

RESEARCH

Open Access



# Portfolio diversification, differentiation and the robustness of holdings networks

Danilo Delpini<sup>1\*</sup> , Stefano Battiston<sup>2†</sup>, Guido Caldarelli<sup>3,4,5†</sup> and Massimo Riccaboni<sup>5†</sup>

\*Correspondence: [ddelpini@uniss.it](mailto:ddelpini@uniss.it)

<sup>†</sup>Stefano Battiston, Guido Caldarelli and Massimo Riccaboni contributed equally to this work.

<sup>1</sup>Dept. of Economics and Business, University of Sassari, Via Muroni 25, 07100 Sassari, Italy

Full list of author information is available at the end of the article

## Abstract

Networks of portfolio holdings exemplify how interdependence both between the agents and their assets can be a source of systemic vulnerability. We study a real-world holdings network and compare it with various alternative scenarios from randomization and rebalancing of the original investments. Scenarios generation relies on algorithms that satisfy the global constraints imposed by the numbers of outstanding shares in the market. We consider fixed-diversification models and diversification-maximizing replicas too. We extensively analyze the interplay between portfolio diversification and differentiation, and how the outreach of exogenous shocks depends on these factors as well as on the type of shock and the size of the network with respect to the market. We find that real portfolios are poorly diversified but highly similar, that portfolio similarity correlates with systemic fragility and that rebalancing can come with an increased similarity depending on the initial network configuration. We show that a large diversification gain is achieved through rebalancing but, noteworthy, that makes the network vulnerable in front of unselective shocks. Also, while the network is riskier in the presence of targeted shocks, it is safer than its random counterparts when it is stressed by widespread price downturns.

**Keywords:** Systemic risk, Network models, Dynamics on networks

## Introduction

When we consider large institutional investors like mutual funds, interconnections arise because of managerial sharing (Augustiani et al. 2015), herding behavior (especially during crises), or simply similarity of investment strategies. Asset-overlap at the global scale exists because of the activity of global and international funds that invest in foreign assets around the world. The effects on systemic riskiness of such interdependence have been emphasized by the Global Financial Crisis 2007–2008. A very active area of research deals with the problem of making quantitative statements about the fragility of financial systems with respect to the propagation of distress. The latter can be negative downturns in asset values or insolvency of financial institutions depending on cases. Network science has provided insights into this topic and, in particular, the role of network topology characterizing mutual ties and exposures has been investigated (Hurd and Rohwedder 2010;

Battiston et al. 2012b; 2012a; Huang et al. 2013; Delpini et al. 2013; Galbiati et al. 2013; Caccioli et al. 2014; Acemoglu et al. 2015; Elliott et al. 2014).

A cornerstone of portfolio management is diversification (Statman 1987; Agnew et al. 2003; Domian et al. 2007; Hu et al. 2014), a well-understood strategy to cut down portfolio risk from idiosyncratic shocks to asset prices. A reduction of individual risk can be achieved by reducing portfolio concentration (Domian et al. 2003; 2007; Statman 2004) but the evidences collected from the crisis suggest that the efficacy of diversification strategies can depend strongly on market conditions. More generally, the systemic implications of diversification and its relationship with the notion of *differentiation* between the holdings of different market players are uncertain. In Delpini et al. (2019) the network of mutual funds portfolios in the United States has been studied across the crisis. It has been found that diversification has increased and, less predictably, investment similarities have decreased. Simulations show that the observed similarities between portfolios are more likely than one can expect by chance and finite-size effects, even when controlling for strongly connected assets, and finally there exist groups of highly similar portfolios. This makes the network riskier with respect to random counterparts to parity of portfolio diversification. Accordingly, the analysis in Fricke (2019) suggests that, while diversification may reduce risk for an individual portfolio, the structure of the similarities across mutual funds could be key in determining systemic risk.

In this paper we study bipartite networks of portfolio holdings (Delpini et al. 2019; Lin and Guo 2019; Braverman and Minca 2018; Guo et al. 2016). We model a holdings network as sub-system of the embedding financial market wherein the numbers of outstanding shares of each stock are treated as constants and provide global market constraints. We aim at analyzing how the outreach of financial shocks in the holdings network depends on network topology, portfolio diversification and differentiation, as well as on the type of shock applied and the relative size of the network. In particular, we want to test if more diversified portfolios can actually reduce risk at the systemic level as one might expect. In order to generate random scenarios for comparisons, we take advantage of algorithms that are required to satisfy the global constraints at any time. Such algorithms mimic the actual process of increasing, reducing or replacing asset positions through buy/sell orders to the market. In particular, we consider explicitly scenarios where holders reallocate their wealth to either the original or randomly chosen assets in a way to achieve the most diversified portfolio compatible with the constraints in force. We develop the analysis in a computational setting, exploiting a snapshot of the US mutual funds holdings network as a test case. We extensively compare the original network with the simulated scenarios and try to better characterize the interplay between portfolio diversification and differentiation. We then simulate the process of propagation of exogenous shocks to asset prices, based on a dominating flow-induced trading dynamics wherein negative fund performances are followed by asset fire-sales (Greenwood et al. 2015; Fricke 2019). We do that for both targeted and unselective shocks and compare the systemic damages registered for each scenario.

In “[Model of bipartite network of holdings](#)” section we formalize the notion of a holdings network within the market and its representation as a bipartite graph. In “[Market model of flow-induced trading](#)” section we describe the dynamics of shock propagation that will be used in the simulations. “[Algorithms](#)” section describes the basic concepts underlying the algorithms for generating rebalanced portfolios or performing

holdings reshuffle. The statistical properties of the generated scenarios, their comparison in terms of diversification, differentiation and riskiness in front of different kinds of shocks are presented in “Results” section. We draw conclusions afterwards.

## Materials and methods

### Dataset

As a case study for our analysis, we employ a snapshot of the bipartite network of US mutual fund holdings corresponding to the third quarter of 2012. The network was inferred from data in the Survivor-Bias-Free US Mutual Funds database provided by the Center for Research in Security Prices, The University of Chicago Booth School of Business. We only considered equity funds with a reported Total Net Assets (TNA) greater than or equal to one million US dollars (USD) and, in addition, we discarded funds that are classified as International and Global. We parsed the data for just the holdings that correspond to stocks with a valid ticker. Assets with coupon or maturity information and derivatives were not considered. This filtered network has 3497 different funds investing in 9015 different assets. The number of holdings is 550,554 for a total asset value of 2.61 trillion USD. In the following, we will use the generic term “holders” to refer, in particular, to US mutual funds and their asset managers, and “holdings” to indicate the different stocks in the funds’ investment portfolios.

### Model of bipartite network of holdings

We model a network of holdings with  $m$  portfolios and an investment universe of  $n$  different stocks as a bipartite graph (Caldarelli 2007; Newman 2010). We first introduce the network’s incidence matrix  $\mathbf{B} = (b_{ij})$ , where  $b_{ij} = 1$  if portfolio  $i$  invests in stock  $j$  and  $b_{ij} = 0$  otherwise. The number of holdings is the number of the non-zero elements of  $B$  and we will indicate it as  $n_h = |E|$ , where  $E = \{(i, j)\}$  stands for the set of the holdings and  $|E|$  for its cardinality. Let  $n_{ij}$  be the number of shares of stock  $j$  in the portfolio and  $s_j$  the stock price. The *share matrix*  $\mathbf{N} = (n_{ij})$  can be assumed as the weighted incidence matrix.

The TNA of portfolio  $i$  is  $p_i = \sum_{j=1}^n n_{ij} s_j = (\mathbf{N} \mathbf{s})_i$  where  $\mathbf{s} = (s_1, s_2, \dots, s_n)$  is the vector of stock prices. By introducing the *value matrix*  $\mathbf{V} = (v_{ij})$  whose elements  $v_{ij} = n_{ij} s_j$  are the holding values, the TNA can be written equivalently as  $p_i = (\mathbf{V} \mathbf{1}_n)_i$  where  $\mathbf{1}_n$  is the vector with all  $n$  components being one. In the following we will also refer to the *portfolio weight*  $w_{ij} = v_{ij}/p_i$  of each position, representing the relative weight of an individual position into a stock with respect to the portfolio’s TNA.

Such network is a subsystem of the whole “market”. The number  $N_j$  of outstanding shares of each stock  $j = 1, \dots, n$  is assumed to be constant in the market. We do not take into account share splitting or new issues. This implies that holders’ portfolios can not hold more shares at any time  $t$  than those outstanding:

$$\sum_i n_{ij}(t) \leq N_j, \quad j = 1, \dots, m. \quad (1)$$

The above inequalities represent global constraints. These need to be satisfied during all stages of the execution of the randomization algorithms and at any time during the propagation of exogenous shocks through the portfolio network.

The exact values of  $N_j$  can not be retrieved easily. However, if we think of the holdings network as a small-scale representation of the market, we can introduce a scaling

parameter  $c$  and imagine that the total value of each stock in holders' portfolios initially is a fraction  $1/c$  of its value in the whole market. In this spirit, and for the purposes of simulations, we make the following choice for the outstanding shares:

$$N_j = c \sum_i n_{ij}(0) . \quad (2)$$

Parameter  $c$  captures the size of the network, in total assets terms, relative to the market: the larger  $c$  the more the holdings network is to be considered small with respect to its embedding market. Considering the total assets value of our test network (see “[Dataset](#)” section), and that the New York Stock Exchange had a market capitalization of 30.1 trillion USD as of February 2018,  $c = 10$  can be considered a sensible choice. We will also compare the results for this value to those for  $c = 2$  and  $c = 100$ , to also account for scenarios where the holdings network represents a large or small fraction of the market respectively.

### Market model of flow-induced trading

We are interested in studying how the aggregate fragility of the network is dependent on the overlap of portfolios and in testing interaction between portfolio diversification and overlap. We consider a case where market-wide negative events can determine mass investor outflows from funds. In such events, a fund will be forced to liquidate part of its asset positions in order to repay leaving investors. This trading activity will have a (negative) impact on stock prices and this will cause funds that were not hit by the original event to experience losses afterwards.

Usually funds do not retain cash: a flow of investors into a fund is followed by an expansion of asset positions while an outflow is associated to a shrinkage of positions. Outflow corresponds to fund investors asking for redemption of their shares and to fund managers liquidating some assets to meet requests. Such trading activity by funds is referred to as *flow-induced trading*. We are interested in studying the network's reaction to idiosyncratic shocks on a small time scale, when portfolio managers fire-sale assets with the effect of amplifying the shocks and triggering new ones. In particular we aim at measuring how fast the total value of the network is eroded and to what extent network fragility depends on asset commonality between portfolios. Accordingly, we consider a model where portfolio managers can only liquidate assets and we make the working assumption that the market outside the network is large and liquid enough to absorb such offer.

Consider the dynamics of a portfolio along a generic trading period. Let  $p(t - 1)$  be the end-of-period TNA for period  $t - 1$  and assume that this value includes the fund's flow along the period. If the portfolio's composition was not altered, its return  $r(t)$  along period  $t$  would depend only on market price variations and the amount  $p(t - 1)$  would grow at rate  $r(t)$  into the amount  $[1 + r(t)] p(t)$ . We define *fund flow* along period  $t$  the difference between the closing TNA and the accrued value:

$$f(t) = p(t) - [1 + r(t)] p(t - 1) .$$

Because flow-induced trading is needed to repay leaving investors, we assume that it comes with no modifications to the portfolios' investment strategies, by which we mean that the stocks in the portfolios will remain the same. This implies that portfolio managers liquidate every asset in the same proportion (*proportional selling*). Let  $\eta_i(t)$  be the fraction of TNA that is liquidated along period  $t$

$$\sum_j n_{ij}(t) s_j(t - 1) = p_i(t - 1) [1 + \eta_i(t)] .$$

Then proportional selling corresponds to solution  $n_{ij}(t) = [1 + \eta_i(t)] n_{ij}(t - 1)$  to the previous equation, with  $\eta_i(t) < 0$ .

Flow-induced trading has an impact on the prices of the securities being traded. The *price impact*  $I_j(x)$  of trading  $x$  shares of a stock is defined as the corresponding relative price variation. In the literature it is often assumed that a linear relationship exists (Kyle 1985) of the form  $I_j(x) = x/\lambda_j$ . The stock-specific parameter  $\lambda_j$  is called the stock's *market depth*: the larger  $\lambda_j$  the more liquid the stock and the smaller the price impact of trading it. We assume that the market depth of a stock can be approximated by the total number of shares that have been issued and purchased by all participants in the market (*outstanding shares*). We take  $\lambda_j = N_j$  and such parameter accounts for heterogeneity in the liquidity characteristics of stocks. The price impact on stock  $j$  along period  $t$  will be

$$I_j(t) = \frac{\Delta s_j(t)}{s_j(t - 1)} = \frac{\sum_i \Delta n_{ij}(t)}{\lambda_j} = \frac{\sum_i \eta_i(t) n_{ij}(t - 1)}{N_j} < 0 . \tag{3}$$

Note that  $\lambda_j = N_j$  and the scaling relationship (2) imply that the larger is  $c$  the smaller is the price impact of asset liquidations.

Taking into account both the price variations and asset liquidations, the portfolio TNA at end-of-period will be

$$p_i(t) = [1 + \eta_i(t)] \sum_j [1 + I_j(t)] n_{ij}(t - 1) s_j(t - 1) .$$

We can motivate a simple choice for the amount of assets liquidated. Consider an individual investor soon after a negative return  $r_i(t - 1)$  has occurred to her fund's portfolio. The larger is  $r_i$ , the higher is the chance that the investor will ask for redemption of one of her fund shares. Because  $0 \leq |r_i| \leq 1$  we can interpret the absolute value of the portfolio (negative) return as a proxy of such probability. If we consider homogeneous investors and a uniform redemption probability across the fund's shares, the average fraction of redemptions will be  $|r_i(t - 1)|$ . Thus, as a basic approximation and discarding fluctuations, we also take  $\eta_i(t) = r_i(t - 1)$ .

To summarize, we consider a dynamics of shock propagation that proceeds as follows:

- 1 A negative shock to stock prices  $\delta(t) = (\delta_1(t), \dots, \delta_n(t))$  hits the market, with  $\delta_j = \Delta s_j/s_j \in [-1, 0]$ ;
- 2 This determines negative portfolio returns  $r_i(t)$  for those portfolios that invest in the stocks hit;
- 3 Negative returns determines an outflow from funds and a corresponding reallocation of portfolio positions  $n_{ij}(t) \rightarrow n_{ij}(t + 1) = [1 + \eta_i(t + 1)] n_{ij}(t)$  where  $\eta_i(t + 1) = r_i(t)$ .
- 4 Flow-induced asset selling has a negative price impact  $\delta(t + 1) = \mathbf{I}(t + 1)$  as given by Eq. (3) and then the process starts again from step 1:

$$\delta(t) \longrightarrow \mathbf{r}(t) \longrightarrow \Delta n_{ij}(t + 1) \longrightarrow \mathbf{I}(t + 1) = \delta(t + 1) .$$

For our purposes we assume that the first shock  $\delta(1)$  is exogenous: no outflow occurs during the first trading period and the first variation of the TNAs is completely determined by the negative returns.

**Systemic damage**

The dynamics described above implies that the value of every portfolio reduces by the end of the trading period, or it stays the same if the portfolio is not investing in the stocks that went through a price downturn. In the same spirit of Delpini et al. (2019), we consider the relative reduction in the total value of the portfolios  $D(t) = |\sum_i p_i(t) / \sum_i p_i(t - 1) - 1|$  as an indicator of network fragility. Comparison of the real network and its counterparts from random allocation scenarios will provide indication of what configuration is more robust in the light of global market constraints as well as insight into diversification–riskiness and differentiation–riskiness correlations.

A comprehensive measure of the fragility of the holding network would require to measure  $D$  with respect to all possible combinations of one or multiple shocks  $\delta_j(1)$ . This is impractical due to computational time constraints and sensible choices are needed to select appropriate values for  $\delta(1)$ . A shock to a single random stock, albeit large, is unlikely to produce large-scale effects in few trading periods. On the other hand a combination of very large shocks to many assets will produce a too fast and unrealistic degradation of portfolio values. A possible strategy is to consider worst case scenarios where a subset of targets is selected according to some criterion of asset network “centrality”. This was done in Delpini et al. (2019) where stress tests are performed applying a uniform shock to the most popular assets. There, the degree centrality  $k$  of a stock in the bipartite network was chosen to select the targets. Let us indicate this choice as the “ $k$ -targets” scheme. We will compare it with the following alternative choices.

Indeed it may not be obvious which measure of centrality is most appropriate. In principle, there could be assets that are very popular but account for a small fraction of the network’s value or, on the contrary, assets that are owned by few large holders with large portfolio weights. In order to take into account the actual monetary weight that stocks have in portfolios as well their popularity, we propose to alternatively select targets by maximizing the following indicator:

$$h_j = \left[ \sum_{i=1}^m \left( \frac{v_{ij}}{s_j N_j} \right)^2 \right]^{-1} = \left[ \sum_{i=1}^m \left( \frac{n_{ij}}{N_j} \right)^2 \right]^{-1} .$$

If holders owned all outstanding shares in the market,  $h_j$  would account for the number of funds that own the largest portions of the stock’s market value, or simply the number of leading holders of stock  $j$ . Taking into account relationship (2), such a number actually gets multiplied by a uniform factor  $c^2$ . Nevertheless we continue to think of it as a number of effective holders and, following this interpretation, we will say that the stocks with the highest  $h_j$  are the “most owned” in the network. In this sense,  $h_j$  is a rescaled version of the Herfindahl–Hirschman index, computed for stock  $j$  in terms of the ratios of the numbers of shares in each portfolio to the total number of outstanding shares. Let us indicate a scenario where target stocks are selected maximizing  $h_j$  as the “ $h$ -targets” scheme. In both this scheme and the “ $k$ -targets” scheme we apply a uniform random shock  $\delta_j \in [\delta_{\min}, \delta_{\max}]$  to the  $n_{\text{target}}$  stocks with the largest values of  $k_j$  and  $h_j$  respectively. We take  $n_{\text{target}} = 0.1 \times n$  and arbitrarily set  $\delta_{\min} = 1\%$  in order to produce an appreciable systemic damage.

Crises can come with negative returns that are widespread across a lot of stocks. In order to account for that, we also consider an “all-targets” scheme where a uniform shock  $\delta_j \in [0, \delta_{\max}]$  is applied to all stocks. This choice also free us from the need of arbitrarily

selecting a subset of targets. In this case we allow stock prices to eventually be unchanged, that is  $\delta_{\min} = 0\%$ .

Trading of a given stock on a stock exchange is suspended if the price relative loss exceeds a stopping threshold. This is dependent of the exchange regulation and usually less than 10%. We therefore set  $\delta_{\max} = 10\%$  for all cases.

For every scheme and different underlying holding network, we perform  $n_{\text{mc}} = 100$  Monte Carlo runs of shock propagation over  $T = 10$  trading periods. Systemic damage is computed as the average value of  $D$  over the runs.

**Portfolio diversification and similarity**

We measure diversification by the portfolio’s *Herfindahl–Hirschman index*

$$h_i = \left[ \sum_{j=1}^n \left( \frac{v_{ij}}{p_i} \right)^2 \right]^{-1} = \left[ \sum_{j=1}^n w_{ij}^2 \right]^{-1},$$

which is interpreted as the number leading holdings in a portfolio.

Similarity between two portfolios can be quantified by their *cosine similarity*:

$$s_{ii'} = \frac{\sum_j v_{ij} v_{i'j}}{\|\vec{v}_i\| \|\vec{v}_{i'}\|} = \frac{\sum_j w_{ij} w_{i'j}}{\|\vec{w}_i\| \|\vec{w}_{i'}\|},$$

where  $\vec{v}_i = (v_{i1}, v_{i2}, \dots)$  is the vector of holdings of portfolio  $i$  and  $\vec{w}_i = (w_{i1}, w_{i2}, \dots)$  is the corresponding vector of portfolio weights. It is worth noticing that both indicators  $h_i$  and  $s_{ii'}$  are independent of the portfolio’s TNA and can be equivalently expressed in terms of either the holding values or the portfolio weights. Throughout the paper, we refer to the subsidiary notion of *differentiation* between two portfolios. In symbols, we define it as the complementary quantity of similarity, or  $1 - s_{ii'}$ , and both quantities take values in  $[0, 1]$ . The less two portfolios are similar, the more they are different.

Consider two portfolios of degree  $k_i$  and  $k_{i'}$  that invest in a perfectly balanced way, with uniform positions  $v_{ij} = p_i/k_i$  and  $v_{i'j} = p_{i'}/k_{i'}$  if stock  $j$  is in the portfolio, and 0 otherwise. Their similarity will be  $s_{ii'} = k_{ii'}/\sqrt{k_i k_{i'}}$  where  $k_{ii'}$  is the number of common assets. For a given degree sequence  $k_1, k_2, \dots, k_m$  the average similarity across the network is

$$\bar{s} = \frac{2}{m(m-1)} \sum_{i=1}^m \sum_{i'=i+1}^m s_{ii'} = \frac{2}{m(m-1)} \sum_{i=1}^m \sum_{i'=i+1}^m \frac{k_{ii'}}{\sqrt{k_i k_{i'}}}.$$

Consider also a network where every portfolio invests selecting uniformly at random the same number of stocks it owns in the real network. Since stocks are treated indifferently, it is natural to expect that such idealized investor would invest the same amount of wealth in every stock. We will refer to such limit scenario as an “unconstrained random holdings” network (URH). Because of the balanced positions, in this network investors maximize diversification conditionally on  $k_i$ , achieving  $h_i = 1/\sum_j (v_{ij}/p_i)^2 = 1/\sum_j (1/k_i)^2 = k_i$ , where the second summation is limited to the stocks  $j'$  that are actually present in the portfolio.

If we consider many realizations of a network like that, the expected value of the number of common assets will be  $\langle k_{ii'} \rangle \approx \frac{k_i k_{i'}}{n}$ , where the approximation holds as long as the probability of choosing the same stock multiple times can be neglected. Under this approximation, the expected value of the average network similarity reads:

$$s_0 = \langle \bar{s} \rangle \approx \frac{2}{m(m-1)} \sum_{i=1}^m \sqrt{k_i} \left( \sum_{i'=i+1}^m \sqrt{k_{i'}} \right), \tag{4}$$

and this can be considered a benchmark value with respect to which we evaluate deviations from randomness. As for the diversification, since in the URH case the degree sequence is constrained and each portfolio achieves the maximum diversification  $h_i = k_i$ , we see that the average network diversification is identically equal to the average degree  $h_0 = \bar{h} = \sum_i h_i/m$ . In the following it will be useful to compare different scenarios in terms of the network’s average diversification and similarity relative to the benchmarks  $h_0$  and  $s_0$ .

**Portfolio rebalancing**

In a perfectly balanced portfolio every stock has the weight  $w = p/k$ . Conditionally on  $k$ , such a portfolio has maximum diversification  $h = k$ . If a holder, starting with an unbalanced portfolio, wants to increase her diversification without opening positions in new stocks, or conditionally on  $k$ , she can try to rebalance her positions in the original assets steering them toward the optimal level  $w = p/k$ . At first, some positions are larger than  $w$ , some are smaller and possibly some are just at the threshold. The portfolio’s TNA can be decomposed in the following way

$$\begin{aligned} p &= \sum_j v_j = \sum_{j \in J^+} v_j + \sum_{k \in J^-} v_k + \sum_{l \in J^0} v_l \\ &= \sum_{j \in J^+} (w + \Delta v_j) + \sum_{k \in J^-} (w - \Delta v_k) + \sum_{l \in J^0} w \\ &= kw + \sum_{j \in J^+} \Delta v_j - \sum_{k \in J^-} \Delta v_k \\ &= p + \sum_{j \in J^+} \Delta v_j - \sum_{k \in J^-} \Delta v_k, \end{aligned}$$

where  $J^+$ ,  $J^-$  and  $J^0$  are the sets of stocks corresponding to redundant, deficient or balanced positions, while  $\Delta v_j$  is the difference between the actual position on stock  $j$  and the balanced position  $w$ . From the previous relationship it follows that

$$\sum_{j \in J^+} \Delta v_j = \sum_{k \in J^-} \Delta v_k,$$

which simply states that the total excess equals the total defect. A holder can liquidate redundant positions and use the liquidity to buy more shares and balance defective positions. In the absence of limits on the number of shares that can be bought or sold, the holder will achieve perfect diversification. Otherwise, she will end with a suboptimal portfolio but the value of  $h$  will have increased anyway.

**Algorithms**

Portfolios can be similar with respect to how many stocks they invest in, what stocks and how much money is invested in each one. We introduce two synthetic models where the original holdings are randomized with different strategies. We will refer to them by the symbols  $H_1$  and  $H_2$  respectively, distinguishing them from the original holdings network  $H_{or}$ . In both of them we assume portfolio TNAs and portfolio degree sequences as given.



In other words, we focus on how stocks are chosen, conditionally on the number of stocks  $k_i$  and the portfolio wealth  $p_i$ .

In model  $H_1$ , each holder goes through a process of replacement of its original stocks with new ones chosen uniformly at random. We refer to this as a “shuffling of holdings”. The degree sequence of stocks is not preserved and the degree of each stock is approximately a binomial random variable with mean  $|E|/n$ . As discussed in Delpini et al. (2019), the degree distribution of the real assets decays slowly and there exist very popular assets owned by thousands of portfolios. In order to account for the role of these hubs, we also consider model  $H_2$ . It differs from  $H_1$  in that stock replacement is performed by means of a double-edge swap strategy that preserves the degree sequence of stocks as well as that of holders. The number of swaps to be performed is fixed upfront and expressed as a fraction  $f$  of the number of holdings, that is  $n_{\text{swap}} = f \times n_h$ .

In both models the ideal investor uses no information during stock selection and treats stocks as being equivalent, if it was not for the number of available shares that is heterogeneous and subject to changes during the allocation process. For both scenarios  $H_1$  and  $H_2$  we consider two variants. In the first one holders tries to re-allocate exactly the same amount of money of each original position. This case will allow us to compare scenarios for equal portfolio concentrations and will be referred to as an “unbalanced scenario”. In the second one funds try to achieve the highest degree of diversification possible by going through the rebalancing process discussed in the previous section. In a way, this is also coherent with the idea of a random allocation: since stocks are treated equivalently, an investors will have no reason to make unbalanced positions, provided that enough shares are available from the market to buy. Rebalancing is performed after randomization: first a random network is generated where portfolio positions are exactly the same as the original ones, be it  $H_1$  or  $H_2$ , and positions are rebalanced afterwards<sup>1</sup>. We will consider a rebalanced network for the original topology as well and we will indicate the three balanced cases as  $H_{\text{or,b}}$ ,  $H_{1,\text{b}}$  and  $H_{2,\text{b}}$  respectively to distinguish them from their original or unbalanced counterparts.

The algorithms for rebalancing and randomization that we use are required to enforce the global constraints (1) at each stage of their execution. They ensure that when holders open new positions no more shares can be bought of a stock than actually available on the market at that very moment. This aspect represents a major contribution of our work. Indeed, the simple randomization algorithms in Delpini et al. (2019) preserved portfolio TNAs and degrees, but they were otherwise unconstrained. Because no constraints were imposed on the numbers of shares, the total value of every asset in the network of portfolios was allowed to change. Those routines were instrumental in comparing the actual network topology with its random counterparts but the corresponding network models are to be considered approximations valid in the limit of infinite numbers of shares. On the contrary, the  $n$  global constraints (1) are required to be satisfied at any time here. In particular, they limit the maximum diversification holders can reach without altering their degree and this can be critical for systemic risk assessment.

<sup>1</sup>The algorithms may be adapted to perform random allocation and rebalancing simultaneously. This increases algorithmic complexity and execution times and more importantly makes it difficult to guarantee a diversification increase upfront.

Both in scenario  $H_1$  and  $H_2$ , one holding at a time is replaced. When position  $v_{ij}$  is to be re-allocated,  $v_{ij}/s_j$  shares are sold of stock  $j$  and  $v_{ij'}/s_{j'}$  shares need to be bought of stock  $j'$  from the market. Since the number of outstanding shares is limited, the random re-allocation of an individual portfolio can fail. It can happen that the shares of  $j$  that are available on the market to buy at a given time are not sufficient to re-allocate the desired position. It can also be the case, in scenario  $H_2$ , that the stock that was sampled as a replacement can not be swapped with the original one without either violating the global constraints or changing the stocks' degree sequence. In such cases, a failure is registered and a new stock is randomly sampled for replacement. When a limit number of consecutive failures is reached during the allocation of a given portfolios, the whole randomization process is reset and restarted from the very beginning after a reshuffle of all stocks and holders. This is done to stochastically escape configurations that would not allow full re-allocation of all portfolios and to find possible solutions that may be reached through a different sequence of holding replacements.

## Results

We first provide a comprehensive summary of the statistical features of the original network and the balanced and/or random scenarios, in terms of portfolio diversification and similarity. We also compare them with the unconstrained random benchmark URH introduced previously and provide a nice model representation in a two-dimensional parameter space. Then we show the results of shock propagation across the different topologies and draw conclusions about the relationship between systemic risk, the selective or unselective nature of the shocks, the degree of diversification and similarity of the portfolios and also the relative size of the network to the market.

All routines used to generate random scenarios suffer from round-off errors to some extent. This means that even though the algorithms are guaranteed to preserve portfolio TNAs and total network value exactly, numerically they do not. In the case of unbalanced scenarios, the  $h$  index of portfolios is not preserved exactly either. We checked numerical accuracy by looking at the following quantities. As for the total network value and the TNAs, we considered the absolute differences  $M_{\text{err}} = |\sum_{ij} v_{ij} - M_{\text{or}}|$  and  $p_{\text{err}} = \max_{i=1, \dots, m} |\sum_j n_{ij} s_j - p_{i, \text{or}}|$ . For the diversification index, we computed the relative error  $h_{\text{err}} = \max_{i=1, \dots, m} |h_i/h_{i, \text{or}} - 1|$  instead. These errors turn out to be negligible with respect to the order of magnitude of the corresponding quantities involved in the analysis, see Table 1. We also include in the table the average relative increase of diversification and the fraction of edges that are different from the original network. These quantities are defined as  $h_{\text{incr}} = \sum_i [(h_i - h_{i, \text{or}})/h_{i, \text{or}}] / m$  and  $E_{\text{diff}} = |E \cap E_{\text{or}}| / |E_{\text{or}}|$  respectively. Rebalancing guarantees a large increase of diversification in all cases. On average, it allows portfolios to more than double their initial value of  $h$ . The larger  $c$ , the larger the diversification gain, as expected because a large  $c$  increases the probability for a holder to attain perfect rebalancing of her positions. As for model  $H_2$ , changing  $f$  from 1 to 10 does not provide a significant diversification gain. The value of  $E_{\text{diff}}$  measures how much the topology of the network has been altered by the randomization procedure. We register a turnover of holdings between 68% and 89%. For given  $c$ , model  $H_2$  has a slightly lower turnover because the random selection of new holdings is limited by the requirement of preserving the stocks' degree sequence. Increasing  $f$  to 10 allows to reduce the turnover

**Table 1** Numerical errors affecting the relevant quantities in the balanced and random scenarios

Network	c	f	$h_{incr}$	$E_{diff}$	$M_{err}$ (USD)	$P_{err}$ (USD)	$h_{err}$
$H_{or,b}$	2		1.13	0	0	$4.6 \cdot 10^{-4}$	
$H_{or,b}$	10		1.34	0	$4.9 \cdot 10^{-4}$	$6.8 \cdot 10^{-4}$	
$H_{or,b}$	100		1.36	0	$4.9 \cdot 10^{-4}$	$6.4 \cdot 10^{-4}$	
$H_1$	2			0.83	$4.9 \cdot 10^{-4}$	$6.1 \cdot 10^{-5}$	$7.3 \cdot 10^{-15}$
$H_1$	10			0.86	0	$1.5 \cdot 10^{-4}$	$1.1 \cdot 10^{-14}$
$H_1$	100			0.89	0	$1.5 \cdot 10^{-4}$	$7.5 \cdot 10^{-15}$
$H_{1,b}$	2		0.72	0.83	0	$2.1 \cdot 10^{-4}$	
$H_{1,b}$	10		1.05	0.86	0	$4.3 \cdot 10^{-4}$	
$H_{1,b}$	100		1.22	0.89	$4.9 \cdot 10^{-4}$	$6.3 \cdot 10^{-4}$	
$H_2$	2	1		0.68	0	$1.1 \cdot 10^{-4}$	$7.9 \cdot 10^{-15}$
$H_2$	2	10		0.74	0	$1.2 \cdot 10^{-4}$	$7.2 \cdot 10^{-15}$
$H_{2,b}$	2	1	1.15	0.68	0	$5.0 \cdot 10^{-4}$	
$H_{2,b}$	2	10	1.18	0.74	$4.9 \cdot 10^{-4}$	$3.3 \cdot 10^{-4}$	
$H_2$	10	1		0.69	0	$7.6 \cdot 10^{-5}$	$7.5 \cdot 10^{-15}$
$H_2$	10	10		0.74	0	$4.6 \cdot 10^{-5}$	$9.4 \cdot 10^{-15}$
$H_{2,b}$	10	1	1.33	0.69	0	$7.9 \cdot 10^{-4}$	
$H_{2,b}$	10	10	1.34	0.74	0	$6.6 \cdot 10^{-4}$	
$H_2$	100	1		0.69	0	$1.5 \cdot 10^{-4}$	$8.0 \cdot 10^{-15}$
$H_2$	100	10		0.74	$4.9 \cdot 10^{-4}$	$1.2 \cdot 10^{-4}$	$1.1 \cdot 10^{-14}$
$H_{2,b}$	100	1	1.36	0.69	0	$1.0 \cdot 10^{-3}$	
$H_{2,b}$	100	10	1.36	0.74	0	$6.2 \cdot 10^{-4}$	

difference with respect to model  $H_1$  and will be assumed as the default in the following discussion of the results.

In Table 2 we provide a 5-number summary of the statistical distribution of  $h$ , plus its interquartile range and mean value  $\bar{h}$ . The first, second and third quartiles, and the 9th and 91st percentiles are reported.

The distribution broadens after rebalancing and the mean more than doubles. Most evidently the distribution’s right tail grows fatter and the upper percentile  $q$  (0.91) can even increase threefold depending on the randomization model and parameters. In Table 3 the same summary is provided for the distribution of the portfolio similarity. We omit the lower percentile and the first quantile because they are zero for all cases.

**Table 2** Summary statistics of the portfolio diversification in the real network and its counterparts in the balanced allocation scenarios

Network	c	f	$q_h(0.09)$	$Q_{h,1}$	$Q_{h,2}$	$Q_{h,3}$	$q_h(0.91)$	$IQR_h$	$\bar{h}$
$H_{or}$			5.9	16.4	37.3	68.1	134.3	51.7	64.0
$H_{or,b}$	2		9.0	27.7	58.0	112.5	405.9	84.9	138.5
$H_{or,b}$	10		10.0	29.0	60.0	118.0	442.4	89.0	155.6
$H_{or,b}$	100		10.0	29.0	60.0	118.0	442.4	89.0	157.2
$H_{1,b}$	2		8.0	24.8	51.7	99.9	310.8	75.1	111.9
$H_{1,b}$	10		9.0	27.5	56.6	109.7	377.0	82.2	134.2
$H_{1,b}$	100		9.5	28.2	58.8	114.5	412.8	86.3	147.1
$H_{2,b}$	2	1	9.1	27.9	58.0	113.5	406.1	85.6	141.1
$H_{2,b}$	2	10	9.1	28.0	58.3	114.1	411.4	86.1	144.1
$H_{2,b}$	10	1	10.0	29.0	60.0	117.8	438.0	88.8	154.7
$H_{2,b}$	10	10	10.0	29.0	60.0	117.7	438.5	88.7	155.5
$H_{2,b}$	100	1	10.0	29.0	60.0	118.0	442.2	89.0	157.2
$H_{2,b}$	100	10	10.0	29.0	60.0	118.0	442.4	89.0	157.3

**Table 3** Summary statistics of the portfolio similarity in the real network and its counterparts in the random and balanced allocation scenarios

Network	c	f	Q <sub>s,2</sub>	Q <sub>s,3</sub>	q <sub>s</sub> (0.91)	IQRs	$\bar{s}$
$H_{or}$			0	$4.0 \cdot 10^{-2}$	$2.3 \cdot 10^{-1}$	$4.0 \cdot 10^{-2}$	$5.9 \cdot 10^{-2}$
$H_{or,b}$	2		0	$6.5 \cdot 10^{-2}$	$1.9 \cdot 10^{-1}$	$6.5 \cdot 10^{-2}$	$5.2 \cdot 10^{-2}$
$H_{or,b}$	10		0	$6.2 \cdot 10^{-2}$	$1.8 \cdot 10^{-1}$	$6.2 \cdot 10^{-2}$	$5.0 \cdot 10^{-2}$
$H_{or,b}$	100		0	$6.2 \cdot 10^{-2}$	$1.8 \cdot 10^{-1}$	$6.2 \cdot 10^{-2}$	$5.0 \cdot 10^{-2}$
$H_{1,b}$	2		$1.2 \cdot 10^{-2}$	$3.7 \cdot 10^{-2}$	$6.4 \cdot 10^{-2}$	$3.7 \cdot 10^{-2}$	$2.3 \cdot 10^{-2}$
$H_{1,b}$	10		$3.9 \cdot 10^{-3}$	$2.8 \cdot 10^{-2}$	$4.9 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$
$H_{1,b}$	100		0	$2.2 \cdot 10^{-2}$	$3.9 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$
$H_1$	2		$3.2 \cdot 10^{-3}$	$2.0 \cdot 10^{-2}$	$4.3 \cdot 10^{-2}$	$2.0 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$
$H_1$	10		$3.4 \cdot 10^{-4}$	$1.2 \cdot 10^{-2}$	$3.0 \cdot 10^{-2}$	$1.2 \cdot 10^{-2}$	$9.0 \cdot 10^{-3}$
$H_1$	100		0	$7.9 \cdot 10^{-3}$	$2.3 \cdot 10^{-2}$	$7.9 \cdot 10^{-3}$	$6.5 \cdot 10^{-3}$
$H_{2,b}$	2	1	$4.5 \cdot 10^{-2}$	$8.1 \cdot 10^{-2}$	$1.2 \cdot 10^{-1}$	$8.1 \cdot 10^{-2}$	$5.4 \cdot 10^{-2}$
$H_{2,b}$	2	10	$4.6 \cdot 10^{-2}$	$8.1 \cdot 10^{-2}$	$1.2 \cdot 10^{-1}$	$8.1 \cdot 10^{-2}$	$5.4 \cdot 10^{-2}$
$H_{2,b}$	10	1	$4.2 \cdot 10^{-2}$	$7.5 \cdot 10^{-2}$	$1.2 \cdot 10^{-1}$	$7.5 \cdot 10^{-2}$	$5.1 \cdot 10^{-2}$
$H_{2,b}$	10	10	$4.2 \cdot 10^{-2}$	$7.5 \cdot 10^{-2}$	$1.2 \cdot 10^{-1}$	$7.5 \cdot 10^{-2}$	$5.1 \cdot 10^{-2}$
$H_{2,b}$	100	1	$4.1 \cdot 10^{-2}$	$7.4 \cdot 10^{-2}$	$1.1 \cdot 10^{-1}$	$7.4 \cdot 10^{-2}$	$5.0 \cdot 10^{-2}$
$H_{2,b}$	100	10	$4.1 \cdot 10^{-2}$	$7.4 \cdot 10^{-2}$	$1.1 \cdot 10^{-1}$	$7.4 \cdot 10^{-2}$	$5.0 \cdot 10^{-2}$
$H_2$	2	1	$2.0 \cdot 10^{-2}$	$5.0 \cdot 10^{-2}$	$8.8 \cdot 10^{-2}$	$5.0 \cdot 10^{-2}$	$3.4 \cdot 10^{-2}$
$H_2$	2	10	$1.8 \cdot 10^{-2}$	$4.4 \cdot 10^{-2}$	$7.6 \cdot 10^{-2}$	$4.4 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$
$H_2$	10	1	$1.8 \cdot 10^{-2}$	$4.7 \cdot 10^{-2}$	$8.2 \cdot 10^{-2}$	$4.7 \cdot 10^{-2}$	$3.1 \cdot 10^{-2}$
$H_2$	10	10	$1.6 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	$7.0 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	$2.6 \cdot 10^{-2}$
$H_2$	100	1	$1.8 \cdot 10^{-2}$	$4.6 \cdot 10^{-2}$	$8.0 \cdot 10^{-2}$	$4.6 \cdot 10^{-2}$	$3.0 \cdot 10^{-2}$
$H_2$	100	10	$1.6 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	$6.9 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	$2.5 \cdot 10^{-2}$

Indeed, many portfolios do not overlap at all with many others and the corresponding similarity matrix is sparse. If we discard zeros, we obtain the conditioned similarity distribution that is summarized in Table 4.

Shuffling of holdings suppresses average similarity the most. However, portfolio overlap is sensibly reduced in model  $H_2$  too. Portfolio rebalancing has mixed effects: it reduces the average similarity in the real network and increases it if performed after randomization of the holdings. We also note that a larger  $c$  typically results in a lower similarity because it corresponds to more outstanding shares and weaker constraints on random stock selection. Moreover, performing more swaps by increasing  $f$  in model  $H_2$  further reduces  $\bar{s}$  in the unbalanced case, but its effects are nearly undetectable after a portfolio rebalance. In Fig. 1 we summarize part of the previous information for the diversification and the conditioned similarity by means of boxplots limited to scenario  $c = f = 10$ .

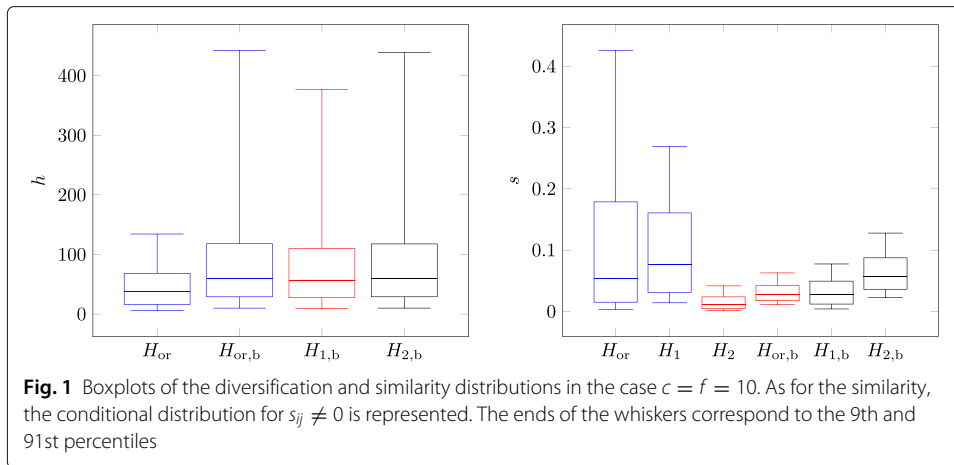
These plots convey the information that portfolio rebalancing stretches the diversification distribution towards larger values, and that the real network topology and its balanced version are characterized by large similarities compared to models  $H_1$  and  $H_2$ . It also shows that all distributions are skewed.

It is useful to represent all models in a parameter space that is specific for the network at hand. Its coordinates are the average values of the diversification index and of the portfolio similarity across the network. We take as a benchmark for these quantities their expected values for a URH network with the same fund degree sequence of  $H_{or}$ , see Eq. (4) and the discussion about the perfectly balanced random holdings network model. For our case network such values are  $h_0 = 157.4$  and  $s_0 = 0.011$  respectively. The first is an exact value because it is completely determined by the funds' degree sequence, while the

**Table 4** Summary statistics of the conditioned portfolio similarity in the real network and its counterparts in the random and balanced allocation scenarios

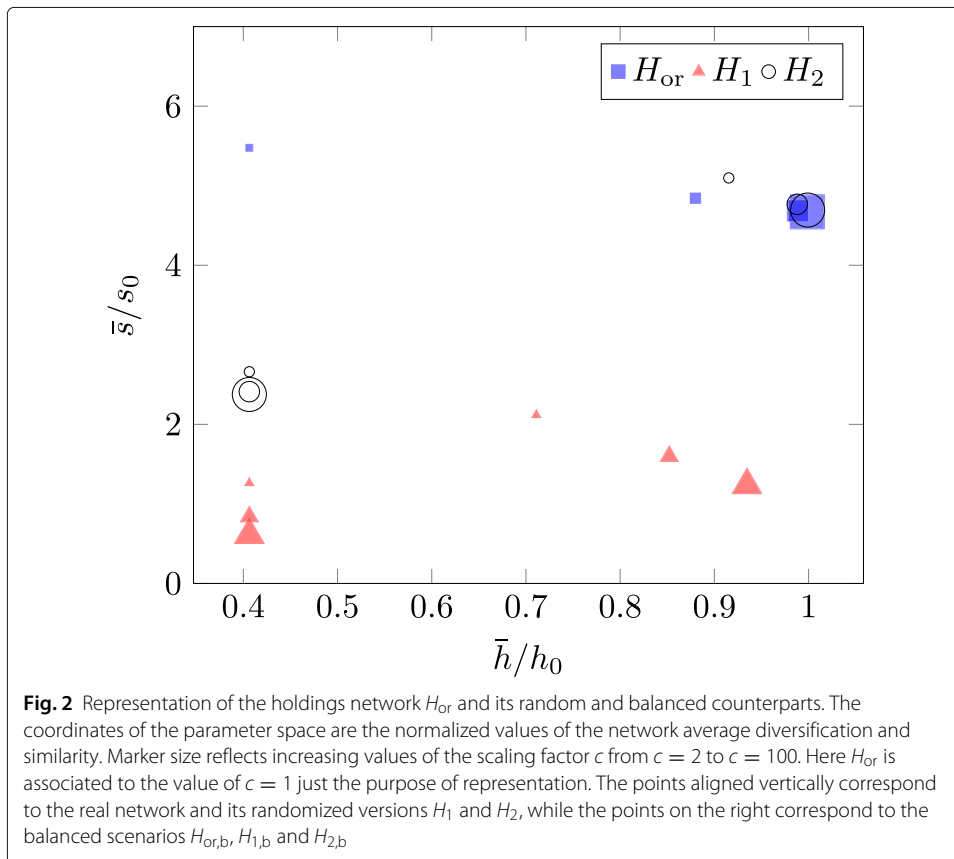
Network	c	f	$q_{s,1}^*$ (0.09)	$Q_{s,1}^*$	$Q_{s,2}^*$	$Q_{s,3}^*$	$q_s^*$ (0.91)	$IQR_s^*$	$s^*$
$H_{or}$			$3.5 \cdot 10^{-3}$	$1.5 \cdot 10^{-2}$	$5.4 \cdot 10^{-2}$	$1.8 \cdot 10^{-1}$	$4.3 \cdot 10^{-1}$	$1.6 \cdot 10^{-1}$	$1.3 \cdot 10^{-1}$
$H_{or,b}$	2		$1.5 \cdot 10^{-2}$	$3.2 \cdot 10^{-2}$	$8.0 \cdot 10^{-2}$	$1.7 \cdot 10^{-1}$	$2.8 \cdot 10^{-1}$	$1.3 \cdot 10^{-1}$	$1.2 \cdot 10^{-1}$
$H_{or,b}$	10		$1.4 \cdot 10^{-2}$	$3.1 \cdot 10^{-2}$	$7.7 \cdot 10^{-2}$	$1.6 \cdot 10^{-1}$	$2.7 \cdot 10^{-1}$	$1.3 \cdot 10^{-1}$	$1.1 \cdot 10^{-1}$
$H_{or,b}$	100		$1.4 \cdot 10^{-2}$	$3.1 \cdot 10^{-2}$	$7.6 \cdot 10^{-2}$	$1.6 \cdot 10^{-1}$	$2.7 \cdot 10^{-1}$	$1.3 \cdot 10^{-1}$	$1.1 \cdot 10^{-1}$
$H_{1,b}$	2		$8.6 \cdot 10^{-3}$	$2.0 \cdot 10^{-2}$	$3.3 \cdot 10^{-2}$	$5.2 \cdot 10^{-2}$	$7.8 \cdot 10^{-2}$	$3.2 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$
$H_{1,b}$	10		$1.1 \cdot 10^{-2}$	$1.8 \cdot 10^{-2}$	$2.7 \cdot 10^{-2}$	$4.2 \cdot 10^{-2}$	$6.3 \cdot 10^{-2}$	$2.5 \cdot 10^{-2}$	$3.3 \cdot 10^{-2}$
$H_{1,b}$	100		$1.0 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$2.3 \cdot 10^{-2}$	$3.5 \cdot 10^{-2}$	$5.2 \cdot 10^{-2}$	$2.0 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$
$H_1$	2		$2.3 \cdot 10^{-3}$	$6.9 \cdot 10^{-3}$	$1.6 \cdot 10^{-2}$	$3.2 \cdot 10^{-2}$	$5.5 \cdot 10^{-2}$	$2.5 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$
$H_1$	10		$1.5 \cdot 10^{-3}$	$4.7 \cdot 10^{-3}$	$1.2 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$	$4.2 \cdot 10^{-2}$	$1.9 \cdot 10^{-2}$	$1.8 \cdot 10^{-2}$
$H_1$	100		$1.0 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$	$9.0 \cdot 10^{-3}$	$1.9 \cdot 10^{-2}$	$3.4 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$1.4 \cdot 10^{-2}$
$H_{2,b}$	2	1	$2.4 \cdot 10^{-2}$	$3.8 \cdot 10^{-2}$	$6.2 \cdot 10^{-2}$	$9.4 \cdot 10^{-2}$	$1.4 \cdot 10^{-1}$	$5.6 \cdot 10^{-2}$	$7.3 \cdot 10^{-2}$
$H_{2,b}$	2	10	$2.4 \cdot 10^{-2}$	$3.9 \cdot 10^{-2}$	$6.2 \cdot 10^{-2}$	$9.4 \cdot 10^{-2}$	$1.4 \cdot 10^{-1}$	$5.5 \cdot 10^{-2}$	$7.3 \cdot 10^{-2}$
$H_{2,b}$	10	1	$2.3 \cdot 10^{-2}$	$3.6 \cdot 10^{-2}$	$5.8 \cdot 10^{-2}$	$8.8 \cdot 10^{-2}$	$1.3 \cdot 10^{-1}$	$5.2 \cdot 10^{-2}$	$6.9 \cdot 10^{-2}$
$H_{2,b}$	10	10	$2.3 \cdot 10^{-2}$	$3.6 \cdot 10^{-2}$	$5.8 \cdot 10^{-2}$	$8.7 \cdot 10^{-2}$	$1.3 \cdot 10^{-1}$	$5.1 \cdot 10^{-2}$	$6.9 \cdot 10^{-2}$
$H_{2,b}$	100	1	$2.2 \cdot 10^{-2}$	$3.5 \cdot 10^{-2}$	$5.7 \cdot 10^{-2}$	$8.7 \cdot 10^{-2}$	$1.3 \cdot 10^{-1}$	$5.1 \cdot 10^{-2}$	$6.8 \cdot 10^{-2}$
$H_{2,b}$	100	10	$2.2 \cdot 10^{-2}$	$3.6 \cdot 10^{-2}$	$5.7 \cdot 10^{-2}$	$8.6 \cdot 10^{-2}$	$1.3 \cdot 10^{-1}$	$5.1 \cdot 10^{-2}$	$6.8 \cdot 10^{-2}$
$H_2$	2	1	$4.9 \cdot 10^{-3}$	$1.5 \cdot 10^{-2}$	$3.4 \cdot 10^{-2}$	$6.2 \cdot 10^{-2}$	$9.9 \cdot 10^{-2}$	$4.7 \cdot 10^{-2}$	$4.6 \cdot 10^{-2}$
$H_2$	2	10	$4.5 \cdot 10^{-3}$	$1.3 \cdot 10^{-2}$	$3.0 \cdot 10^{-2}$	$5.4 \cdot 10^{-2}$	$8.4 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	$3.8 \cdot 10^{-2}$
$H_2$	10	1	$4.4 \cdot 10^{-3}$	$1.4 \cdot 10^{-2}$	$3.1 \cdot 10^{-2}$	$5.8 \cdot 10^{-2}$	$9.2 \cdot 10^{-2}$	$4.4 \cdot 10^{-2}$	$4.2 \cdot 10^{-2}$
$H_2$	10	10	$4.1 \cdot 10^{-3}$	$1.2 \cdot 10^{-2}$	$2.7 \cdot 10^{-2}$	$4.9 \cdot 10^{-2}$	$7.7 \cdot 10^{-2}$	$3.7 \cdot 10^{-2}$	$3.5 \cdot 10^{-2}$
$H_2$	100	1	$4.3 \cdot 10^{-3}$	$1.3 \cdot 10^{-2}$	$3.1 \cdot 10^{-2}$	$5.7 \cdot 10^{-2}$	$9.0 \cdot 10^{-2}$	$4.3 \cdot 10^{-2}$	$4.1 \cdot 10^{-2}$
$H_2$	100	10	$4.0 \cdot 10^{-3}$	$1.2 \cdot 10^{-2}$	$2.7 \cdot 10^{-2}$	$4.9 \cdot 10^{-2}$	$7.7 \cdot 10^{-2}$	$3.7 \cdot 10^{-2}$	$3.4 \cdot 10^{-2}$

Conditioned statistics are distinguished from the unconditioned ones by an asterisk superscript



second value was obtained as a sample mean of the average network similarity for 100 Monte Carlo realizations of model URH. We then consider rescaled coordinates  $\bar{h}/h_0$  and  $\bar{s}/s_0$  for all scenarios under investigation. These are useful to see at a glance how much actual values deviate from a completely random case. The original network, the balanced and randomized counterparts are then represented as in Fig. 2, where the marker size is increasing with the value of the scaling factor  $c$ .

The plot supports the previous considerations about the statistical distributions of the  $h$  and  $s$  and provides some more insight. In the original network and the randomized



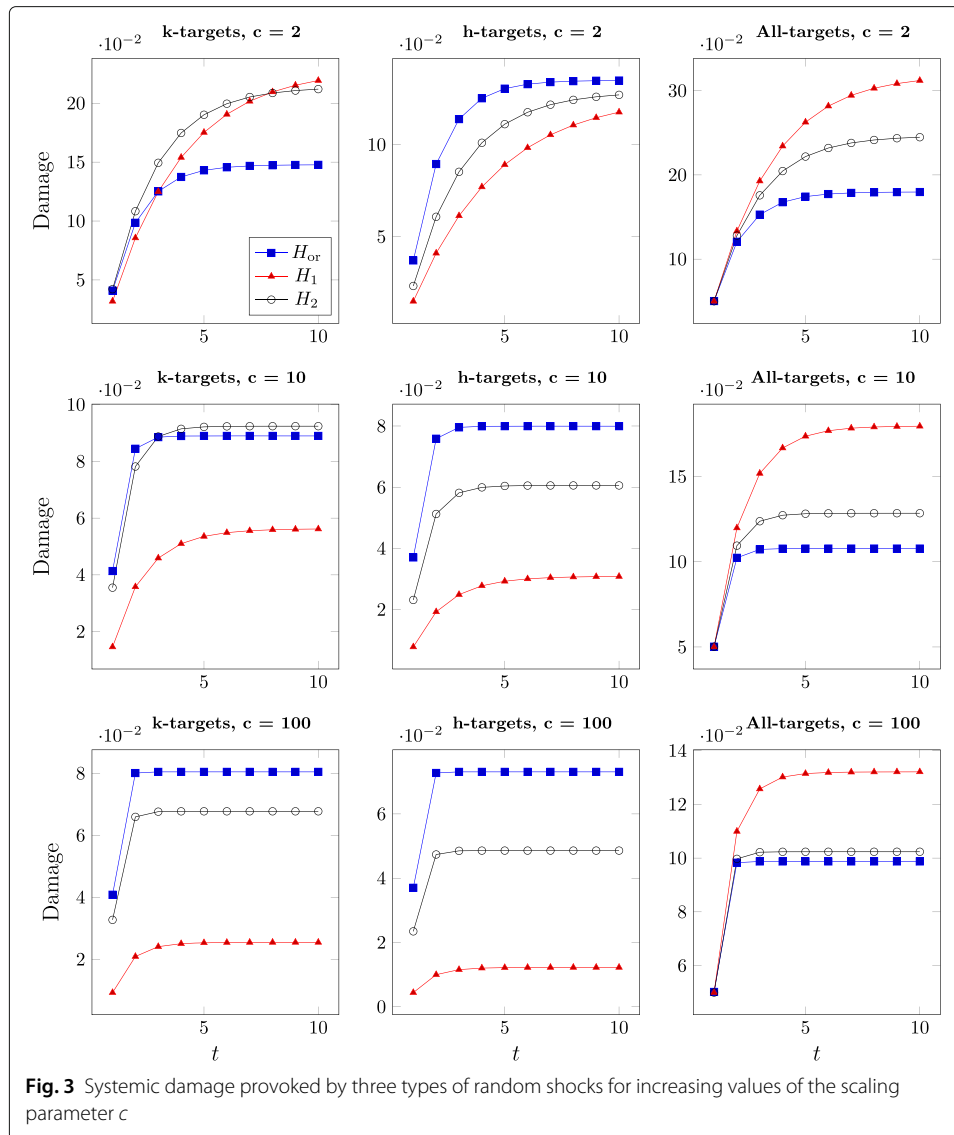
scenarios without a rebalancing, the average diversification is considerably smaller than for a URH network. The balanced models get a higher diversification that converges to the benchmark  $h_0$  for increasing  $c$ . The real network exhibits an average similarity that is more than five times larger than  $s_0$ , in agreement with the findings of Delpini et al. (2019). Model  $H_2$  preserves the strongly connected stocks but still the average similarity is reduced noticeably. Model  $H_1$  has similarity close to  $s_0$ . Rebalancing increases  $\bar{s}$  for both  $H_1$  and  $H_2$  and the similarity gain is decreasing with  $c$ . Noteworthy, it has a slightly suppressive effect for the original network, as noted earlier. This indicates that holders correlate with respect to both the choice of which stocks to buy and their proportions, and that unselective rebalancing of all positions mitigate such correlation. We also notice that for large  $c$ , scenarios  $H_{or,b}$  and  $H_{2,b}$  tend to converge to the same region of the parameter space. Overall, this plot shows that it is difficult to steer the real network toward benchmark values of both diversification and similarity: the holding reshuffling brings the average similarity closer to  $h_0$  but the rebalancing needed to raise  $\bar{h}$  comes with a new similarity gain.

A major goal of this analysis is the study of systemic fragility in the light of the global market constraints and the different diversification and similarity profiles of the considered scenarios. To this end we performed simulations of the flow-induced trading dynamics of shock propagation. This was done for three kinds of exogenous shocks corresponding to the schemes “ $k$ -targets”, “ $h$ -targets” and “all-targets” introduced in “Systemic damage” section. In Fig. 3 the damage curves of the original network and the random unbalanced scenarios are compared for varying  $c$ .

The actual fragility crucially depends on how the exogenous shock is applied and the relative market size of the network. When the most popular stocks get hit and  $c = 10$ ,  $H_{or}$  and  $H_2$  are the most risky scenarios and undergo similar damages. The shuffled network  $H_1$  is significantly safer. As  $H_2$  preserves strongly connected stocks while shuffling avoids them, we guess that in this case the effects of hubs dominate. When the shock affects the most owned stocks (those with the highest  $h$ ), the systemic damage reduces, especially in scenarios  $H_1$  and  $H_2$ , while the real network is the most risky. Under scheme “ $h$ -targets”, weights play a crucial role and the gap between  $H_{or}$  and  $H_2$  is most evident. However, when all stocks in the market receive a shock (“all-targets” case), we observe a reversal of the damage curves, with  $H_{or}$  becoming the safest configuration and  $H_1$  the most fragile. This is a major point, because market crashes can come with negative returns that are widespread across stocks. The figure also shows that the systemic damage becomes saturated after just few periods, which supports consideration of our framework for stress testing holdings networks over short time horizons. Similar trends are observed when  $c = 100$  but with a reduced damage, as expected since stocks have larger market depths. When  $c = 2$ , the network accounts for half the market and  $H_{or}$  still provides the most robust scenario in front of a shock that interests all the stocks. Unexpectedly, in this case it is also the most robust under the “ $k$ -targets” stress test.

What are the systemic effects of making portfolios less concentrated? Figure 4 shows the damage ratio for each balanced scenario to its unbalanced counterpart and provides an answer.

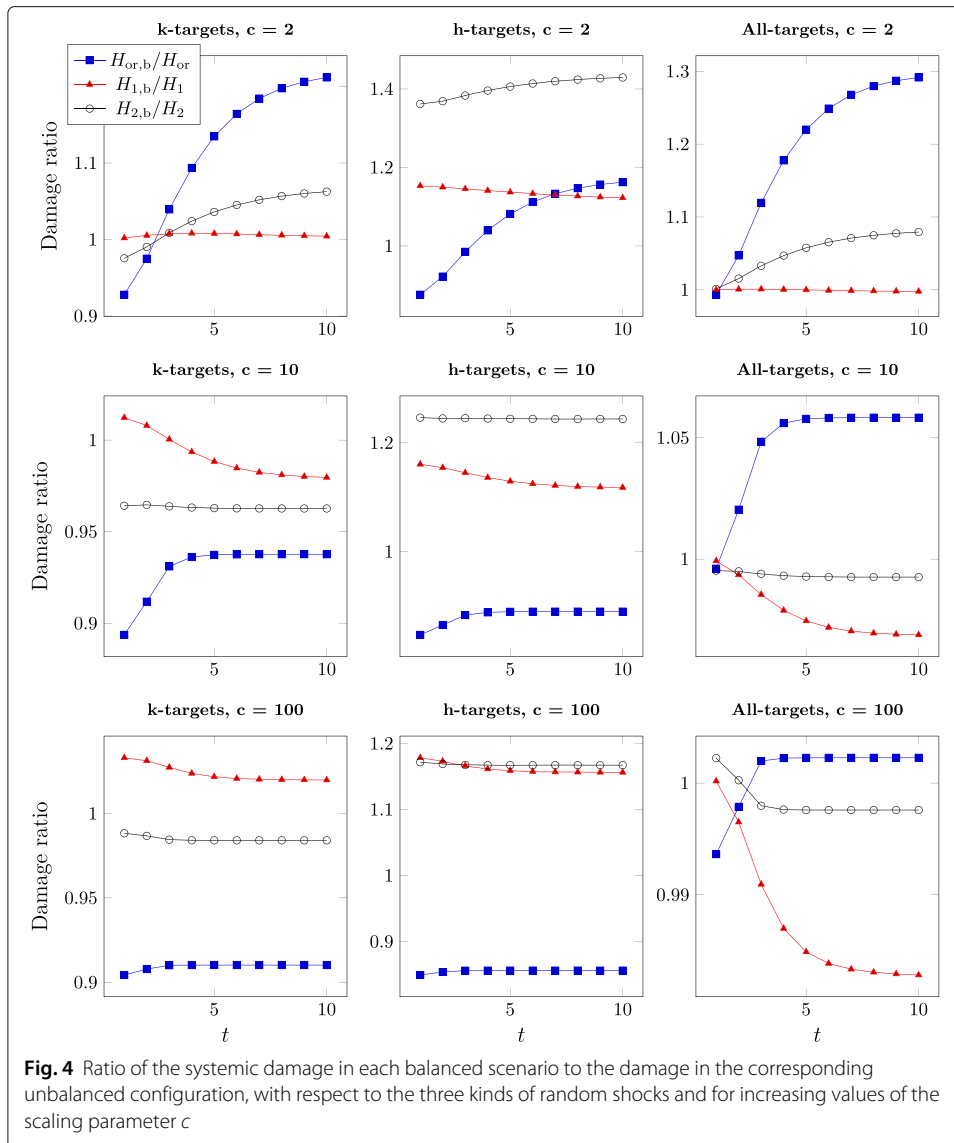
Again, it depends on the network topology and portfolio weights distribution, as well as strongly on how we stress the system. For targeted shocks, rebalancing real portfolios is beneficial as long as the network is relatively small, while for  $c = 2$  a more diversified



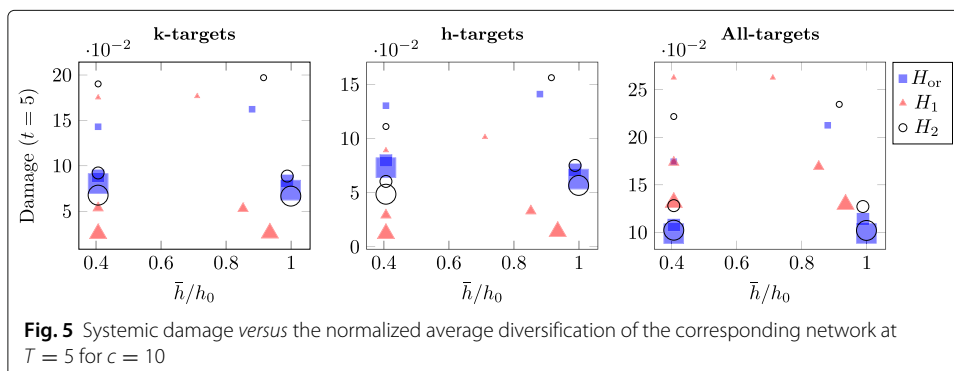
network with the real topology eventually becomes more risky. In front of a widespread shock, rebalancing is always detrimental to  $H_{or}$  after a while. The systemic loss is between few points and 30% larger than it was for the original network, depending on  $c$ , and such negative systemic effect persists over time. Turning to random model  $H_1$ , rebalancing has slightly positive effects under unselective shocks. Under targeted shocks, increasing diversification results in a more risky network, with the exception of  $c = 10$  where  $H_{1,b}$  turns safer after a transient. Model  $H_2$  behaves qualitatively like  $H_{or}$  in the “ $k$ -targets” stress test, while the effects of rebalancing for this model are more similar to  $H_1$  in the other cases. Noteworthy, when we target the most owned stocks, the negative effects of rebalancing for  $H_2$  are severe and persistent over time for all values of  $c$ . It also worth noting that in some circumstances, the effect of rebalancing depends critically on the number of periods elapsed. It can start positive and become negative after a transient, or the other way round.

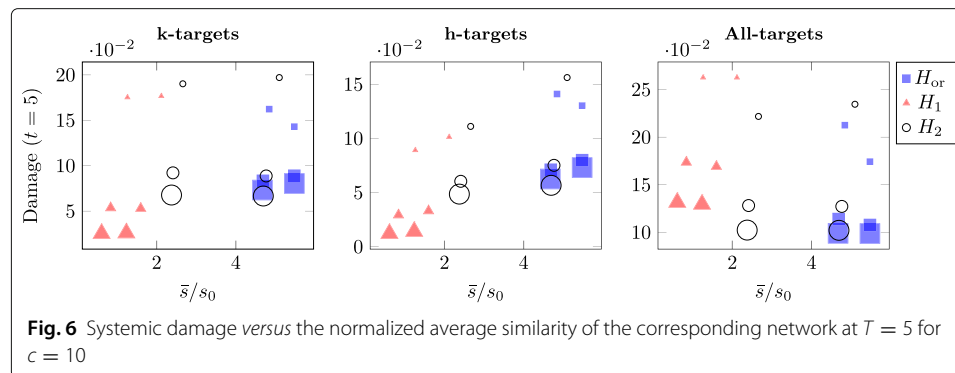
We can abridge the most of these findings in the representations of Figs. 5 and 6.





In the first figure the systemic loss after half the trading periods is shown against the normalized network diversification. No evident correlation can be detected in this case and, to parity of diversification, fragility is strongly mediated by the type of shock applied and the relative size of the network. In Fig. 6 a comparison of the systemic loss and the corresponding normalized network similarity is performed.





We still have a strong interplay between fragility, network size and shock type. However, a positive correlation between the systemic damage and the average similarity is observed under targeted shocks, with a Pearson correlation coefficient of  $\rho_k = 0.3$  and  $\rho_h = 0.6$  respectively. Such correlation becomes negative under unselective shocks, corresponding to  $\rho_{\text{all}} = -0.3$ . This figure delivers useful insight on the effects of investment similarities and we believe that the previous findings represent a major contribution of this work.

## Conclusions

Nowadays, it is of the utmost importance to quantify riskiness in financial systems, especially when investments in foreign assets can provide a global outreach to the propagation of financial distress. Depending on market conditions and the simultaneous activity of investors, a systemic risk component may emerge if diversification strategies are similar. Such effect was suggested in Delpini et al. (2019), and supported by simulations of distress propagation in the holdings network of US mutual funds. The random models behind that analysis were admittedly simple. It was assumed that funds can reallocate their positions freely, without any constraints on the amount of shares available from the market. Also, that analysis did not take into consideration the possibility for asset managers to increase diversification by balancing positions without modifying the network's topology.

In this paper, we improved and extended the previous study of a bipartite network of portfolio holdings. We adopted a more general point of view and modeled the network as a subsystem of the whole market. We explicitly took into account the global constraints posed by the limited numbers of shares in the market. Such number changes from stock to stock and accounts for heterogeneity in stock liquidity characteristics. We exploited more sophisticated algorithms that are required to satisfy such global constraints at any time of their execution. By their means, we could generate synthetic scenarios of both balanced and unbalanced portfolios and perform an extensive computational analysis of how such scenarios react to different types of exogenous shocks. We considered a dynamics of distress propagation where the numbers of shares determine the sensitivity of each stock's price. Such parameters are the same for the original network and its random counterparts, which now allows to perform a more consistent comparison. For every scenario we also simulated its balanced counterpart, representing a case where asset managers try to rebalance all their positions in the quest for a higher degree of diversification.

We found that randomization of the original holdings for fixed portfolio diversification has a strong suppressive effect on similarity even under the global constraints. This effect becomes stronger as the relative market value of the network decreases. We showed that

a large increase of the average diversification can be achieved by portfolio rebalancing, even for a moderate number of outstanding shares. We also provided a convenient representation of scenarios in a diversification–similarity space of coordinates relative to an unconstrained random model of balanced investments. With respect to such benchmark, the real holdings network has small average diversification but strong average portfolio similarity, and a significant increase in both differentiation and diversification can not be achieved trivially.

We then performed an extensive comparison of the different scenarios in terms of the systemic damage from an exogenous shock applied to the corresponding network. We considered both the case of targeted shocks to the most popular or most owned stocks, as well as a widespread random shock to the prices of all stocks. Results show that there is an interplay between diversification and investment differentiation, which varies across the different network topologies. Network's fragility can depend to a large extent both on the way shocks are applied and on the relative size of the network, and results can be unexpected to a degree. Overall, a correlation is found between systemic risk and portfolio similarity. Such correlation is positive under targeted shocks and negative under unselective ones. A rebalancing of the real portfolios makes the network less prone to large damages in front of targeted shocks, provided that the network can be considered small with respect to the market. Remarkably, that also makes it riskier under widespread shocks. This means that increasing diversification may possibly be detrimental in systemic terms depending on market conditions. An initial transient of time where rebalancing changes from beneficial to detrimental, or the opposite, can be observed. Additionally, our results show that holdings reshuffling do not automatically results in a safer network. In particular, in a scenario where all stocks may undergo a negative downturn, the real network provides a safer environment regardless of its relative size.

We believe that our findings can be of major interest when it comes to assessing systemic risk properly and for informing effective policy actions. As a further perspective, the effects of diversification-increasing strategies that change the holders' degree sequence could be investigated. Also, a higher-statistics Monte Carlo study of each random scenario for a fine-grained grid of model parameters would provide a better coverage and broader insight.

#### **Abbreviations**

TNA: Total net assets; URH: Unconstrained random holdings; USD: US dollar

#### **Acknowledgements**

Not applicable.

#### **Authors' contributions**

DD devised the modeling aspects and the visualizations, wrote the algorithms, performed the calculations and the simulations. All authors conceived the idea, analyzed and interpreted the results, and wrote the manuscript. The author(s) read and approved the final manuscript.

#### **Funding**

DD acknowledges funding "Fondo di Ateneo per la Ricerca 2019" provided by the University of Sassari.

#### **Availability of data and materials**

The findings of this study are based on data from the Survivor-Bias-Free US Mutual Funds database © 2012 Center for Research in Security Prices (CRSP), The University of Chicago Booth School of Business. Restrictions apply to the availability of these data, which were used under license and so are not publicly available.

#### **Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Dept. of Economics and Business, University of Sassari, Via Muroni 25, 07100 Sassari, Italy. <sup>2</sup>Dept. of Banking and Finance, University of Zurich, Plattenstrasse 32, 8032 Zurich, Switzerland. <sup>3</sup>Dept. of Molecular Sciences and Nanosystems, and European Centre for Living Technology, Ca' Foscari University of Venice, Ca' Bottacin, Dorsoduro 3911, Calle Crosera, 30123 Venice, Italy. <sup>4</sup>Institute for Complex Systems (CNR) UoS Università "Sapienza", P.le A. Moro 2, 00185 Rome, Italy. <sup>5</sup>IMT School for Advanced Studies, P.zza S. Francesco 19, 55100 Lucca, Italy.

Received: 11 March 2020 Accepted: 30 June 2020

Published online: 14 July 2020

**References**

- Acemoglu D, Ozdaglar A, Tahbaz-Salehi A (2015) Systemic Risk and Stability in Financial Networks. *Am Econ Rev* 105(2):564–608. <https://doi.org/10.1257/aer.20130456>
- Agnew J, Balduzzi P, Sundén A (2003) Portfolio Choice and Trading in a Large 401(k) Plan. *Am Econ Rev* 93(1):193–215. <https://doi.org/10.1257/000282803321455223>
- Augustiani C, Casavecchia L, Gray J (2015) Managerial Sharing, Mutual Fund Connections, and Performance: Mutual Fund Connections and Performance. *Int Rev Finan* 15(3):427–455. <https://doi.org/10.1111/irfi.12054>
- Battiston S, Gatti DD, Gallegati M, Greenwald B, Stiglitz JE (2012a) Default cascades: When does risk diversification increase stability?. *J Financ Stab* 8(3):138–149. <https://doi.org/10.1016/j.jfs.2012.01.002>
- Battiston S, Gatti DD, Gallegati M, Greenwald B, Stiglitz JE (2012b) Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *J Econ Dyn Control* 36(8):1121–1141. <https://doi.org/10.1016/j.jedc.2012.04.001>
- Braverman A, Minca A (2018) Networks of common asset holdings: aggregation and measures of vulnerability. *J Netw Theory Financ* 4(3):53–78. <https://doi.org/10.21314/JNTF.2018.045>
- Caccioli F, Shrestha M, Moore C, Farmer JD (2014) Stability analysis of financial contagion due to overlapping portfolios. *J Bank Financ* 46:233–245. <https://doi.org/10.1016/j.jbankfin.2014.05.021>
- Caldarelli G (2007) Scale-free Networks. *Complex Webs in Nature and Technology*. Oxford University Press, Oxford
- Delpini D, Battiston S, Caldarelli G, Riccaboni M (2019) Systemic risk from investment similarities. *PLoS ONE* 14(5):0217141. <https://doi.org/10.1371/journal.pone.0217141>
- Delpini D, Battiston S, Riccaboni M, Gabbi G, Pammolli F, Caldarelli G (2013) Evolution of Controllability in Interbank Networks. *Sci Rep* 3(1):1626. <https://doi.org/10.1038/srep01626>
- Domian DL, Louton DA, Racine MD (2003) Portfolio diversification for long holding periods: how many stocks do investors need? *Stud Econ Financ* 21(2):40–64. <https://doi.org/10.1108/eb028774>
- Domian DL, Louton DA, Racine MD (2007) Diversification in Portfolios of Individual Stocks: 100 Stocks Are Not Enough. *Financ Rev* 42(4):557–570. <https://doi.org/10.1111/j.1540-6288.2007.00183.x>
- Elliott M, Golub B, Jackson MO (2014) Financial Networks and Contagion. *Am Econ Rev* 104(10):3115–3153. <https://doi.org/10.1257/aer.104.10.3115>
- Fricke D (2019) Are specialist funds “special”? *Financ Manag* 48(2):441–472. <https://doi.org/10.1111/fima.12257>
- Galbiati M, Delpini D, Battiston S (2013) The power to control. *Nat Phys* 9(3):126–128. <https://doi.org/10.1038/nphys2581>
- Greenwood R, Landier A, Thesmar D (2015) Vulnerable banks. *J Financ Econ* 115(3):471–485. <https://doi.org/10.1016/j.jfineco.2014.11.006>
- Guo W, Minca A, Wang L (2016) The topology of overlapping portfolio networks. *Stat Risk Model* 33(3-4). <https://doi.org/10.1515/strm-2015-0020>
- Hu J-L, Chang T-P, Chou RY (2014) Market conditions and the effect of diversification on mutual fund performance: should funds be more concentrative under crisis? *J Prod Anal* 41(1):141–151. <https://doi.org/10.1007/s11123-012-0331-x>
- Huang X, Vodenska I, Havlin S, Stanley HE (2013) Cascading Failures in Bi-partite Graphs: Model for Systemic Risk Propagation. *Sci Rep* 3(1):1219. <https://doi.org/10.1038/srep01219>
- Hurd M, Rohwedder S (2010) Effects of the Financial Crisis and Great Recession on American Households. Working Papers. RAND Corporation, Santa Monica. <https://doi.org/10.7249/WR810>. [https://www.rand.org/pubs/working\\_papers/WR810.html](https://www.rand.org/pubs/working_papers/WR810.html)
- Kyle AS (1985) Continuous Auctions and Insider Trading. *Econometrica* 53(6):1315. <https://doi.org/10.2307/1913210>
- Lin L, Guo X-Y (2019) Identifying fragility for the stock market: Perspective from the portfolio overlaps network. *J Int Financ Mark* 62:132–151. <https://doi.org/10.1016/j.intfin.2019.07.001>
- Newman MEJ (2010) *Networks: An Introduction*. Oxford University Press, Oxford
- Statman M (1987) How Many Stocks Make a Diversified Portfolio?. *J Financ Quant Anal* 22(3):353. <https://doi.org/10.2307/2330969>
- Statman M (2004) The Diversification Puzzle. *Financ Anal J* 60(4):44–53. <https://doi.org/10.2469/faj.v60.n4.2636>

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.