SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
APPLIED PHYSICS
DATA MINING
APPLIED MATHEMATICS
COMPLEX NETWORKS

Received
17 December 2012

Accepted
28 March 2013

Published
16 April 2013

Correspondence and requests for materials should be addressed to T.A. (t.aste@ucl.ac.uk)

Spread of risk across financial markets: better to invest in the peripheries

F. Pozzi[1], T. Di Matteo[2] & T. Aste[1,3]

[1]Department of Applied Mathematics, Research School of Physical Sciences, The Australian National University, 0200 Canberra, ACT, Australia, [2]Department of Mathematics, King's College London, The Strand, London, WC2R 2LS, UK, [3]Department of Computer Science, UCL, Gower Street - London - WC1E 6BT, UK.

Risk is not uniformly spread across financial markets and this fact can be exploited to reduce investment risk contributing to improve global financial stability. We discuss how, by extracting the dependency structure of financial equities, a network approach can be used to build a well-diversified portfolio that effectively reduces investment risk. We find that investments in stocks that occupy peripheral, poorly connected regions in financial filtered networks, namely Minimum Spanning Trees and Planar Maximally Filtered Graphs, are most successful in diversifying, improving the ratio between returns' average and standard deviation, reducing the likelihood of negative returns, while keeping profits in line with the general market average even for small baskets of stocks. On the contrary, investments in subsets of central, highly connected stocks are characterized by greater risk and worse performance. This methodology has the added advantage of visualizing portfolio choices directly over the graphic layout of the network.

In times of market instabilities managing risk is a top priority for the financial industry[1,2]. In this paper we investigate how financial filtered networks, namely Minimum Spanning Trees (*MST*)[3] and Planar Maximally Filtered Graphs (*PMFG*)[4], can be used to characterize the heterogeneous spreading of risk across a financial market and how this information can be employed to reduce investment risk by constructing well-diversified portfolios. Let us recall that financial filtered networks are constructed by retaining the highest correlated links while constraining some overall property of the network without need to specify any threshold[5,6]. Specifically, the *MST* is a spanning tree (a connected network with no loops or cycles) which maximizes the sum of the correlations over the connections in the tree[3]. Similarly the *PMFG* is a maximal planar graph that contains the *MST* as a subgraph and retains the largest correlations across edges[4]. The topology of these networks efficiently encodes the complex dependency structure of financial equities extracting hierarchical and clustering properties, reducing data complexity while preserving the fundamental characteristics of the dataset[3–6]. The underlying idea that we develop in this work is that stocks differently positioned within a financial filtered graph exhibit different patterns of behavior and therefore the selection of stocks from a plurality of alternative regions of the network can be used to set up efficiently diversified portfolios.

As widely accepted since Markowitz seminal work[7], an efficient diversification should aim to select stocks as anti-correlated as possible and remaining consistently anti-correlated over time[1,2]. Identifying, from the study of historical behavior prior to the investment, baskets of stocks with a good likelihood to remain well-diversified over the future investment period is very challenging. Indeed, the structure of correlations between stocks is evolving over time and changes markedly during crises. For this reason the Markowitz approach is normally applied to a selection of stocks identified by using different criteria including the industrial sector and other macro- or micro-economic considerations. In this way, a relatively small set of stocks (typically 10 to 50) is individuated and on such 'basket' the Markowitz optimal portfolio is determined.

In this paper we propose a method to identify such 'basket' of stocks directly from the dependency structure provided by the financial filtered networks. In the present study we investigated a set of highly capitalized stocks in the American Stock Exchange market in the time period ranging from 1981 to 2010 ($T = 7570$ market days). For each market day, t, we investigated the behavior of a selection of $N = 300$ stocks with high capitalization and largest performances over the previous year ($t \in \{\Delta t+1, \ldots, T-\Delta t+1\}$, $\Delta t = 250$ market days, see details in Methods section). Specifically, we computed correlations over a window of six months, reducing the excessive influence of remote market shocks on present correlations by using exponential smoothing[8] (which assigns higher weights to more recent events and incrementally lower weights to past events). We then improved the estimator by computing the average correlation matrix with shrinkage[9] over a period of six months obtaining in this way a robust estimate of the correlations over the year preceding the investment day t (see details in Methods section).

Such matrix shows a remarkable persistence, with autocorrelation values ranging around 50% even after one year. (The autocorrelation of a correlation matrix is defined as the correlation between the vectors of the $N (N - 1)/2$ correlation coefficients at time t and at time $t + \tau$.) This high persistence is a very important fact implying that measurements from the past are likely to forecast the future and the ordering of the correlations is expected to remain rather stable. We then used these average weighted correlations with shrinkage to construct *MST* and *PMFG* financial filtered networks[3,4,10]. An example of *PMFG* is shown in Figure 1.

We now discuss how an efficient investment strategy can benefit from the knowledge of such market dependency structure. In particular, we set up portfolios by selecting stocks from the peripheral regions of the financial filtered networks and we compared the performance of these portfolios with the performance of portfolios set up by selecting central stocks, or random stocks or by using other traditional methods. To this purpose, we first distinguish between stocks lying in the networks' central regions and those lying in the peripheries. Numerous centrality/peripherality measures have been proposed in the literature[11–16]; they reflect different criteria and it is not unusual that a vertex results central for one measure and peripheral for another. In particular, centrality measures on *MST* and *PMFG* tend to distinguish well the few central vertices, highly connected, important and influential, but they are less effective in ranking the different levels of peripherality of non-central vertices. We have therefore adopted an 'agnostic' perspective by looking at some of the most common centrality/peripherality measures (namely Degree (D), Betweenness Centrality (BC), Eccentricity (E), Closeness (C) and Eigenvector Centrality (EC)[15]) computed for both the weighted *MST* and *PMFG* and their unweighted counterparts. Specifically, we elaborated two hybrid centrality indices, X and Y, which group together the rankings of the previous measures (see details in Methods section). In terms of these hybrid measures, small values of $(X + Y)$ are associated with central vertices whereas large values are associated with peripheral vertices. From the study of the variation of these centrality indices over time we observed that central stocks are more persistent whereas peripheral stocks have a larger variability (see details in supporting information). We observe that, in terms of industrial sectors[17], the peripheries are mainly populated by companies belonging to "Electric, Gas, and Sanitary Services" (representing 20% of peripheral companies vs. 11% of all companies), "Oil and Gas Extraction" (7.0% vs. 4.8%), "Petroleum Refining and Related Industries" (2.3% vs. 1.7%) or "Metal Mining" (2.1% vs. 1.0%) while the core is mainly populated by "Depository Institutions" (14% vs. 6.4%), "Security and Commodity Brokers, Dealers, Exchanges, and Services" (6.6% vs. 1.4%) or "Holding and Other Investment Offices" (7.8% vs. 3.0%). These findings are consistent with analyses reported in references[18–20]. We observed that the use of this hybrid centrality measure consistently provides more stable and robust results than the use of any of the centrality measures in isolation. This is due to the different sensitivity of each centrality measure to outliers and noise[21].

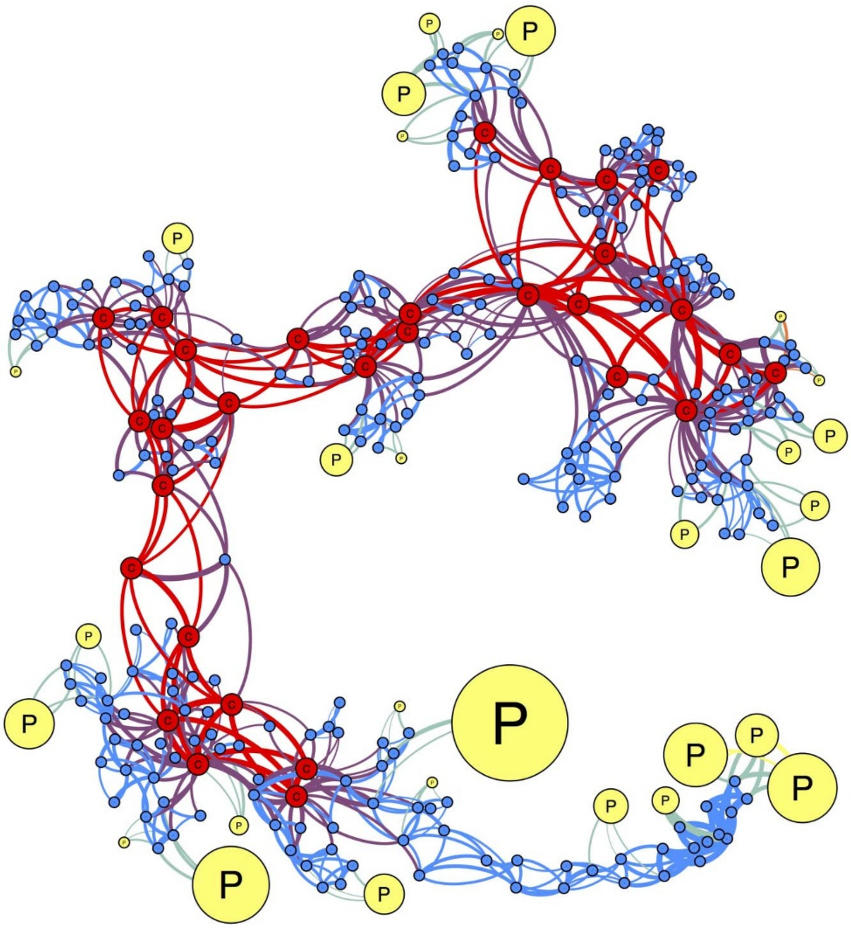

Figure 1 | Example of *PMFG*, a maximally filtered planar graph with vertices 300 stocks, selected among ordinary common shares listed in the American Stock Exchange market and edges associated with the structure of strongest correlations between stocks (in the time period from 1981 to 2010). A portfolio made of 30 peripheral stocks is represented by circles marked with "P"; their area is proportional to the Markowitz weights in the portfolio composition. Circles marked with "C" represent a basket of 30 central stocks. The thickness of the edges is proportional to the correlation coefficients. Names of the stocks corresponding to each vertex are provided in the supporting information.

For each day t, we constructed the *MST*, *PMFG* financial filtered networks by using the average correlations with shrinkage computed over the previous year; we then selected the m most peripheral stocks (with the largest values of $X + Y$) and set up portfolios with either uniform weights or Markowitz weights[7], with or without short-selling (in the present study this corresponds to a total of 7071×3 portfolios). For each portfolio we have observed the returns, defined as $r_t(\tau) = [Price(t+\tau) - Price(t)]/Price(t)$, over a year ($\tau = 1, .., 250$) following the investment date. The performance of each investment strategy is measured by computing the average $\bar{r}(\tau)$ and the standard deviation $s(\tau)$ of the returns over the 7071 investment dates. We have then chosen the 'signal-to-noise ratio' (also known as 'information ratio'), $\frac{\bar{r}(\tau)}{s(\tau)}$, as proxy for performance: good investment strategies must consistently produce high returns associated with small fluctuations being therefore characterized by large $\frac{\bar{r}(\tau)}{s(\tau)}$ ratios; conversely, bad investment strategies produce small returns and larger fluctuations (larger risk) yielding small signal-to-noise ratios.

Before presenting the results on portfolio performance, let us here address the question whether risk is uniformly spread through individual vertices of financial filtered graphs. To this purpose we measured the correlations between the centrality indexes and the signal-to-noise ratios of each stock finding that there is no significant relation between the two. Therefore, we can conclude that, at an individual stock level, the risk is uniformly distributed across financial graphs. In the following sections we shall see that this conclusion is reversed once we consider groups of stocks (i.e. portfolios) rather than individual stocks.

Results

Average performance of different portfolios. We measured the performance of portfolios composed of the $m = 5, 10, 20, 30$ most peripheral stocks within *MST* and *PMFG* graphs (m stocks with largest $X + Y$) and compared it with that of portfolios made of the m most central stocks (m stocks with smallest $X + Y$); we also considered portfolios of m stocks chosen at random and m stocks characterized by the best performance over the period preceding the investment date. All these portfolios were also compared with the performance of the whole 'market' of the 300 stocks. Figure 2 reports results for the signal-to-noise ratios for the case of a basket of m stocks from the *PMFG* where the relative contribution of each stock to the portfolio is weighted uniformly. We can observe that peripheral portfolios systematically outperform central ones, and also outperform portfolios made of randomly chosen stocks and those made of stocks achieving the best performance over the previous period. Notably, the performance of peripheral portfolios is comparable -and often better- than the market performance

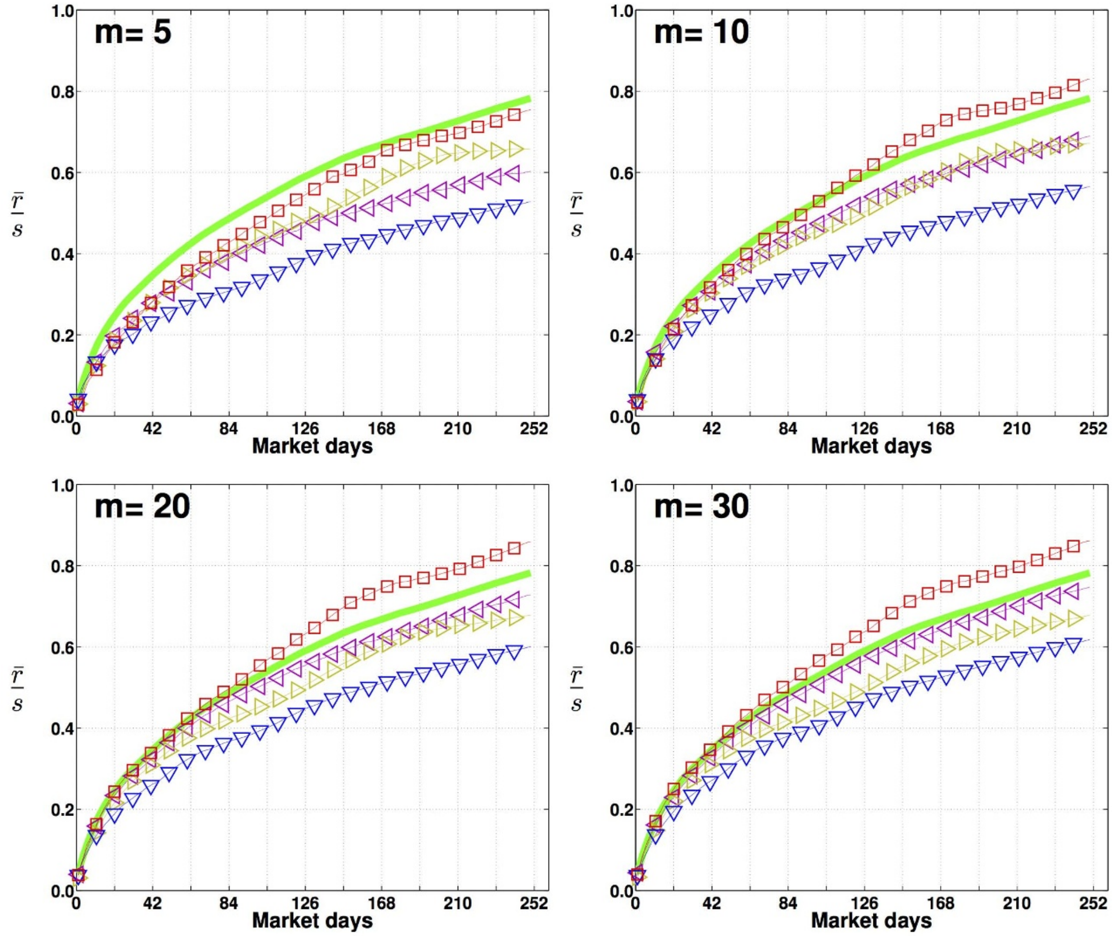

Figure 2 | **Demonstration that portfolios made with peripheral stocks (□) perform better than portfolios made with central stocks (▽).** Portfolio sizes are respectively $m = 5, 10, 20, 30$ stocks; weights are uniform. The plots report the 'signal-to-noise ratio' $\frac{\bar{r}(\tau)}{s(\tau)}$ (average return divided by its standard deviation) for $\tau = 1, .., 250$ days following the investment day. The performance is compared with: (◁) portfolios made of m randomly chosen stocks; (▷) portfolios made with the m stocks that have achieved the best performance over the period preceding the investment date. The thick line is a 'market portfolio' made by taking all 300 stocks.

obtained from all 300 daily stocks. Let us stress that peripheral portfolios, with as little as five stocks, already achieve competitive outcomes. Very similar results are obtained for portfolios set up by using the *MSTs* instead of the *PMFGs*, however the portfolio compositions are different revealing that the two filtered graphs provide alternative investment options (further details are provided in the supporting information). We also considered portfolios weighted by using the Markowitz method with and without short-selling. Figures 3 reports their performance in the case with no-short-selling; the case with short-selling is very similar and it is reported in the supplementary information (Figure S.3). Details on Markowitz portfolio optimization and discussion of portfolio variances are also reported in supplementary information (Sections S.6, S.7 Figures S.4, S.5 and S.6). We note that the results are similar to those with uniform weights, with 'peripheral' portfolios systematically outperforming portfolios of 'central', 'random' and 'best' stocks, and performing competitively with portfolios selected from the whole market. The main difference is that Markowitz weighting significantly improves the performance of all portfolios with the exception of central ones. In particular, the Markowitz method mostly improves the performance of the 'market' portfolio with all 300 stocks. However, it should be stressed that Markowitz solutions for a large number of stocks tend to be avoided by operators because a large system is harder to control and could become more costly to manage[2]. Furthermore, in the case of Markowitz portfolios with short-selling, we observed that the leverage, measured as the sum of all weights in absolute value, is large for 'market' portfolios of 300 stocks (290%). Conversely, Markowitz solutions for *PMFG* peripheral portfolios exhibit very limited leverage levels of: 100%, 102%, 109%, 116%, 124% respectively for $m = 5, 10, 20, 30, 40$. Therefore *PMFG* peripheral portfolios are less exposed to risk, because leverage itself is a measure of risk with high leverages making the investment more vulnerable to large losses. In addition we note that, for the case of Markowitz solutions with all 300 companies and no short-selling, the average number of non-null weights is 32 (with interquartile range between 24 and 41). Analogous averages for *PMFG* peripheral portfolios, for $m = 5, 10, 20, 30, 40$, are respectively equal to 4.9, 9.1, 15.5, 19.8, 22.9, with very narrow interquartile ranges, showing that the basket of stocks selected from *PMFG* peripheries is already well balanced also from the Markowitz perspective. *PMFG* peripheral portfolios are also characterized by small average 'maximum weights'; in the case with no short sales these are 0.42, 0.30, 0.23, 0.21, 0.19 respectively for $m = 5, 10, 20, 30, 40$ with narrow confidence intervals. The case with short sales is identical to all practical effects. From these results, we also conclude that a reasonable number of peripheral companies should be around $m = 20$, ensuring in this way competitive signal-to-noise ratios, together with few non-null Markowitz weights with

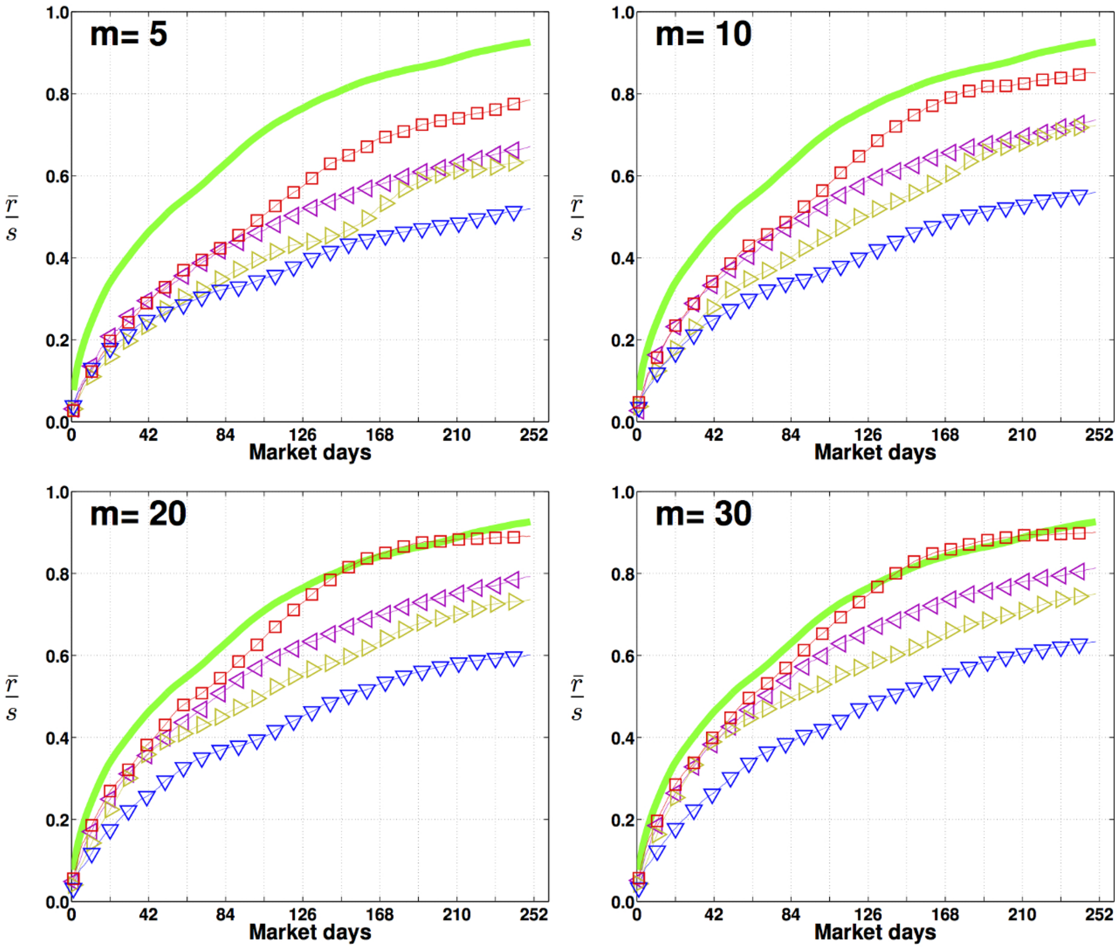

Figure 3 | Demonstration that portfolios made with peripheral stocks (□) perform better than portfolios made with central stocks (▽) also in the case of weights obtained by solving the Markowitz problem with no short-selling. Portfolio sizes are respectively $m = 5, 10, 20, 30$ stocks. The plots report the 'signal-to-noise ratio' $\frac{\bar{r}(\tau)}{s(\tau)}$ (average return divided by its standard deviation) for $\tau = 1, .., 250$ days following the investment day. The performance is compared with: (◁) portfolios made of m randomly chosen stocks; (▷) portfolios made with the m stocks that have achieved the best performance over the period preceding the investment date. The thick line is a 'market portfolio' made by taking all 300 stocks.

Table 1 | Test of the hypothesis that yearly signal-to-noise ratio of peripheral portfolios is superior to that of central portfolios in sub periods of six months. Values are percentages of cases in which the hypothesis is not rejected by a t-test with 5% significance level. The total number of cases is $7071 - 124 = 6947$. Symbols in the first column refer to the portfolios weights: uniform (*u*), Markowitz with no short-selling (*ns*) and Markowitz with short-selling (*s*). Cases with $m = 5, 10, 20, 30, 40$ are reported

	5 stocks	10 stocks	20 stocks	30 stocks	40 stocks
u	50.14%	53.94%	58.33%	62.08%	65.99%
ns	57.43%	56.48%	67.99%	73.07%	73.46%
s	60.46%	65.58%	70.92%	69.61%	67.99%

relatively small maximum weights and small leverages in case of short sales. A comparison with the performance of the benchmark S&P 500 Composite index reveals that *PMFG* peripheral portfolios have larger average yearly excess returns (the difference between portfolios and benchmark returns[2]) than the central ones and comparable values with the market ones (see supporting information). Similarly the Sharpe Information Ratio (information ratio of the excess yearly returns) also shows that *PMFG* peripheral portfolios perform better than the central ones (see also supporting information, S.2, S.3). Consistently, the 'beta coefficients' (the slope of the best-fit regression of the excess returns over the 'risk free' rate[2]) reveal an anti-cyclic pattern for the excess returns of *PMFG* peripheral portfolios with respect to the benchmark S&P 500 Composite index, i.e. they increase when the market goes down and vice-versa, thus showing a fair ability to absorb the financial systematic risk (see supporting information).

Performance over shorter sub-periods. The previous results demonstrate that -on average, over the whole period- the performance of portfolios made of peripheral stocks is superior to that of portfolios made of central stocks. We now investigate whether these good outcomes are also consistently obtained within shorter sub-periods. To this purpose we computed, for each day t, the yearly returns in the preceding six months (i.e. 125 returns $r(s, 250)$ in the period $s = \{t - 124, \ldots, t\}$) and performed an out-of-sample t-test to measure the likelihood that in the following period peripheral portfolios are superior to central portfolios. The proportions of cases in which the signal-to-noise ratio of peripheral portfolios is significantly larger than that of central portfolios, at a 5% significance level, are reported in Table 1. These results reveal that, indeed, in most sub-periods, portfolios made of peripheral stocks have better performances than portfolios made of central stocks. This is consistent with what obtained for the whole period. We note that differences are less accentuated in portfolios with uniform weights and more evident when weights are determined with Markowitz solutions.

Likelihood of negative returns. Another measure of risk is the likelihood of negative returns - which investors wish as tiny as possible. We therefore computed the empirical probability of non-negative returns after six and twelve months from the date when the investment was initially made. We found that investing in the peripheries of financial filtered networks provides a larger likelihood of achieving positive results after both six and twelve months with respect to investments in central stocks. This is consistently verified for portfolios of various sizes from 5 to 40 stocks and for both *PMFG* and *MST* graphs. The results are shown in Fig. 4 where one can note that investments, with portfolios of only 20 stocks selected from the peripheries of the financial filtered graphs, have a comparable -and sometimes higher- likelihood of positive returns with respect to investments made of all 300 stocks in the market. This is consistent with the signal-to-noise ratios discussed previously.

Likelihood of higher returns. We have so far established that peripheral portfolios are exposed to lower risk than central portfolios. In an investor's perspective it is also important to establish whether or not peripheral portfolios can provide higher returns than other investments. For this purpose we tested the hypothesis that the difference between returns of peripheral and central portfolios is positive or null. Specifically, for each day, we performed an out-of-sample t-test on the yearly returns in the preceding 125 days. The results reveal that portfolios made of peripheral stocks consistently yield equal or better returns than

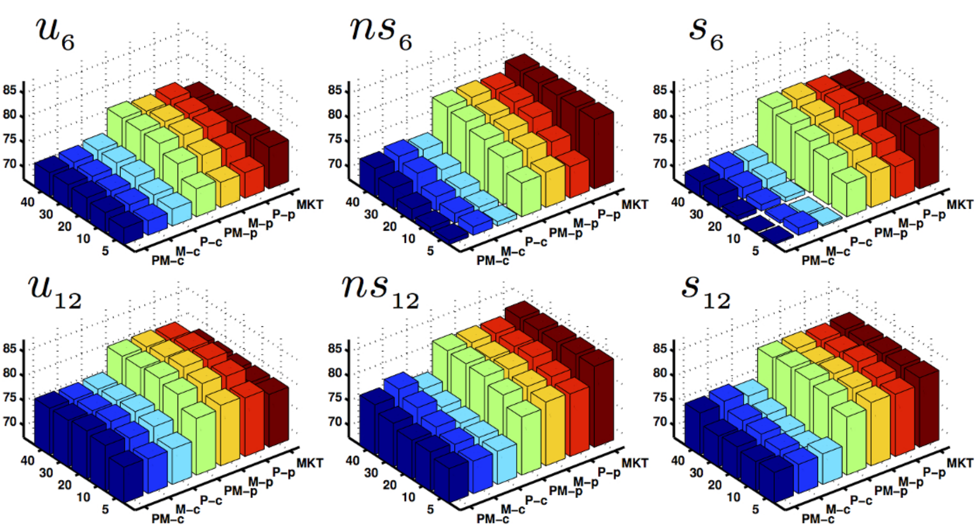

Figure 4 | Demonstration that peripheral portfolios have larger likelihood of non-negative returns than central portfolios. (Upper panel) probability of non-negative returns (expressed in per-cent values) after six months from the date when the investment was made; (lower panel) after one year from the date when the investment was made. Cases with uniform weights (u), Markowitz solutions with no short-selling (ns) and with short-selling (s) are shown. Investments based on portfolios of $m = 5, 10, 20, 30, 40$ stocks selected from central (c) and peripheral (p) regions of the financial filtered graphs *MST* (*M-c* and *M-p*), *PMFG* (*P-c* and *P-p*) and the combination of the two (*MST-PMFG*, i.e. *PM-c* and *PM-p*) are compared with the investment made over all the 300 stocks (*MKT*).

Table 2 | Test of the hypothesis that, over sub-periods of six months, yearly returns obtained with peripheral portfolios are larger or equal than that of portfolios set up with central stocks. Values are percentages of cases in which the hypothesis is not rejected by a t-test with 5% significance level. The total number of cases is $7071 - 124 = 6947$. Symbols in the first column refer to the portfolios weights: uniform (*u*), Markowitz with no short-selling (*ns*) and Markowitz with short-selling (*s*). Cases with $m = 5, 10, 20, 30, 40$ are reported

	5 stocks	10 stocks	20 stocks	30 stocks	40 stocks
u	62.26	62.50	63.06	62.65	63.29
ns	67.11	67.63	70.33	71.79	70.07
s	68.63	70.76	71.90	69.76	67.30

portfolios made of central stocks. Table 2 reports the percentage of cases in which the hypothesis is not rejected (i.e. peripheries give equal or higher returns than centers) for different weightings and for different sizes ($m = 5, 10, 20, 30, 40$). Significance level was set at 5%.

Portfolios from other regions of the financial filtered graphs. We also investigated other regions of the financial filtered graphs by looking at the positions of all companies in the plane defined by the axes $(X + Y)$ and $(X - Y)$. Specifically we investigated the four sides of the square of coordinates $A = (2, 0)$, $B = (1, 1)$, $C = (0, 0)$, $D = (1, -1)$. In this map the 'peripheral' regions used in the previous investment strategies are around the corner A and the 'central' regions lie around C. For each side ($\overline{AB}$, $\overline{BC}$, $\overline{CD}$ and $\overline{AD}$) we selected the m companies which lay closer to each of these sides and set up the optimal portfolios by using the same methodology described previously. We found that sides $\overline{AB}$ and $\overline{AD}$ perform better than $\overline{BC}$ and $\overline{CD}$ but worse than the 'peripheral' corner A; $\overline{AB}$ performs better than $\overline{AD}$ in terms of signal-to-noise returns but worse in terms of total returns. Overall, the results are analogous to those described previously for the central/peripheral (C/A) regions.

Discussion

We have shown that financial filtered graphs can be used to select portfolios with lower risk and better returns than those obtained with other traditional methods. This has been achieved by first defining suitable correlation matrices, then constructing *MST* and *PMFG* financial filtered graphs and finally establishing appropriate indices to select portfolios made of stocks located in either central or peripheral regions. We have quantified the investment performance by using a large range of measures, including: 'signal-to-noise' ratio between average returns and their standard deviations; portfolio variance; probability to obtain larger returns; likelihood of non-negative returns; average returns and Sharpe information ratio (see supporting information). All results consistently show that portfolios set up from a selection of peripheral stocks have lower risk and better returns than portfolios set up from a selection of central stocks. Poor performances of the central portfolios might be consequence of the fact that the center of the network is more likely to be subject to sudden perturbations due to the herd effect: during periods of booms and crashes the system gets highly correlated and investors simultaneously rush in the same direction, buying or selling, respectively. Hence, portfolios containing companies that are at the center of these irrational moods are more likely to carry larger risk. An efficient diversification is possible if the portfolio is composed of stocks characterized by both low correlations and high expected returns' signal-to-noise ratios. We have shown that these securities are located in the peripheries of the financial filtered graphs.

There is a large scope of applicability and testing for the present method within a variety of different domains including FX markets and the vast field of derivatives where it can be combined with traditional pricing methods. Further studies will focus on the application of a newly introduced clustering method[22] which can be used for further distinguishing between peripheral and central stocks in the portfolio selection. Another investigation will be dedicated to verify whether the risk of a company default is uniformly distributed across financial networks.

Table 3 | Correlation matrix for the rankings of the centrality indices (Degree *D*, Betweenness Centrality *BC*, Eccentricity *E*, Closeness *C* and Eigenvector Centrality *EC*[15]) calculated on *PMFG*

	C_D^w	C_D^u	C_{BC}^w	C_{BC}^u	C_E^w	C_E^u	C_C^w	C_C^u	C_{EC}^w	C_{EC}^u
C_D^w	1.00	0.97	0.82	0.94	0.37	0.23	0.37	0.33	0.34	0.36
C_D^u	0.97	1.00	0.85	0.97	0.33	0.22	0.35	0.32	0.35	0.36
C_{BC}^w	0.82	0.85	1.00	0.82	0.31	0.22	0.33	0.30	0.32	0.33
C_{BC}^u	0.94	0.97	0.82	1.00	0.35	0.25	0.37	0.35	0.35	0.36
C_E^w	0.37	0.33	0.31	0.35	1.00	0.94	0.85	0.84	0.70	0.71
C_E^u	0.23	0.22	0.22	0.25	0.94	1.00	0.81	0.81	0.65	0.66
C_C^w	0.37	0.35	0.33	0.37	0.85	0.81	1.00	0.99	0.91	0.92
C_C^u	0.33	0.32	0.30	0.35	0.84	0.81	0.99	1.00	0.91	0.92
C_{EC}^w	0.34	0.35	0.32	0.35	0.70	0.65	0.91	0.91	1.00	0.99
C_{EC}^u	0.36	0.36	0.33	0.36	0.71	0.66	0.92	0.92	0.99	1.00

Methods

Additional material can be found in supporting information.

Data and daily selection of 300 stocks. We studied all ordinary common shares in the American Stock Exchange market in the period from 1981 to 2010 for a total of $T = 7570$ market days (data from the *CRSP*[23], ordinary common shares of "Americus Trust Components, Primes and Scores", "Closed-end funds", "Real Estate Investment Trusts" have been excluded from the dataset). We performed our analysis on moving time-windows of $\Delta t = 250$ days (one market year). Contiguous missing prices for less than five consecutive dates have been replaced with the previous value and, for each day t, stocks with less than Δt contiguous observations until t and Δt after t were discarded (note that keeping these stocks does not affect significantly results[21]). For each market day we have then selected the first 600 stocks by capitalization. We further reduced the dataset by retaining only the top half 'best performing' subset of stocks over the previous Δt period. To this purpose, for each stock and for each time t we computed the daily returns $r(t, 1)$ and calculated their average, $\langle r(1)\rangle_{\Delta t}$, and their standard deviation, $s_{\Delta t}(1)$, over the previous Δt days. We then selected the half stocks with highest $\frac{\langle r(1)\rangle_{\Delta t}}{s_{\Delta t}(1)}$ ratios (i.e. those on average with the highest daily performance over the previous Δt days); leaving us with $N = 300$ stocks for each time. Note that the daily set of stocks changes very slowly, with the daily average replacement rate (ratio between number of new companies, from a day to the next, and total number of companies) being just 3.7%; the weekly average replacement rate 8.1%; the monthly rate 15.6%; and the yearly rate 58.4%. In terms of industrial sectors, our selection is not neutral, with stocks belonging to major industrial groups such as Electric, Gas, and Sanitary Services and Chemicals and Allied Products being most likely to be selected. With this procedure we considered a total of 2286 different stocks over the whole period.

Dependency measure. In order to reduce the excessive influence of remote events on present correlations, we used exponential weights (defined as $w_t = w_0 \exp\left(\frac{t-\tau}{\theta}\right)$, such that $w_t > 0$ and $\sum_{t=1}^{\tau} w_t = 1$) so that past observations count less than recent ones[8]. Here, $t = 1,2,\ldots,\tau$ and $\theta > 0$ is the characteristic time horizon. Weighted sample means, variance, covariance and correlation are defined from the weighted averages from: $\bar{f}^w(t) = \sum_{s=t-\tau+1} w_s f(r(s))$[8]. We used this exponentially smoothed averages to compute, for each t, weighted Pearson's correlation coefficients $R_{ij}^w(t)$ over a window of six months ($\tau = \theta = 125$). For each day t we monitored these correlations in the previous six months and we computed their average values with shrinkage[9]:

$$\bar{\mathbf{R}}_{ij}^w(t) = \frac{1}{2(\tau+1)}\left[\sum_{s=t-\tau}^{t} R_{ij}^w(s) + \sum_{i=1}^{j-1}\sum_{j=2}^{N}\sum_{s=t-\tau}^{t}\frac{2R_{ij}^w(s)}{N(N-1)}\right]. \quad (1)$$

The shrinkage significantly improves the numerical significance of the correlation matrix (the condition number[24] is reduced by two orders of magnitude).

Centrality and Peripherality measures. We computed the Degree (D), the Betweenness Centrality (BC), the Eccentricity (E), the Closeness (C) and the Eigenvector Centrality (EC)[15] for both weighted and unweighted graphs for both *MST* and *PMFG*. For the weighted degree (often called strength) and the weighted Eigenvector Centrality the weight between vertex i and vertex j is $1+\bar{\mathbf{R}}_{ij}^{w}$. Whereas, for the weighted Betweenness Centrality, Eccentricity and Closeness the weight is $\sqrt{2\left(1-\bar{\mathbf{R}}_{ij}^{w}\right)}$ (i.e. the Euclidean distance). These measures of centrality/peripherality have been sorted, respectively, in descending order for the centrality measures (D, BC, EC) and in ascending order for the peripherality measures (E, C). Then, for each measure, tied ranks (or midranks)[25] have been calculated so that central vertices have been assigned higher rankings and peripheral vertices lower ones. Note that very similar rankings are found for both *PMFG* and *MST*. All these measures of centrality/peripherality are clearly not independent and indeed they all result positively correlated among each other. The structure of their correlation matrix, for 300 NYSE firms over the period 2001–2003, is reported in Table 3 for the case of *PMFG*. We note that the matrix has two diagonal blocks containing high correlation values (all larger than 0.65) while the outer block contains low values (all smaller than 0.37) indicating the presence of two clusters, made respectively of the rankings of D and BC and the rankings of E, C and EC, which are strongly correlated within their cluster and scarcely correlated between clusters. Therefore, we defined two combined measures as follows: $X=\frac{c_D^w+c_D^u+c_{BC}^w+c_{BC}^u-4}{4\times(N-1)}$ and $Y=\frac{c_E^w+c_E^u+c_C^w+c_C^u+c_{EC}^w+c_{EC}^u-6}{6\times(N-1)}$, where we denoted with c_D^w the tied ranking of the weighted Degree (D) and with c_D^u its unweighted counterpart; for all other measures, we used the corresponding symbol (BC, E, C, EC) instead of D. These two hybrid measures distinguish between highly connected vertices connected to other highly connected vertices (small X, small Y); highly connected vertices connected to scarcely connected vertices (small X, large Y); scarcely connected vertices connected to highly connected vertices (large X, small Y); scarcely connected vertices connected to scarcely connected vertices (large X, large Y). We therefore considered as hybrid measures of centrality the sum and the difference between X and Y. The value of $X+Y$ is small for central vertices and large for peripheral vertices; whereas the value of $X-Y$ is large if the vertex has few important connections and it is small if it has many unimportant connections. A Matlab code to calculate centrality and peripherality indices is reported in the supporting information. The choice of a hybrid measure is heuristic, based on the observation that by using it we consistently obtain better performing portfolios than those from the centrality and peripherality measures in isolation or in different combinations. A comparison between performances of portfolios constructed by using alternative combinations of centrality measures is reported in the supplementary information S.4 (Figures S.7–10; see also reference[21]). Let us stress that, although the use of the proposed hybrid measure gives best performances, the main result of this paper, that investments in peripheral equities are better than investments in central ones, is consistently obtained for all centrality measures[21].

1. Meucci, A. *Risk and asset allocation*, (Springer Berlin, 2009).
2. Hull, J. C. *Options, Futures, and Other Derivatives*, (Prentice Hall, 2012).
3. Mantegna, R. N. Hierarchical structure in financial markets. *European Physical Journal B* **11**, 193–197 (1999).
4. Tumminello, M., Aste, T., Di Matteo, T. & Mantegna, R. N. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences* **102/30**, 10421–10426 (2005).
5. Aste, T. & Di, M. T. Dynamical networks from correlations. *Physica A* **370**, 156–161 (2006).
6. Tumminello, M., Lillo, F. & Mantegna, R. N. Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization* **75**, 40–58 (2010).
7. Markowitz, H. Portfolio selection. *The Journal of Finance* **7**, 77–91 (1952).
8. Pozzi, F., Di Matteo, T. & Aste, T. Exponential Smoothing Weighted Correlations. *European Physical Journal B* **85**, 175 (2012).
9. Ledoit, O. & Wolf, M. Honey, I shrunk the sample covariance matrix, UPF Economics and Business Working Paper 691 (2003).
10. Aste, T. Matlab code for computation of *PMFG*s, Matlab File Exchange 27360.
11. Sabidussi, G. The Centrality Index of a Graph. *Psychometrika* **31**, 581–603 (1966).
12. Freeman, L. C. Centrality in networks: I. Conceptual clarification. *Social Networks* **1**, 215–239 (1979).
13. Bonacich, P. Power and centrality: a family of measures. *American Journal of Sociology* **92**, 1170–1182 (1987).
14. Borgatti, S. P. Centrality and network flow. *Social Networks* **27**, 55–71 (2005).
15. Newman, M. E. J. *The mathematics of networks, in S. N. Durlauf, L. E. Blume* (Eds.), The New Palgrave Encyclopedia of Economics, 2nd ed. (Palgrave Macmillan, Basingstoke, 2008).
16. Caldarelli G. *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford Univesity Press, Oxford 2007).
17. U. S. Securities and Exchange Commision, U.S. Standard Industrial Classification codes, http://www.sec.gov/info/edgar/siccodes.htm (last accessed 22/03/2012).
18. Pozzi, F., Di Matteo, T. & Aste, T. Centrality and Peripherality in filtered graphs from dynamical financial correlations. *Advances in Complex Systems* **11**, 927–950 (2008).
19. Di Matteo, T., Pozzi, F. & Aste, T. The use of dynamical networks to detect the hierarchical organization of financial market sectors. *European Physical Journal B* **73**, 3–11 (2010).
20. Pozzi, F., Aste, T., Shaw, W. & Di Matteo, T. The use of topological quantities to detect hierarchical properties in financial markets: the Financial sector in NYSE, Proceedings of 10th WSEAS international conference on Mathematics and computers in business and economics, Recent Advances in Computer Engineering 301–304 (2009).
21. Pozzi, F. Filtering financial networks and optimal portfolio selection. Ph.D. Thesis (The Australian National University, Canberra, 2013).
22. Song, W.-M., Di Matteo, T. & Aste, T. Hierarchical information clustering by means of topologically embedded graphs. *PLoS One* **7**, e31929 (2012).
23. Wharton Research Data Services, https://wrds-web.wharton.upenn.edu/wrds/, CRSP US Stock Database (last accessed 22/03/2012).
24. Anderson, E. *et al. LAPACK User's Guide, Third Edition* (SIAM, Philadelphia, 1999).
25. Gibbons, J. D. & Chakraborti, S. *Nonparametric statistical inference, 5th edition* (Chapman & Hall/CRC, 2011).

Acknowledgments

We thank the School of Finance at The Australian National University for providing data queried from the *CRSP* US STOCK database accessed via the subscription to the Wharton Research Data Services.

Author contribution

All the authors contributed equally to the manuscript writing the paper, revising it and preparing the figures.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Pozzi, F., Di Matteo, T. & Aste, T. Spread of risk across financial markets: better to invest in the peripheries. *Sci. Rep.* **3**, 1665; DOI:10.1038/srep01665 (2013).