The Network of U.S. Mutual Fund Investments: Diversification, Similarity and Fragility throughout the Global Financial Crisis

Danilo Delpini,[1,2*] Stefano Battiston,[3] Guido Caldarelli,[2,4] Massimo Riccaboni[2,5]

[1]Dept of Economics and Business, University of Sassari, Sassari, Italy

[2]IMT Institute for Advanced Studies, Lucca, Italy

[3]Dept of Banking and Finance, University of Zurich, Zurich, Switzerland

[4]ISC-CNR Uos "Sapienza", Roma, Italy

[5]Dept of Managerial Economics, Strategy and Innovation, Katholieke Universiteit Leuven, Leuven, Belgium

[*]E-mail: ddelpini@uniss.it

Network theory proved recently to be useful in the quantification of many properties of financial systems. The analysis of the structure of investment portfolios is a major application since their eventual correlation and overlap impact the actual risk diversification by individual investors. We investigate the bipartite network of US mutual fund portfolios and their assets. We follow its evolution during the Global Financial Crisis and analyse the interplay between diversification, as understood in classical portfolio theory, and similarity of the investments of different funds. We show that, on average, portfolios have become more diversified and less similar during the crisis. However, we also find that large overlap is far more likely than expected from models of random allocation of investments. This indicates the existence of strong correlations between fund portfolio strategies. We introduce a simplified model

1

of propagation of financial shocks, that we exploit to show that a systemic risk component origins from the similarity of portfolios. The network is still vulnerable after crisis because of this effect, despite the increase in the diversification of portfolios. Our results indicate that diversification may even increase systemic risk when funds diversify in the same way. Diversification and similarity can play antagonistic roles and the trade-off between the two should be taken into account to properly assess systemic risk.

Introduction

One of the most important issues that the Global Financial Crisis (GFC) of 2007-2008 has highlighted concerns the effects on systemic risk of the increasing interdependence both between large institutional investors and among global assets. On the one hand, large institutional investors allow for a better diversification of individual risk: the larger the number of different assets in a portfolio, the smaller the fraction of an idiosyncratic shock an investor has to bear. On the other hand, the GFC has shown that cross-sectional dependencies between assets can cause idiosyncratic shocks (i.e. related to the distress/bankruptcy of a single specified asset) to spread, ultimately substantially threatening the stability of the entire financial system. According to classical investment portfolio theory, risk depends on the share of individual stock holdings and the variance–covariance matrix among its holdings [1]. Hence, theoretical models imply that a portfolio should be fully diversified to reduce risks (unsystematic risks), but how to construct a well-diversified portfolio still remains to be studied. There is consensus in the literature that a strong risk reduction of holdings can be realized by increasing the number of assets in a portfolio [2, 3, 4]. Indeed, between 1997 and 2012, assets in the equity, balanced, and fixed income mutual funds have increased by more than 400 percent [5]. The efficacy of diversification strategies depends, however, on market conditions and systemic risk that may

moderate the relationship between diversification, fund performance, and risk. Recently, it has been shown that the benefit of diversification increases within higher market volatility conditions, such as the GFC of 2007-2008 meaning that the number of stocks needed to achieve a well-diversified portfolio increases under those market conditions [6]. Unfortunately, the implications of such an increase of portfolio diversification during crises are not fully understood. Moreover, the empirical evidence accumulated during the GFC has raised legitimate doubts on the effectiveness of portfolio diversification strategies to reduce risk.

Systemic risk in financial systems has been increasingly investigated through the lens of network theory [7, 8, 9, 10, 11, 12, 13, 14]. Most of the work so far has focused on the network of interbank loans, even though there is a paucity of data about the real-world structure of financial networks. Interestingly, it has been found that, when the magnitude of negative shocks is large and the network is scale-free, a more densely connected financial network (corresponding to a more diversified pattern of investment) serves as a mechanism for the propagation of shocks, leading to a more fragile financial system, thus increasing systemic risk.

Depending on the level of similarities across mutual funds, increasing portfolio diversification during a crisis might increase the cross-correlation among assets thus amplifying systemic risk. Therefore, the exact role played by the evolution of the financial network during crisis in creating systemic risk remains, at best, imperfectly understood. The same holds for the role of global fund managers which appears to have been studied mostly in simulated scenarios [15]. So far, most of the papers in the literature assume that the relationship between diversification and risk is constant in the long run, regardless of market conditions. We contribute to fill this gap in the literature by analysing the bipartite network of US mutual funds over time and throughout the Global Financial Crisis.

Here, we provide an extensive characterization of the mutual fund universe covered by the CRSP Survivor-Bias-Free Mutual Funds Database in 2005-2010. In particular, we consider the

bipartite network of portfolio holdings and we perform a temporal analysis of its topology, focusing on the interplay between the complementary notions of diversification and similarity of portfolio investments. We study how the systemic fragility of the system depends on the overlap between portfolios due to correlated investment strategies. This is done through simulations of a simple dynamics of distress propagation and through comparison with null models of random investments. The fund holding network has been considered previously in [16] where the correlation between changes in a firm's position in the network and future stock market performance is considered, and in [17] where the CRSP database was studied to detect possible conflicts of interest in the strategies of multi-fund managers. In [18] the size distribution of funds has been investigated in detail, while in [19] a size-dependent model of fund growth has been proposed to explain its shape. However, to the best of our knowledge, this is the first attempt to provide a comprehensive description of this system across years, from the point of view of network science, and to disentangle the roles of diversification and similarity in determining distress propagation and systemic fragility.

Results

We analyse the mutual funds market as a bipartite network where one of the two sets is formed by the US mutual fund portfolios and the other set is made by the assets in those portfolios. A detailed formalization is provided in Materials and Methods. Within this framework, we perform an extensive statistical analysis of portfolio diversification from a twofold perspective. On one side, we consider the usual notion of diversification *within* a single portfolio, dating back to Markowitz [20]. This depends on the number of different assets in the portfolio, as well as on their weights, and on the correlation between asset price variations. On the other side, we add a second dimension to the analysis, by studying diversification *across* portfolios, defined as the degree of similarity of the assets held by mutual funds and measured by suitable overlap

4

indices. We compare the real data with two null models of random allocation of investments to test for heterogeneity and non trivial correlations between portfolios. Finally, introducing a simple model of distress propagation and exploiting the random benchmarks, we explore the connection between similarity of portfolios, distress propagation and systemic fragility.

The above characterization is performed with special attention to the evolution of the network throughout the GFC. In particular, we show how the systemic fragility of the network has changed. As it is well known, the peak of the crisis has been reached in 2008, culminating in the collapse of Lehman Brothers in September. In order to compare the structure of the network over time, we select three *reference quarters*, 2006Q2, 2008Q2, 2010Q2, where the notation "2006Q2", for instance, stands for the second quarter of year 2006. Our goal is to identify three moments in time *before*, *during* and *after* crisis.

Topology of fund holdings, concentration and similarity

In this Section we investigate diversification within and across portfolios, through the lens of network science. These concepts are related to the *degree* of funds and assets in the holding network.

The degree k_i of fund i represents the number of different assets held at a given moment in time, while the degree k_α of an asset represents the number of funds that hold asset α in their portfolio.

The probability distributions $p(k_{i,\alpha})$ for the fund and asset degree, respectively, are shown in Fig. 1 for the selected reference quarters. The probability densities extend over several orders of magnitude in both cases, with fat tails that are the typical signature of topologically heterogeneous systems. This means that most funds manage portfolios of just few tens of different assets while the average degree is in the hundreds. However, there are funds for which the number of assets is larger than the average by orders of magnitude (hubs). In between these two extremes
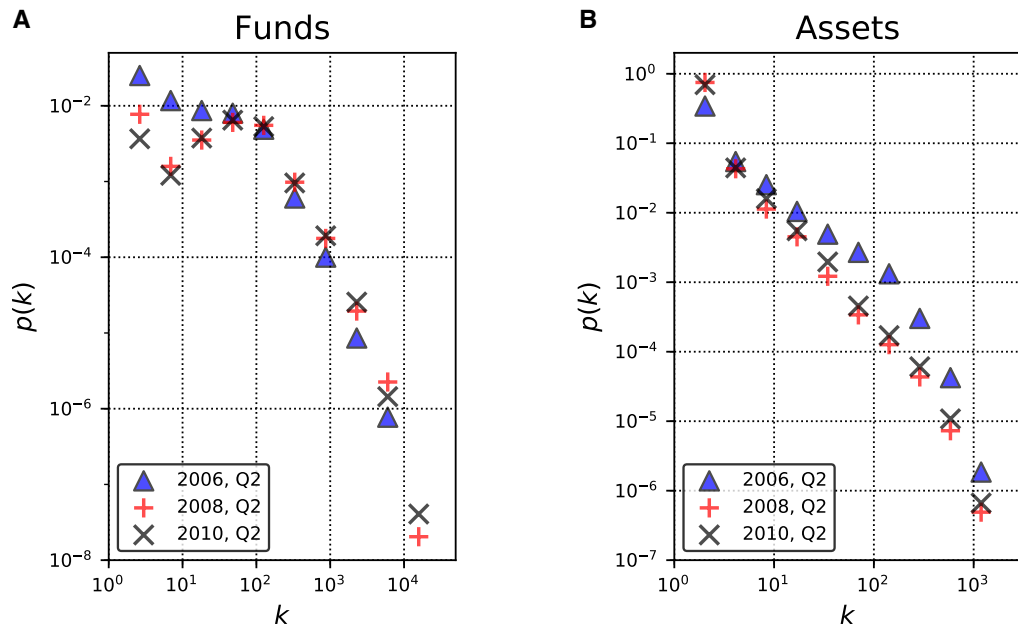
5

Figure 1: **Degree distributions.** Probability density functions of fund degree (**A**) and asset degree (**B**) for the three reference quarters. Most funds invest in few tens of different assets but some manage very diversified portfolios, with thousands of different holdings. The assets present similar heterogeneity. During the crisis, the probability of holding many assets has increased, while the probability for an asset to be held by many funds has fallen.

we find a wide spectrum of intermediate investment strategies. A similar heterogeneity characterizes the assets: most assets are found in the portfolios of few funds, but some assets enjoy huge popularity and are held by almost every fund.

During crisis, the tail of the distribution for funds has grown "fatter", while an opposite behavior is observed for assets. We reckon that, during crisis, the probability of finding portfolios with a small degree has decreased and, correspondingly, the probability for a fund to invest in a large number of different financial products has increased. At the same time, the probability of finding the same asset in many portfolios has become smaller, suggesting that the average overlap between portfolios has decreased.

Fund diversification

Fund-degree provides a first topological notion of diversification of an individual portfolio. However, from a financial point of view we must take into account the weights of each asset in a portfolio and consider a specialized index of concentration.

The Herfindahl–Hirschman index is a typical measure of concentration [21, 22, 23] and we use its inverse to compare the diversification of mutual fund portfolios. Precisely, we consider the following index

$$h_i = \left(\sum_\alpha w_{i\alpha}^2 \right)^{-1}. \tag{1}$$

If a fund manager invests primarily in a single asset, h_i will be close to one; the opposite case is represented by a uniform investment, with the same fraction of portfolio wealth invested in each asset, when we have $h_i = k_i$. In this sense, h_i can be regarded as the number of leading assets in the portfolio. We will refer to the quantity defined in Eq. (1) as the "inverse Herfindahl index".

The larger the value of h, the more the portfolio can be considered as diversified. This notion of diversification does not take into account how similar are different portfolios. Even though diversification is believed to lower market risk in a portfolio, we will show that a systemic risk component arises since portfolios tend to diversify in the same way. This adds a second systemic dimension to the notion of diversification, that is of major interest in a network perspective.

Portfolio similarity

Portfolio similarity depends indeed on two factors: the number of assets they have in common, and the similarity of the weights attached to common assets. Given two funds i and j, let P_i and P_j be the sets of the indices of the assets in their portfolios, that is $P_i = \{\alpha_l^{(i)}\}_{l=1,\dots,k_i}$ and $P_j = \{\alpha_l^{(j)}\}_{l=1,\dots,k_j}$, with $\alpha_l \in [1, N_a]$. We indicate as $|P|$ the cardinality of the set of assets P. A measure of overlap between the various choices is provided by the *Jaccard Index* [24], which

is the size of the intersection of the two sets divided by the size of their union:

$$J_{ij} = \frac{|P_i \cap P_j|}{|P_i \cup P_j|}.$$

(2)

This shows which percentage of the assets included in P_i or P_j belongs to both. When two portfolios contain exactly the same assets, the Jaccard Index equals one, $J = 1$.

We notice that asset degree is connected to the notion of overlap between portfolios. Indeed, the larger the degree of an asset, the more "popular" it is in investment strategies and the larger the probability of finding it in different portfolios.

The Jaccard index refers to a topological notion of similarity but portfolio weights must be taken into account to provide a more relevant measure. Indeed, when two funds invest in the same assets, their portfolios should be considered dissimilar in financial terms if the distribution of the weights are different. To capture this, we sum the smallest of the weights attached to each asset in the intersection and multiply this by the Jaccard index, obtaining a *Similarity Index* [16]. Formally, the latter can be defined as follows:

$$s_{ij} = J_{ij} \times \sum_{\alpha \in P_i \cap P_j} \min\left(w_{i\alpha}, w_{j\alpha}\right).$$

(3)

This index is now equal to one for two portfolios that contain exactly the same assets in exactly the same proportions, and it will be smaller than one otherwise. When the two portfolios do not overlap at all its value is zero.

In Fig. 2 we show the level of concentration and similarity in the network over time, measured by the average of the indicators introduced before. A considerable increase of diversification has taken place during the crisis, with an increment of 27% in the h-index in 2007Q2–2008Q2. Conversely, the average amount of overlap between two portfolios has reduced, with percentage variations of -13% and -17% for the Jaccard index and the similarity index respectively. These results support the scenario outlined from the analysis of degree distributions:
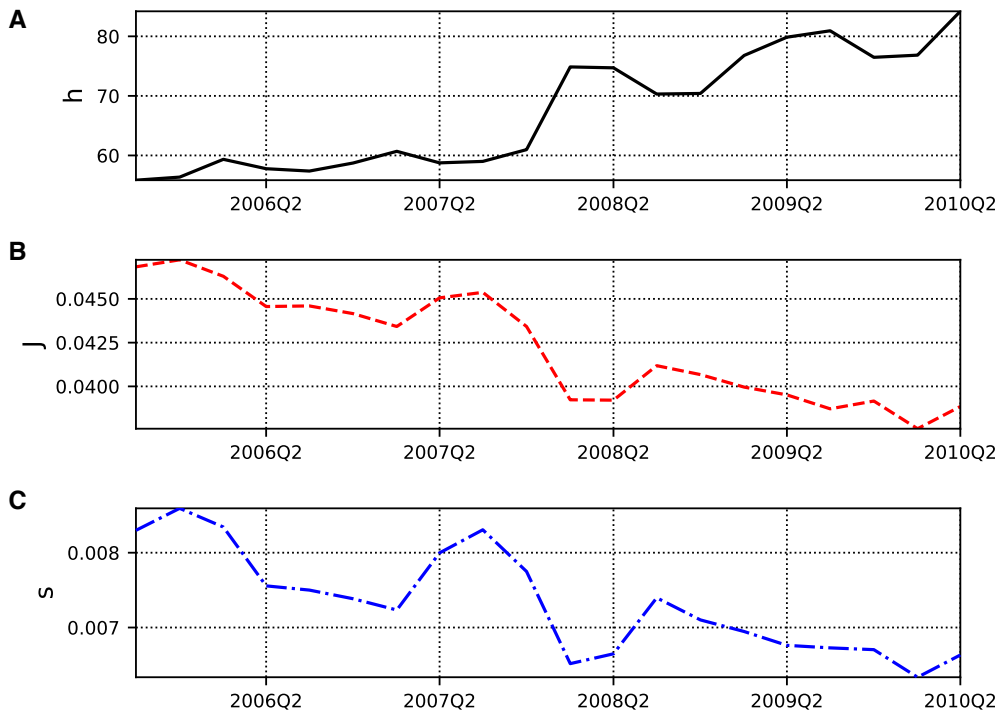
Figure 2: **Portfolio diversification and similarity over time.** Average values of Herfindahl index (**A**), Jaccard index (**B**) and similarity index (**C**). During the global financial crisis, the level of diversification within portfolios has increased while the level of similarity across portfolios has reduced.

during crisis, fund managers raised the level of diversification within their portfolios; at the same time, investment decisions of individual managers are more "diversified" from each other. This picture is also in good agreement with trends in the basic quantities that define the topology of the network (see the summary statistics reported in Table S1 in the Supplementary Materials section). During the crisis the average degree of funds has nearly doubled. More evidently, the total number of assets in the network has grown by an order of magnitude and the network density has reduced correspondingly. The growth in the number of assets in mutual funds during the crisis has been recently documented also by using other data sources, such as the Pensions and Investments survey [5]. It is worth noting that the enlargement of the set of different assets is crucial for understanding the opposite trends of diversification and similarity in Fig. 2. Indeed,

9

if the number of available asset stayed the same, we would expect an increase of diversification of the portfolios to be reflected, to some extent, in an increase of their overlap.

From Fig. 2 we can also observe that the timeline of the similarity s closely mimics the Jaccard index J. This suggests that, on average, the similarity between two portfolios in the holding network is mainly determined by the commonality of their assets. This was only partly predictable. Indeed, conditionally on the weights $w_{i,\alpha}$, the value of s is proportional to J (see Eq. (3)). However, in principle the two indicators could have different behaviors and we could imagine cases where, for instance, the relative size of the intersection of two portfolio increases but the value of s decreases if the weights attached to common assets are tuned suitably.

Comparison to null models

We analyse the level of heterogeneity and information content in the network of holdings by comparing the structure of the real network with two different "null" models. These models are obtained through randomization of the original investments by two alternative strategies.

In the first case, we start from a graph with N_f funds, N_a assets and no links, and then we reassign the original E edges (with their original market values attached) drawing at random and without repetition from the possible fund–asset pairs. Naively, we can think of this null model as a representation of a scenario where portfolio managers choose at random in which assets to invest and how much money to put on each of them. In the text and the figures we will identify this null model with the abbreviation *Rnd-1*.

In the other case, we perform the randomization subjected to the constraint that the degree and portfolio weights of each fund do not change. This is accomplished by reassigning the original investments of each fund to randomly selected assets (by construction, this strategy preserves the original value of the Herfindahl index of portfolios). This model represents a network where fund managers set the level of diversification in advance; that is, they decide the

number of distinct assets to invest in, as well as the relative proportions, and they then choose the assets at random. We indicate in the following with the abbreviation *Rnd-2* the second null model. Since values of portfolio concentration in this model are the same as in the original network, we shall exploit it to separate the contribution of similarity to systemic fragility. In summary, while the number of edges is preserved in the both cases, the degree sequence of funds is preserved only in the second one.

In Fig. 3 we compare the complementary cumulative distribution functions ($CCDF$) of the different indices (h, J, s) in a snapshot of the real network (2006Q2) compared with their distribution measured in the two null models. The $CCDF(x)$ gives the probability of finding a value of the index that is greater or equal to x. The same analysis done in the other quarters provides similar results. In particular, we find that the distribution of h provides a clear picture of the system's heterogeneity. In the real network, $p(h)$ decays slowly and the number of effective assets in portfolios ranges from few to thousands. Conversely, the distribution for model Rnd-1 is concentrated and the probabilities of large h are smaller by orders of magnitude. In the real case, the average value of leading assets in a portfolio is $\bar{h} \approx 58$, while it is only 9 for model Rnd-1. Similar considerations apply when we inspect the probability distributions of J and s. In these cases, a comparison is also performed with model Rnd-2, which retains the same values of diversification of the original network. We see that large values of the indices are more likely to occur for model Rnd-2 than Rnd-1, but both models fail to capture the probability of very large overlap that characterizes the real network. In particular, only in the real case we find pairs of almost identical portfolios. As a matter of fact, community detection on the network reveals the tendency of funds to cluster in small groups of portfolios with very similar structures.

As far as diversification of individual portfolios is concerned, the result of the comparison supports intuition. Fund managers are professional investors whose diversification strategies cannot be reduced to random selection of assets. As for the overlap of different portfolios, we
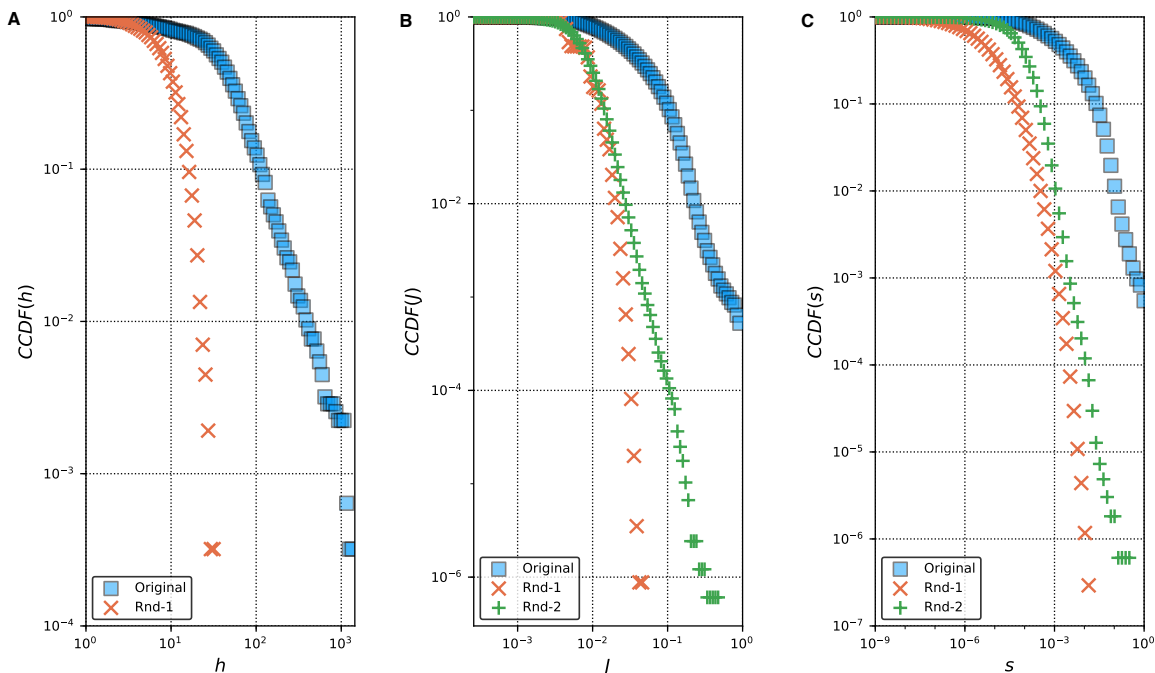
Figure 3: **Diversification and similarity: comparison to the null models.** Complementary cumulative distribution functions of Herfindahl index, h, (**A**), Jaccard index, J (**B**) and similarity index, s (**C**). The real network in 2006Q2 is compared to the null models as described in the text. The latter ones fail in reproducing the heterogeneous behaviour observed in the real systems. In particular, large values of the number of effective assets (h) and of the similarity between portfolios (s) are much more likely in the real world than in the randomised sets.

conclude that large similarities are considerably more likely than could be expected by chance. This result still holds when the values of h of real portfolios are kept fixed during the randomization procedure. This indicates that the effects of diversification *per se* are not sufficient to explain the probability of large similarity observed in the network and suggests that investment strategies of different funds can be strongly correlated. Many different mechanisms have been suggested in the literature to account for such a high degree of similarity across portfolios including connections between mutual fund managers and corporate board members, herding behavior and imitation of successful diversification strategies. We will show that such corre-

lation between portfolios contributes to a large extent to the system's vulnerability to financial shocks.

Systemic risk and similarity

To test the effects of similarity on the network fragility, we consider a basic model of distress propagation. An idiosyncratic shock in the value of an asset spreads across the network due to common exposures of portfolios (more details are provided in Materials and Methods). The shock propagates because individual investors sell their portfolio shares in response to a drop in the value of the fund they invest in. Selling shares entails selling the individual assets in the portfolio. This process, produces a negative effect on the market value of assets being sold and triggers a new round of losses for the portfolios owning those assets. This is, of course, a simplified view of the network dynamics which does not consider portfolio reallocations. However, reallocation of mutual fund investments is a slow process which might only partially ameliorate the negative impact of the asset sell-off process.

While being an approximation, the assumption of a static network can be justified with the following arguments. First, we expect portfolio reallocations by fund managers to take place on a time scale larger than the one of asset price movements on the market and of buying/selling decisions of individual investors. Moreover, keeping the topology fixed, we can provide a simple measure of fragility for each snapshot of the network and study its evolution during crisis.

In the context of such dynamics, we measure network fragility by considering the total percentage "loss of value" of the system. Since the total loss depends strongly on the asset receiving the initial shock, we compute the average D of the losses induced by a shock to one of the "top assets". We define the latter as the assets whose market value in the portfolios of all funds is larger than the percentile of order 0.999 of the distribution. The quantity D provides a
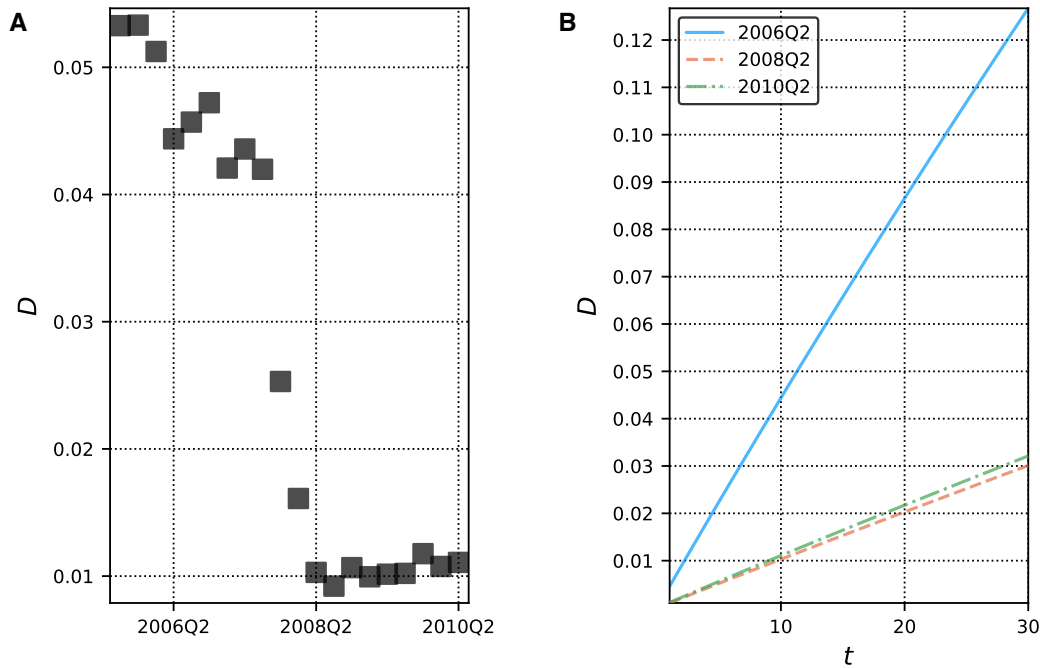
Figure 4: **Shock propagation before and after the global financial crisis.** Systemic fragility has fallen during the global financial crisis (**A**). Before the crisis, a single shock would have provoked a considerable loss for the whole system after few time steps (**B**). The velocity of damage propagation after the crisis is *1/4* than before but a systemic vulnerability still exists. Both panels correspond to the results of an initial shock $\delta_0 = 50\%$; panel **A** reports the damage after $t = 10$ steps.

measure of the *systemic damage* produced by a single shock, as a function of the time elapsed and of the shock entity.

Previously, we have shown that the average similarity between portfolios has decreased during crisis, see Fig. 2. We observe from Fig. 4 that similarity is correlated with fragility, since the network has indeed become more robust. The potential systemic damage in the network has been progressively reduced in the years 2005–2010 (panel **A**). Before the crisis, a 50% shock to just one of the top assets would give a 4–6% loss in the whole system's value after 10 time steps. To put this into perspective, such a loss would correspond to 120–180 billion dollars
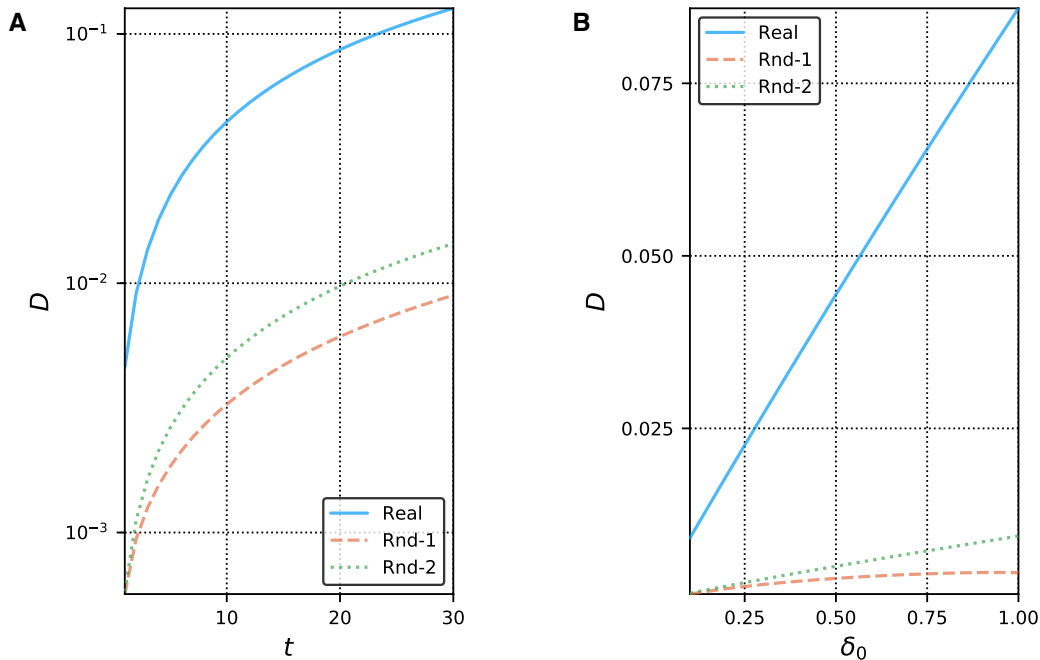
Figure 5: **Effect of similarity on systemic fragility.** Values of the systemic damage as a function of time for $\delta_0 = 0.5$ **(A)** and as a function of shock entity for $t = 10$ **(B)**. Randomization has a positive impact: damage propagation in the real case is 8.8 times faster than for model Rnd-2 and 14.5 time faster than model Rnd-1. After randomization, the propagation is also less sensitive to shock entity. Comparison of models Rnd-1 and Rnd-2 indicates that systemic fragility can be ascribed mainly to the similarity between portfolios and that systemic damage is minimum for completely random investments (Rnd-1).

in currency amount in 2006Q2. The loss reduces to 1% after crisis. As already observed for the average similarity, this trend has already started at the beginning of the observation period but has undergone a strong acceleration during the global financial crisis. Moreover, before the GFC distress propagation was a fast process (panel **B**) and damage could hit the 10% barrier in few steps. After the crisis the velocity of damage propagation, estimated by the slope of $D(t)$, is four times smaller. Nonetheless, a vulnerability still exists.

In the two panels of Fig. 5 we compare the systemic damage in a sample snapshot (2006Q2) of the real system with its random counterparts.

The comparison is performed for a fixed value of the shock ($\delta_0 = 0.5$) as a function of time, and for a fixed time ($t = 10$) as a function of the shock, respectively.

The result is that systemic fragility reduces by at least an order of magnitude when a randomization is performed. In model Rnd-1, where the diversification profile of the original portfolios is altered (we obtain a model of random investments), the values of h are concentrated, the average is smaller that in the original network, and large values of h are simply not observed (see Fig. 3). Furthermore, the overlap between portfolios is also largely suppressed. In model Rnd-2, the original values of h are preserved and only the similarity is reduced since assets are reassigned randomly. From that property follows that the reduction of systemic damage that we observe for model Rnd-2 can be considered a direct effect of the reduction in the similarities between portfolios. The damage for Rnd-1 is even smaller, the difference getting larger with both time and shock entity. Counter-intuitively, a model of random investments, where portfolios tend to have similar diversification and a small number of leading assets, is more robust with respect to the simulated dynamics of shock propagation.

We conclude that significant similarities between portfolios, a statistical signature of the holding network, is crucial in the transmission of financial distress and can make the network more fragile. Moreover, in a network of random investments (Rnd-1) this systemic risk component is reduced further and the network appears less fragile even though portfolios are more concentrated.

Discussion

We perform the first extensive study of the structure and the evolution of the US mutual fund network throughout the Global Financial Crisis (GCF) of 2007–2008. Even though in normal times households rarely rebalance their retirement saving portfolios [25], it has been found that 21% of them changed investment strategy between February and November 2009 [26]. Such a

dramatic recomposition of investment portfolios during crises can have severe consequences on financial stability. The size of the "ecosystem" of different fund investments has grown steeply over time [5] and, as an average, in the aftermath of the crisis mutual funds have become better diversified and less similar. However, mean values do not tell the whole story. Inspection of the probability density functions shows evidence of an heterogeneous system, with few largely diversified hubs and many specialized funds. Moreover, the probability of the similarity between portfolios decays slowly and large similarities are far more likely that can be expected from benchmark models of a random network of investments. We conclude that a high degree of correlation exists between investment decisions of different funds. This correlation limits the effectiveness of fund of funds diversification strategies.

One of the leading forces behind the emergence of such correlation can be found in the social network of relationships between fund managers [27] and the effects of managerial sharing. Other reasons may be due to herding behavior and the fact that professional investors with similar targets and risk profiles are likely to adopt similar investment strategies. For instance, portfolio managers try to maximize profits and the strategies of many of them will likely include those assets that have proved to be profitable or that can be selected by shared quantitative analysis techniques. Extreme market uncertainty can act as a driver for fund investments during a crisis, when an important fraction of their invested capital is moved from equity mutual funds to fixed-income mutual funds. During the crisis, defined contribution equity mutual funds experienced a large outflow of more that -15%, while flow into the fixed income mutual funds reached an historical peak of +20% [5]. Similarly, many funds might be damaged during crisis at the same time and trigger a second-order effect by which other funds get hit in a failure cascade. In our stylized representation of distress propagation, such second-order effect is induced by the many individual investors that simultaneously sell fund shares. However, a complete representation would also consider changes in the network topology as a result of the fact that portfolio

17

managers will try to rebalance their portfolios. Massive co-movements in fund allocations as a response to crisis may have an even higher impact on the market value of securities. This, in turn, may result in significant effects back to the mutual fund network and possibly lead to higher levels of overlap between funds.

Similarity of portfolios provides a different notion of diversification of investments. Exploiting a stylized model of shock propagation on a static network, we have shown how strongly this sort of systemic coupling can impact the fragility of the network. We find that the systemic damage in the network is much larger than the damage that can be procured to a network of random investments. Thus we quantify the systemic risk induced by the similarity of portfolio investment strategies. The comparison between a completely random model and a model where the diversification of original portfolios is preserved shows that similarity between portfolios represents a major source of systemic risk that can make the network fragile. It also emerges that completely random investments, even though corresponding to more concentrated portfolios, would result in a less vulnerable network. That is not to say that diversification is self-defeating *per se*, but indicates that it can make the system more fragile as funds tend to diversify their portfolios with similar investment patterns. The global financial crisis has stimulated an increase in the diversification of portfolios, but a systemic risk component still exists because of the similarity of investments.

We believe that the evidences presented in our study have implications for both the modeling and the regulation of financial networks. They show that similarity induces a systemic vulnerability and that portfolio diversification may even play antagonistically when strong correlations between investment strategies exist. Both aspects should be taken into account for the purpose of assessing systemic risk in holding networks and for devising effective policy actions.

Materials and Methods

Dataset   The US Mutual Fund market is the largest in the world. With 15\$ trillion assets under management at year-end 2013, it accounts for about half the total value in mutual fund assets worldwide. In this study, we analyse data from the CRSP Survivor-Bias-Free Mutual Fund Database. It includes open-ended mutual funds registered with the Securities and Exchange Commission and provides detailed information about the composition of managed portfolios. A mapping of funds to the portfolios of assets they manage is provided and detailed information about portfolio holdings, including the market value of each holding, is delivered on a quarterly basis.

In the following, we provide a formal construction of the bipartite network of holdings. We also present a schematic description of the model of distress propagation considered in the main text.

Bipartite graph of portfolio holdings   We represent portfolio holdings in terms of a bipartite graph. The two vertex classes are the set of mutual funds $i = 1, \ldots, N_f$, and the set of the different assets $\alpha = 1, \ldots, N_a$ in their portfolios. The degree k_i of vertex i is exactly the number of distinct assets held in the portfolio of fund i. Edges incident at vertex i can be assigned weights $W_{i\alpha}$ equal to the total market value of the shares of asset α and the graph is also conveniently represented by a $N_f \times N_a$ matrix B with elements $B_{i\alpha} = W_{i\alpha}$. We indicate as $G(N_f, N_a)$, or simply G, the undirected bipartite network of portfolio holdings. If we wanted to retain just the topological information of which asset is owned by which fund, we could define an unweighted graph $G^0(N_f, N_a)$. This would correspond [28, 29] to the incidence matrix B^0 whose elements $B^0_{i\alpha}$ are 1 if asset α is in the portfolio of fund i and 0 otherwise. A schematic picture of the network of holdings is provided in Fig. 6.
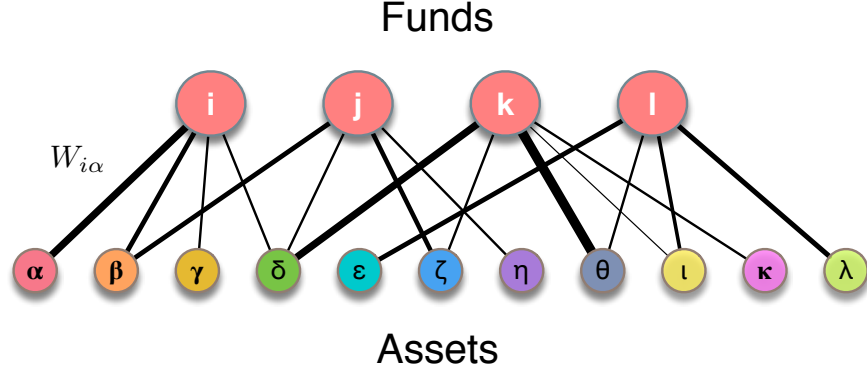
Funds

Figure 6: **Graph of portfolio holdings.** The network of portfolio holdings can be represented as a bipartite graph. The two vertex classes are the *funds* $\{i, j, k, \dots\}$ and the *assets* $\{\alpha, \beta, \gamma, \dots\}$ in their portfolios. Each edge (i, α) represents a specific holding relationship. The edge weight $W_{i\alpha}$ is equal to the total market value of security α owned by fund i in its portfolio.

The *size* of fund i is given by

$$S_i = \sum_{\alpha=1}^{N_a} W_{i\alpha} \,,$$

where it is understood that $w_{i\alpha} = 0$ if asset α is not in the portfolio. The quantities $W_{i\alpha}$ We easily see that, in network terms, fund size corresponds to *node strength* [30, 31]. We also indicate as M_α the aggregate market value of an asset

$$M_\alpha = \sum_{i=1}^{N_f} W_{i\alpha} \,.$$

The sum $S_{tot} = \sum_i S_i$ represents the *total value* of the system. By construction it holds $\sum_{i=1}^{N_f} S_i = \sum_{\alpha=1}^{N_\alpha} M_\alpha$.

The quantities S_i and $W_{i\alpha}$ are expressed in currency units. We also define the *portfolio weights* $w_{i\alpha} = W_{i\alpha}/S_i$, that represent the fraction of portfolio wealth corresponding to each asset. The indices of diversification and similarity discussed in the main text are all expressed in terms of the $w_{i\alpha}$s.

Due to portfolio reallocations, the set of assets in a portfolio and the edge weights can change with time; also the number of indexed funds in the dataset depends on time. Report-

ing of portfolio reallocations by funds are not synchronous and some choices are due in order to aggregate the information about holdings over some time horizon and construct the graph representing the system. Since portfolio composition is reported on a quarterly basis, in our analysis we choose a three-month time frame for aggregation. We create quarterly snapshots of the bipartite network by means of the following procedure. For each quarter, we consider the set of the funds $i = 1, \ldots, N_f$ for which a mapping to a portfolio exists. For the given quarter and for each fund i, we retrieve the holding information at the most recent report date t_i. Basically, this information is given by the set of assets in the portfolio and the corresponding market values. Fund i is inserted into the graph for the current quarter and so are the links (i, α), with $\alpha \in [1, N_a]$, for all assets held by the fund. Each link is assigned weight $W_{i\alpha}$ equal to the market value of the holding it represents.

When parsing the holding relationships in the dataset, attention has to be paid to the issue of *fund classes*. As a matter of fact, a fund may issue different types of shares all corresponding to the same underlying portfolio. In the database, different classes of a funds are associated to different unique identifiers, as if they were distinct funds. For the purpose of our analysis, we consolidated information about fund classes, to avoid including the same fund multiple times when constructing the network snapshots.

Basic model of distress propagation Let us suppose that at time t the prices of some securities undergo a negative shock $\delta_\alpha(t) = [v_\alpha(t+1) - v_\alpha(t)]/v_\alpha(t) \leq 0$, where $v_\alpha(t)$ is the market price of α. These downward moves produce a negative variation in the values of the portfolios that hold the securities in object. As a consequence, we expect individual fund investors to disinvest their shares, effectively selling the individual securities of the corresponding portfolio.

Generally, we expect that to a larger drop Δ_i of the portfolio of fund i, corresponds a higher probability $p(\Delta_i)$ for an investor to disinvest its shares of the fund. Let V_α be the total market

21

value of α disinvested in the process of selling fund shares. The instantaneous increase in the offer of α will have a negative impact in its value on the market. This reflects in a new drop $\delta_\alpha(t+1) = \lambda(V_\alpha(t))$, where λ is some *market impact* function (deterministic or stochastic). In the absence of large random effects and portfolio reallocations, in a connected network of portfolio holdings such a recursive dynamics will bring a progressive reduction of the total value of the system.

For the purpose of simplicity and clarity, the simulations discussed in the main text assume the fraction of fund shares liquidated to be equal to the portfolio drop and the market impact to be proportional to V_α. Such a model provides a simplified, schematic representation of shock propagation within a bipartite network of portfolios and their holdings. It should be considered an approximation of the dynamics in the time span that divides subsequent reallocations executed by fund managers, wherein the network can be regarded as static. It incorporates *contagion effects* due to common exposures of overlapping portfolios and is a multiplicative process where initial shocks get amplified over time.

To compare the systemic fragility of the network at different times, we performed simulations where a single security α is given an initial shock. This shock propagates for a number of steps and at each time the percentage total loss of value of the system is computed. So, the damage at time t given a shock to α is computed as $D_\alpha(t) = (S_{tot}(t+1) - S_{tot}(0))/S_{tot}(0)$. The process is repeated for a given set A of securities and we define the *systemic damage* by the average $D(t) = 1/|A| \sum_{\alpha \in A} D_\alpha(t)$. In order to reduce the computational time and speed up the process, we choose A as the set of assets for which the total value in the portfolios of funds is individually above the quantile of order 0.999 (*top assets*).

List of Supplementary Materials items

- Table S1: **Summary statistics of the bipartite network of holdings.**

References

[1] M. Statman, *Journal of Financial and Quantitative Analysis* **22**, 353 (1987).

[2] D. L. Domian, D. A. Louton, M. D. Racine, *Studies in Economics and Finance* **21**, 40 (2003).

[3] D. L. Domian, D. A. Louton, M. D. Racine, *The Financial Review* **42**, 557 (2007).

[4] M. Statman, *Financial Analysts Journal* **60**, 44 (2004).

[5] C. Sialm, L. Starks, H. Zhang, *American Economic Review* **105**, 432 (2015).

[6] J.-L. Hu, T.-P. Chang, R. Y. Chou, *Journal of Productivity Analysis* **41**, 141 (2014).

[7] S. Battiston, *et al.*, *Science* **351**, 818 (2016).

[8] S. Battiston, G. Caldarelli, R. M. May, T. Roukny, J. E. Stiglitz, *Proceedings of the National Academy of Sciences* **113**, 10031 (2016).

[9] D. Acemoglu, A. Ozdaglar, A. Tahbaz-Salehi, *American Economic Review* **105**, 564 (2015).

[10] S. Battiston, M. Puliga, R. Kaushik, P. Tasca, G. Caldarelli, *Scientific Reports* **2**, 541 (2012).

[11] S. Battiston, D. D. Gatti, M. Gallegati, B. Greenwald, J. E. Stiglitz, *Journal of Financial Stability* **8**, 138 (2012).

[12] M. Elliott, B. Golub, M. O. Jackson, *American Economic Review* **104**, 3115 (2014).

[13] M. Galbiati, D. Delpini, S. Battiston, *Nature Physics* **9**, 126 (2013).

[14] D. Delpini, *et al.*, *Scientific Reports* **3**, 1626 (2013).

[15] T. Feldman **75**, 192 (2010).

[16] P. S. Wool, Essays Concerning the Network Structure of Mutual Fund Holdings and the Behavior of Institutional Investors, eScholarship, University of California (2013).

[17] G. Abdesaken, Conflicts of interest in multi-fund management conflicts of interest in multi-fund management (2013).

[18] Y. Schwarzkopf, J. D. Farmer, *Phys. Rev. E* **81**, 066113 (2010).

[19] Y. Schwarzkopf, J. D. Farmer, What drives mutual fund asset concentration? (2016). Available at SSRN: https://ssrn.com/abstract=1173046.

[20] H. Markowitz, *The Journal of Finance* **7**, 77 (1952).

[21] S. Battiston, *Physica A: Statistical Mechanics and its Applications* **338**, 107 (2004).

[22] J. Glattfelder, S. Battiston, *Physical Review E* **80**, 036104 (2009).

[23] W. Woerheide, *Financial Services Review* **2**, 73 (1993).

[24] P. Jaccard, *Bulletin de la Société Vaudoise des Sciences Naturelles* **37**, 547 (1901).

[25] J. Agnew, P. Balduzzi, A. Sundén, *American Economic Review* **93**, 193 (2003).

[26] M. D. Hurd, S. Rohwedder, Effects of the financial crisis and great recession on american households, *Working Paper 16407*, National Bureau of Economic Research (2010).

[27] C. Augustiani, L. Casavecchia, J. Gray, *International Review of Finance* **15**, 427 (2015).

[28] G. Caldarelli, *Scale-free networks. Complex webs in nature and technology* (Oxford University Press, Oxford, 2007).

[29] M. E. J. Newman, *Networks: an introduction* (Oxford University Press, Oxford; New York, 2010).

[30] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, *Proceedings of the National Academy of Sciences of the United States of America* **101**, 3747 (2004).

[31] S.-H. Yook, H. Jeong, A.-L. Barabasi, *Proceedings of the National Academy of Sciences of the United States of America* **99**, 13382 (2002).

Supplementary Materials

Quarter	N_f	N_a	ρ	$\bar{k}$	S_{tot} ($B)
2005Q3	2474	7498	0.015	109	2123.7
2005Q4	2233	7117	0.016	113	2523.8
2006Q1	2545	7810	0.015	121	2767.7
2006Q2	3128	10743	0.011	114	2944.9
2006Q3	3197	10878	0.010	112	2980.4
2006Q4	3186	11471	0.010	117	3221.5
2007Q1	3155	12410	0.010	123	3042.2
2007Q2	3609	12482	0.010	121	3738.0
2007Q3	3492	16303	0.008	128	3848.2
2007Q4	3414	37694	0.004	144	4319.9
2008Q1	3773	74875	0.003	189	4983.3
2008Q2	5007	158894	0.001	198	5617.8
2008Q3	5411	197897	0.001	194	4626.6
2008Q4	5453	136511	0.001	189	3718.1
2009Q1	5689	164507	0.001	200	4653.8
2009Q2	5733	152444	0.001	203	5355.9
2009Q3	5689	147706	0.001	205	6066.5
2009Q4	5751	119412	0.002	182	6036.0
2010Q1	5511	131137	0.001	187	6300.3
2010Q2	5022	122458	0.002	202	5957.7

Table S1: **Summary statistics of the bipartite network of holdings.** Number of funds N_f and assets N_a, the network density ρ, the average number of assets in portfolios $\bar{k}$ and the total value of the network expressed in billion dollars. Network density is defined as $\rho = E/(N_f N_a)$ where E is the total number of investments (number of portfolio–asset holding relationships). The number of indexed funds and different assets in their investments have grown systematically. The average degree of funds nearly doubles during crisis and the network becomes less dense, with ρ decreasing by an order of magnitude.