for a system of 16 Eurodollar interest rates's daily quotations and one more system of 34 interest rates's weekly quotations.[9–14] We obtain in both cases the opposite result: a positive slope on the mean-σ plane. This is not surprising because these are highly regulated systems indeed, monitored under strict control and strongly influenced by an international institutional system made of many cooperating national Central Banks. During a turbulent period, interest rates time series get partially decorrelated until agents and authorities adjust their positions, and then the system gets correlated again. Conversely, stock markets are highly competitive, hardly controllable, with dynamics hardly manageable and predictable, so they are much more complex and turbulent systems. During calm periods, the system is less correlated than during turbulent ones, when agents are driven by euphoria or panic; in such a system, public authorities have a lower control.

5. FREQUENCIES OF SUB-GRAPHS'S EDGES

Both MST and $PMFG$ select many statistically significant edges with high positive correlations but also some residual edges with lower weights. The dynamical graphs have some edges which appear often and others that are inserted only rarely. In order to detect significant edges, a frequency has been computed for each edge and for each Δt.

Both relative and absolute edge frequencies for dynamical MSTs are shown in Table 3 and Table 4, analogously for dynamical $PMFG$s in Table 5 and Table 6.

We find that, when $\Delta t = 251$ days, that is when the filtering is particularly robust, several edges that never appear in the dynamical MSTs appear very often (more than 70% of cases) in the dynamical $PMFG$s instead. We observe that more than 99.90% of all edges for dynamical MSTs and more than 99.50% of all edges for

Table 3. Relative frequencies, at each Δt, for edges belonging to the dynamical MSTs.

Δt	$= 0$	< 0.1	< 0.2	...	< 0.7	< 0.8	< 0.9	$= 1$
1	42.22%	99.31%	99.81%	...	99.99%	99.99%	100%	0.00%
2	66.68%	98.92%	99.59%	...	99.98%	99.99%	99.99%	0.00%
3	77.73%	98.68%	99.44%	...	99.98%	99.98%	99.99%	0.01%
4	83.62%	98.49%	99.34%	...	99.96%	99.98%	99.99%	0.01%
6	89.45%	98.25%	99.18%	...	99.93%	99.95%	99.98%	0.01%
12	95.48%	98.37%	98.96%	...	99.81%	99.86%	99.92%	0.03%

Table 4. Absolute frequencies, at each Δt, for edges belonging to the dynamical MSTs.

Δt	$= 0$	> 0.1	> 0.2	...	> 0.7	> 0.8	> 0.9	$= 1$
1	18,936	310	85	...	5	4	2	0
2	29,907	485	182	...	7	6	5	2
3	34,864	593	250	...	11	7	6	3
4	37,505	675	295	...	18	11	6	5
6	40,120	785	370	...	33	23	9	6
12	42,821	731	466	...	84	61	36	14

Table 5. Relative frequencies, at each Δt, for edges belonging to the dynamical $PMFG$s.

Δt	$= 0$	< 0.1	< 0.2	...	< 0.7	< 0.8	< 0.9	$= 1$
1	8.76%	97.65%	99.03%	...	99.97%	99.98%	99.99%	0.00%
2	20.58%	96.93%	98.5%	...	99.88%	99.92%	99.96%	0.00%
3	34.38%	96.33%	98.21%	...	99.81%	99.88%	99.94%	0.01%
4	44.74%	95.92%	98.04%	...	99.74%	99.84%	99.92%	0.02%
6	58.82%	95.61%	97.72%	...	99.58%	99.75%	99.85%	0.06%
12	79.14%	95.56%	97.21%	...	99.23%	99.44%	99.63%	0.19%

Table 6. Absolute frequencies, at each Δt, for edges belonging to the dynamical $PMFG$s.

Δt	$= 0$	> 0.1	> 0.2	...	> 0.7	> 0.8	> 0.9	$= 1$
1	3,931	1,054	435	...	15	9	3	0
2	9,232	1,377	673	...	54	36	16	2
3	15,420	1,646	803	...	86	52	26	5
4	20,068	1,828	879	...	115	73	37	11
6	26,379	1,967	1,024	...	189	113	69	25
12	35,495	1,992	1,252	...	347	249	166	84

Table 7. Dynamical MST and $PMFG$ edges, with 100% frequency. Δt =251 days.

i	CODE	SECTOR	SUBSECTOR	CODE	SECTOR	SUBSECTOR
1	*SBC*	*Services*	*CommunicationServices*	*BLS*	*Services*	*CommunicationServices*
2	*FNM*	*Financial*	*ConsumFinancServ*	*FRE*	*Financial*	*ConsumFinancServ*
3	*LEH*	*Financial*	*InvestmentServices*	*BSC*	*Financial*	*InvestmentServices*
4	*MBI*	*Financial*	*InsProp.&Casualty*	*ABK*	*Financial*	*InsProp.&Casualty*
5	*NEM*	*BasicMaterials*	*Gold&Silver*	*ABX*	*BasicMaterials*	*Gold&Silver*
6	*RD*	*Energy*	*Oil&Gas − Integrated*	*TOT*	*Energy*	*Oil&Gas − Integrated*
7	**WLP**	**Financial**	**InsAccidental&Health**	**HMA**	**Healthcare**	**HealthcareFacilities**
8	*LIZ*	*ConsumerCyclical*	*Apparel/Accessories*	*VFC*	*ConsumerCyclical*	*Apparel/Accessories*
9	*CTX*	*CapitalGood*	*ConstructionServices*	*PHM*	*CapitalGood*	*ConstructionServices*
10	**JP**	**Financial**	**InsLife**	**TMK**	**Financial**	**InsAccidental&Health**
11	*BJS*	*Energy*	*OilWellServ&Equip*	*SII*	*Energy*	*OilWellServ&Equip*
12	*KRI*	*Services*	*Printing&Publishing*	*DJ*	*Services*	*Printing&Publishing*
13	*WLP*	*Financial*	*InsAccidental&Health*	*HUM*	*Financial*	*InsAccidental&Health*
14	*WHR*	*ConsumerCyclical*	*Appliance&Tool*	*MYG*	*ConsumerCyclical*	*Appliance&Tool*

Table 8. Dynamical $PMFG$ edges, with high frequencies for $PMFG$s and 0% frequency for MSTs. Δt = 251 days.

i	CODE	SECTOR	SUBSECTOR	CODE	SECTOR	SUBSECTOR	PMFG
1	*GCI*	*Services*	*Printing&Publishing*	*DJ*	*Services*	*Printing&Publishing*	0.996
2	*SPG*	*Services*	*RealEstateOperations*	*DRE*	*Services*	*RealEstateOperations*	0.9738
3	*BLS*	*Services*	*CommunicationServices*	*CTL*	*Services*	*CommunicationServices*	0.9698
4	*ABK*	*Financial*	*InsProp.&Casualty*	*JP*	*Financial*	*InsLife*	0.9054
5	*MCD*	*Services*	*Restaurants*	*EAT*	*Services*	*Restaurants*	0.8672
6	**UTX**	**Conglomerates**	**Conglomerates**	**GD**	**CapitalGood**	**Aerospace&Defense**	**0.8672**
7	*PFE*	*Healthcare*	*MajorDrugs*	*ABT*	*Healthcare*	*MajorDrugs*	0.8632
8	*MBI*	*Financial*	*InsProp.&Casualty*	*TMK*	*Financial*	*InsAccidental&Health*	0.8491
9	**HMA**	**Healthcare**	**HealthcareFacilities**	**HUM**	**Financial**	**InsAccidental&Health**	**0.8491**
10	*IP*	*BasicMaterials*	*Paper&PaperProducts*	*PPG*	*BasicMaterials*	*ChemicalManifacturing*	0.8471
11	**MAS**	**ConsumerCyclical**	**Furniture&Fixtures**	**CTX**	**CapitalGood**	**ConstructionServices**	**0.829**
12	*HMA*	*Healthcare*	*HealthcareFacilities*	*MME*	*Healthcare*	*N\A*	0.8048
13	*VFC*	*ConsumerCyclical*	*Apparel/Accessories*	*JNY*	*ConsumerCyclical*	*Apparel/Accessories*	0.7928
14	*IP*	*BasicMaterials*	*Paper&PaperProducts*	*PD*	*BasicMaterials*	*MetalMining*	0.7928
15	*CMA*	*Financial*	*RegionalBanks*	*UPC*	*Financial*	*N\A*	0.7928
16	*PX*	*BasicMaterials*	*ChemicalManifacturing*	*ROH*	*BasicMaterials*	*Chemical − Plastic&Rubber*	0.7827
17	*RD*	*Energy*	*Oil&Gas − Integrated*	*KMG*	*Energy*	*Oil&GasOperations*	0.7807
18	*GGP*	*Services*	*RealEstateOperations*	*DRE*	*Services*	*RealEstateOperations*	0.7746
19	*GIS*	*ConsNonCycl*	*FoodProcessing*	*CPB*	*ConsNonCycl*	*FoodProcessing*	0.7485
20	*BR*	*Energy*	*Oil&GasOperations*	*UCL*	*Energy*	*Oil&GasOperations*	0.7163
21	*OXY*	*Energy*	*Oil&GasOperations*	*TOT*	*Energy*	*Oil&Gas − Integrated*	0.7163
22	*NSM*	*Technology*	*Semiconductors*	*LSI*	*Technology*	*Semiconductors*	0.7143
23	*MHP*	*Services*	*Printing&Publishing*	*DJ*	*Services*	*Printing&Publishing*	0.7022

dynamical $PMFG$s have persistence lower than 80%. More than 99% of all edges for dynamical MSTs and more than 97% of all edges for dynamical $PMFG$s have persistence lower than 20%. It is noteworthy to observe that, when Δt = 21 days, 42.2% of all edges for dynamical MSTs but only 8.8% of all edges for dynamical $PMFG$s never appear. While, when Δt = 251 days, 95.5% of all edges for dynamical MSTs and 79.1% of all edges for dynamical $PMFG$s never appear. It is also remarkable that, when Δt = 21 days, 99.3% of all edges for dynamical MSTs and 97.7% of all edges for dynamical $PMFG$s are selected in less than 10% of cases.

The most significant dynamical MST and $PMFG$ edges, with 100% frequencies, are shown in Table 7. We notice that all edges identify a specific economic activity: in the large majority of cases, the two nodes belong to the same sector and sub-sector, and when this is not so, as in rows 7 and 10, the two activities are strictly related in a specific economic sense (ie. Insurance Accidental & Health linked to Healthcare Facilities in the first case, Insurance Accidental & Health linked to Insurance Life in the second case).

In Table 8 some of the most significant edges are shown which are often selected by $PMFG$s (with a frequency of more than 70%) but never selected by MSTs. All edges are, again, strictly associated to a specific economic activity: in most of them the two nodes belong to the same sector and sub-sector. When this is not so, as in rows 6, 9 and 11, the two activities are strictly related in a specific economic sense: UTX "provides a broad range of high-technology products and support services to customers in the aerospace and building industries" and it is linked to GD that is an industry specialized in "Aerospace design, Combat Systems, Marine Systems design, Information Systems and Technology". Similarly, row 9 links the same sectors and subsectors as row 7 of Table 7. Analogously, MAS is in the field of "Furniture & Fixtures (faucets, kitchen, bath cabinets, bath and shower units, spas and hot tubs, shower and plumbing specialties, electronic lock sets and other builders' hardware, air treatment products, ventilating equipment and pumps)"and it is linked to home building company CTX whose main field is "Construction Services" and whose "principal activities are to provide residential and

Table 9. Dynamical MST and $PMFG$ edges, with 100% frequency for $PMFG$s and different frequencies for MSTs. $\Delta t = 251$ days.

i	CODE	SECTOR	SUBSECTOR	CODE	SECTOR	SUBSECTOR	MST
1	PPG	BasicMaterials	ChemicalManifacturing	WY	BasicMaterials	Forestry&WoodProducts	0.004
2	TXN	Technology	Semiconductors	NSM	Technology	Semiconductors	0.012
3	SLB	Energy	OilWellServ&Equip	BJS	Energy	OilWellServ&Equip	0.0141
4	OXY	Energy	Oil&GasOperations	UCL	Energy	Oil&GasOperations	0.0703
5	SBC	Services	CommunicationServices	AT	Services	CommunicationServices	0.1044
6	AHC	Energy	Oil&Gas − Integrated	KMG	Energy	Oil&GasOperations	0.1225
7	MER	Financial	InvestmentServices	AGE	Financial	InvestmentServices	0.1687
8	BNI	Transportation	Railroad	NSC	Transportation	Railroad	0.1888
9	GCI	Services	Printing&Publishing	TRB	Services	Printing&Publishing	0.241
10	PPG	BasicMaterials	ChemicalManifacturing	PX	BasicMaterials	ChemicalManifacturing	0.2671
11	**MME**	**Healthcare**	**N\A**	**HUM**	**Financial**	**InsAccidental&Health**	**0.2871**
12	PG	ConsNonCycl	Personal&HouseholdProducts	CLX	ConsNonCycl	Personal&HouseholdProducts	0.3133
13	BHI	Energy	OilWellServ&Equip	BJS	Energy	OilWellServ&Equip	0.3253
14	MER	Financial	InvestmentServices	BSC	Financial	InvestmentServices	0.3414
15	PPG	BasicMaterials	ChemicalManifacturing	APD	BasicMaterials	ChemicalManifacturing	0.3514
16	MBI	Financial	InsProp.&Casualty	MTG	Financial	InsProp.&Casualty	0.3735
17	UNP	Transportation	Railroad	NSC	Transportation	Railroad	0.3835
18	DD	BasicMaterials	Chemical − Plastic&Rubber	PPG	BasicMaterials	ChemicalManifacturing	0.4157
19	SLB	Energy	OilWellServ&Equip	SII	Energy	OilWellServ&Equip	0.4438
20	EQR	Services	RealEstateOperations	AIV	Services	RealEstateOperations	0.4839
21	UNP	Transportation	Railroad	BNI	Transportation	Railroad	0.5141
22	UNP	Transportation	Railroad	CSX	Transportation	Railroad	0.5402
23	APA	Energy	Oil&GasOperations	KMG	Energy	Oil&GasOperations	0.5422
24	SLB	Energy	OilWellServ&Equip	BHI	Energy	OilWellServ&Equip	0.5442
25	LNC	Financial	InsLife	TMK	Financial	InsAccidental&Health	0.5582
26	TRB	Services	Printing&Publishing	KRI	Services	Printing&Publishing	0.5622
27	FPL	Utilities	ElectricUtilities	CIN	Utilities	ElectricUtilities	0.5703
28	MBI	Financial	InsProp.&Casualty	JP	Financial	InsLife	0.5723
29	IP	BasicMaterials	Paper&PaperProducts	WY	BasicMaterials	Forestry&WoodProducts	0.5783
30	NSM	Technology	Semiconductors	TER	Technology	Semiconductors	0.6205
31	MER	Financial	InvestmentServices	LEH	Financial	InvestmentServices	0.6225
32	KR	Services	RetailGrocery	ABS	Services	RetailGrocery	0.6265
33	HD	Services	RetailHomeImprovement	LOW	Services	RetailHomeImprovement	0.6305
34	SWY	Services	RetailGrocery	ABS	Services	RetailGrocery	0.6606
35	PPG	BasicMaterials	ChemicalManifacturing	EC	BasicMaterials	ChemicalManifacturing	0.6627
36	BHI	Energy	OilWellServ&Equip	SII	Energy	OilWellServ&Equip	0.6727
37	BNI	Transportation	Railroad	CSX	Transportation	Railroad	0.6767
38	**WLP**	**Financial**	**InsAccidental&Health**	**MME**	**Healthcare**	**N\A**	**0.6767**
39	NSC	Transportation	Railroad	CSX	Transportation	Railroad	0.6968
40	**IP**	**BasicMaterials**	**Paper&PaperProducts**	**TIN**	**Conglomerates**	**Conglomerates**	**0.7048**
41	KR	Services	RetailGrocery	SWY	Services	RetailGrocery	0.7129
42	**WY**	**BasicMaterials**	**Forestry&WoodProducts**	**TIN**	**Conglomerates**	**Conglomerates**	**0.7169**
43	**GP**	**BasicMaterials**	**Paper&PaperProducts**	**TIN**	**Conglomerates**	**Conglomerates**	**0.7329**
44	PX	BasicMaterials	ChemicalManifacturing	APD	BasicMaterials	ChemicalManifacturing	0.7771
45	GCI	Services	Printing&Publishing	KRI	Services	Printing&Publishing	0.7992
46	PG	ConsNonCycl	Personal&HouseholdProducts	CL	ConsNonCycl	Personal&HouseholdProducts	0.8072
47	MRK	Healthcare	MajorDrugs	BMY	Healthcare	MajorDrugs	0.8153
48	CL	ConsNonCycl	Personal&HouseholdProducts	CLX	ConsNonCycl	Personal&HouseholdProducts	0.8333
49	FD	Services	RetailDepartment&Discount	JCP	Services	RetailDepartment&Discount	0.8394
50	ADI	Technology	Semiconductors	IRF	Technology	Semiconductors	0.8474
51	EMC	Technology	ComputerStorageDevices	ADI	Technology	Semiconductors	0.8494
52	UCL	Energy	Oil&GasOperations	AHC	Energy	Oil&Gas − Integrated	0.8775
53	LTR	Financial	InsProp.&Casualty	JP	Financial	InsLife	0.8876
54	BLS	Services	CommunicationServices	AT	Services	CommunicationServices	0.8956
55	CL	ConsNonCycl	Personal&HouseholdProducts	AVP	ConsNonCycl	Personal&HouseholdProducts	0.9016
56	ITW	CapitalGood	Misc.CapitalGoods	ETN	CapitalGood	Misc.CapitalGoods	0.9116
57	MRK	Healthcare	MajorDrugs	ABT	Healthcare	MajorDrugs	0.9157
58	PPG	BasicMaterials	ChemicalManifacturing	ROH	BasicMaterials	Chemical − Plastic&Rubber	0.9378
59	SPG	Services	RealEstateOperations	EQR	Services	RealEstateOperations	0.9438
60	CB	Financial	InsProp.&Casualty	JP	Financial	InsLife	0.9478
61	TXN	Technology	Semiconductors	ADI	Technology	Semiconductors	0.9558
62	**ETN**	**CapitalGood**	**Misc.CapitalGoods**	**PH**	**BasicMaterials**	**Misc.FabricatedProducts**	**0.9719**
63	APA	Energy	Oil&GasOperations	APC	Energy	Oil&GasOperations	0.9759
64	UCL	Energy	Oil&GasOperations	KMG	Energy	Oil&GasOperations	0.9819
65	PFE	Healthcare	MajorDrugs	MRK	Healthcare	MajorDrugs	0.99
66	SPG	Services	RealEstateOperations	GGP	Services	RealEstateOperations	0.996
67	AT	Services	CommunicationServices	CTL	Services	CommunicationServices	0.996
68	**PPG**	**BasicMaterials**	**ChemicalManifacturing**	**TIN**	**Conglomerates**	**Conglomerates**	**0.996**
69	CAT	CapitalGood	Constr.&Agric.Machinery	DE	CapitalGood	Constr.&Agric.Machinery	0.998
70	APA	Energy	Oil&GasOperations	BR	Energy	Oil&GasOperations	0.998

commercial constructions" for families and firms (details have been retrieved from companies's Web pages).

Table 9 reports some of the most significant edges that are always selected by $PMFG$s but not always selected by MSTs. Once more, we see clearly that most of these edges have both nodes belonging to the same sector and sub-sector, showing that the system of correlations is highly clustered. For instance, rows 11 and 38 are similar to row 7 of Table 7 and row 9 of Table 8. We find particularly interesting the edges involving Temple-Inland (TIN) (rows 40, 42, 43, 68), from the Conglomerates sector, that is always linked to companies belonging to the sector of Basic Materials and subsectors Forestry, Wood, Paper and Chemical Products. Temple-Inland, indeed, engages in corrugated packaging and forest products (real estate and financial services businesses). It manufactures a range of building products including lumber, studs, gypsum wallboard, engineered wood siding and trim, fiberboard sheathing.

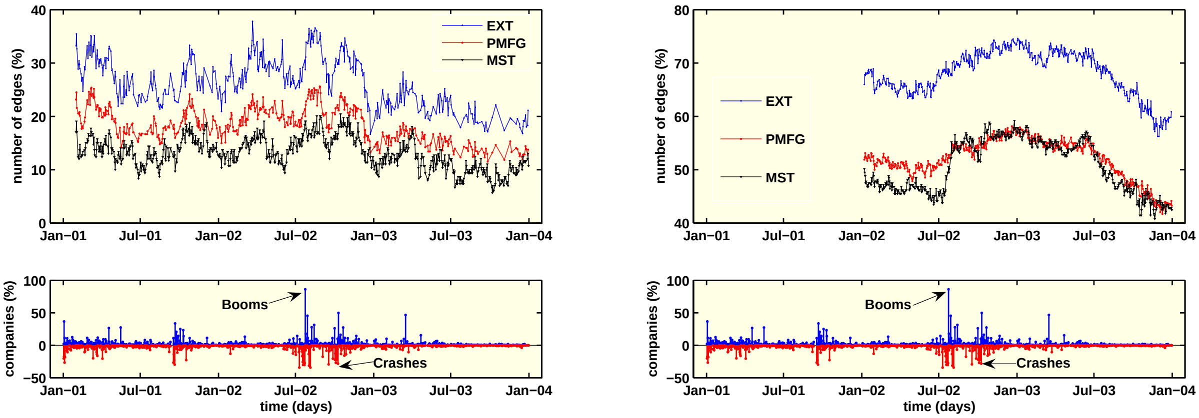

Figure 4. Percentages of persistent edges belonging (from top to bottom) to: graphs obtained by T1 expansion of $PMFG$s; dynamical $PMFG\,(t, \Delta t)$, dynamical $MST\,(t, \Delta t)$. On the left: $\Delta t = 21$; on the right: $\Delta t = 251$. At the bottom, percentages of companies whose standardized return, at each time t, exceeds two standard deviations or falls below minus two standard deviations.

From Table 8 and Table 9, we see that the $PMFG$ procedure selects some especially high quality edges that are missing, always or most of the times, from the MST.

6. LONG RUN TIME PERSISTENCES FOR EDGES OF SUB-GRAPHS

Onnela[15] and Johnson[16] introduced some interesting measures of survival for edges belonging to dynamical graphs: in particular they propose to calculate the common edges between $G(t+k)$ and $G(t)$ (single step survival ratio); or between $G(t+k)$, $G(t+k-1)$, ..., $G(t+1)$ and $G(t)$ (k multi-step survival ratio). These are short-run measures of persistence, weak in the former case; stronger, and rather restrictive, in the latter.

In this paper we have further considered the intersections between dynamical subgraphs and their corresponding static subgraphs. We have then calculated, for each t and Δt, the number of common edges between dynamical $MST\,(t, \Delta t)$ and the Static MST divided by the length of the MST; the number of common edges between dynamical $PMFG\,(t, \Delta t)$ and the Static $PMFG$ divided by the length of the $PMFG$. As we can see in $Figure$ 4, when $\Delta t = 21$ days the dynamical $PMFG\,(t, \Delta t)$ seem to be more stable than the dynamical $MST\,(t, \Delta t)$ and still slightly more stable also in the case $\Delta t = 251$ days.

Following an insight from Ohlenbusch et al.[4] and Aste et al.,[5] we have considered for each t and Δt, all local $T1$ elementary topological movements for all edges of the $PMFG$s. A $T1$ movement is an edge-switching process consisting in joining nodes c and d if and only if they are common neighbors of nodes a and b, where a and b are already linked by an edge in the graph. After joining all such nodes, we obtain a new expanded graph that contains all possible evolutions of the original planar through local $T1$ elementary topological movements. The procedure described for the planar graphs cannot be carried out for trees, since if two nodes have two common neighbors then there must be a cycle in the graph, so this cannot be a tree.

We find that the persistence of edges belonging to the new dynamical expanded planar graphs with respect to the static planar is higher than the others when $\Delta t = 21$ days and sensitively higher when $\Delta t = 251$ days.

7. CONCLUSIONS AND FUTURE RESEARCH

Financial systems are highly complex systems. Available data need to be filtered in order to be able to extract relevant and meaningful information out of an extremely huge amount of data.

In this paper we have shown that both MST and $PMFG$ reproduce pretty well the properties of the system and are structurally robust. They both select some particularly significant edges of the economic underlying system. We have seen that edges selected by both MST and $PMFG$ are impressively clustered within economic sectors and subsectors, with the $PMFG$ having a richer number of high-quality details on the financial system with respect to the MSTs.

We have introduced a new measure of survival for edges of a graph that catches their long-run persistence. We have found that the $PMFG$ seems to be slightly more persistent from a structural point of view, in the long-run, even in the case $\Delta t = 251$ days when both subgraphs are particularly robust. We have seen that, if we expand the $PMFG$ by adding edges through local $T1$ elementary topological movements, we obtain a graph that retains, in the most robust case, most of the edges belonging to the Static $PMFG$.

Further steps will be taken to investigate the robustness and the meaning of those edges that show high clustering power from an economic sectorial point of view.

ACKNOWLEDGMENTS

This work was partially supported by the ARC Discovery Projects DP0344004 (2003), DP0558183 (2005) and COST P10 "Physics of Risk" project.
F. Pozzi kindly acknowledges the Department of Public Economics and the Doctoral School of Economics of the University of Rome "La Sapienza" for a one-year research scholarship and for the authorization to spend this period at the ANU; he also acknowledges the hospitality of The Australian National University.

REFERENCES

1. J. C. Gower, G. J. S. Ross, "Minimum Spanning Trees and Single Linkage Cluster Analysis", Applied Statistics, **18/1**, 54-64, 1969.
2. R. N. Mantegna, "Hierarchical structure in financial markets", EPJB, **11**, 193-197, 1999.
3. M. Tumminello, T. Aste, T. Di Matteo and R. N. Mantegna, "A tool for filtering information in complex systems", PNAS, **102/30**, 10421-10426, 2005.
4. H. M. Ohlenbusch, T. Aste, B. Dubertret, N. Rivier, "The topological structure of 2D disordered cellular systems", EPJB, **2**, 211-220, 1998.
5. T. Aste, D. Sherrington, "Glass transition in self-organizing cellular patterns", J. Phys. A, **32**, 7049-56, 1999.
6. J. Eisner, "State-of-the-art Algorithms for MSTs: A Tutorial Discussion", Manuscript, University of Pennsylvania, 1997.
7. B. Chazelle, "A Minimum Spanning Tree Algorithm with Inverse-Ackermann Type Complexity", Journal of the ACM (JACM), **47/6**, 1028-1047, 2000.
8. A planar graph is a network that can be represented on an Euclidean plane with no intersections between edges.
9. T. Di Matteo, T. Aste, "How does the Eurodollar Interest Rate behave?", International Journal of Theoretical and Applied Finance, **5/1**, 107-122, 2002.
10. T. Di Matteo, T. Aste, R. N. Mantegna, "An interest rates cluster analysis", Physica A, **339**, 181-188, 2004.
11. T. Aste, T. Di Matteo, M. Tumminello, R. N. Mantegna, "Correlation filtering in financial time series", Proceedings-SPIE The International Society For Optical Engineering, **5848**, 100-109, 2005.
12. T. Di Matteo, T. Aste, S. T. Hyde, S. Ramsden, "Interest rates hierarchical structure", Physica A, **355**, 21-33, 2005.
13. T. Aste, T. Di Matteo, "Dynamical networks from correlations", Physica A, **370**, 156-161, 2006.
14. T. Di Matteo, T. Aste, "Extracting the correlation structure by means of planar embedding", Proceedings-SPIE The International Society For Optical Engineering, **6039**, 60390P-1, 2006.
15. J.-P. Onnela, "Taxonomy of Financial Assets", M.Sc. Thesis, Department of Electrical and Communications Engineering, Helsinki University of Technology, 2002.
16. M. McDonald, O. Suleman, S. Williams, S. Howison, N. F. Johnson, "Detecting a currency's dominance or dependence using foreign exchange network trees", Physical Review E **72**, 046106, 2005.
17. T. Aste, T. Di Matteo, S. T. Hyde, "Complex networks on hyperbolic surfaces", Physica A, **346**, 20-26, 2005.
18. M. Tumminello, T. Aste, T. Di Matteo, R. N. Mantegna, "Correlation based networks of equity returns sampled at different time horizons", EPJB, **55**, 209-217, 2007.