PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Cite this article: Bozhidarova M, Ball F, van Gennip Y, O'Dea RD, Stupfler G. 2024 Describing financial crisis propagation through epidemic modelling on multiplex networks. *Proc. R. Soc. A* **480**: 20230787. https://doi.org/10.1098/rspa.2023.0787

Received: 23 October 2023
Accepted: 29 February 2024

Subject Areas:
applied mathematics, mathematical finance, mathematical modelling

Keywords:
epidemic modelling, financial crisis, multiplex network, SIR model, tail dependence

Author for correspondence:
Malvina Bozhidarova
e-mail:
Malvina.Bozhidarova@nottingham.ac.uk

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7131861.

THE ROYAL SOCIETY PUBLISHING

Describing financial crisis propagation through epidemic modelling on multiplex networks

Malvina Bozhidarova[1], Frank Ball[1], Yves van Gennip[2], Reuben D. O'Dea[1] and Gilles Stupfler[3]

[1]School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, UK
[2]Delft Institute of Applied Mathematics (DIAM), Technische Universiteit Delft, Delft, The Netherlands
[3]Univ Angers, CNRS, LAREMA, SFR MATHSTIC, 49000 Angers, France

MB, 0009-0005-4339-9925; FB, 0000-0002-5599-2903; YvG, 0000-0003-4953-8314; RDO, 0000-0002-1284-9103; GS, 0000-0003-2497-9412

This paper proposes a novel framework for modelling the spread of financial crises in complex networks, combining financial data, Extreme Value Theory and an epidemiological transmission model. We accommodate two key aspects of contagion modelling: fundamentals-based contagion, where the transmission is due to direct financial linkages, and pure contagion, where a crisis might trigger additional crises due to global effects. We use stock price, geographical location and economic sector data for a set of 398 companies to construct multiplex networks of four layers, on which a susceptible-infected-recovered transmission model is defined, in order to model the spread of financial shocks between companies by accounting for their interconnected nature. By utilizing stock price data for the 2008 and 2020 financial crises, we investigate and assess the effectiveness of our model in forecasting the propagation of financial shocks through the network, where a shock is detected by measuring stock price volatility. The results suggest that the proposed framework is effective in predicting the spread of financial crises. Our findings demonstrate the significance of each layer of the multiplex network

structure, which differentiates between various transmission pathways, for predicting the number of affected companies, as well as for company-, sector- or location-specific predictions.

1. Introduction

Global financial stability has become one of the key concerns of economic policy-makers and decision-makers due to the increasing frequency, magnitude and international scope of financial crises [1]. Interconnectedness is a key feature of the global financial system, in which companies can be connected in multiple ways, such as via their claims and obligations towards one another [2], or through transactions between them [3], forming a network structure. Understanding how financial crises experienced by certain companies or sectors can spread, potentially leading to wider crises, is self-evidently of interest to policy-makers, investors and business owners, as even minor disruptions in a single company can result in long-term issues and losses, as well as a global financial crisis [4]. The network structure is crucial in determining how the initial shock spreads across the system. Therefore, rather than being viewed as a standalone entity, a company and the risk it confronts should be assessed in conjunction with the network of companies with which it interacts and the wider financial environment in which it operates. Allen & Gale [5] explore the influence of network topology on the propagation of risk in financial systems and they emphasize that the existence of network connections can generate channels for the spread of contagion, leading to an increased probability of risk transmission within the network. Since the publication of this seminal paper, network models have become increasingly common in theoretical and empirical studies of financial contagion.

The term 'financial contagion' first appeared in 1997 during the Asian crisis, alongside which, the Russian Default of 1998 and the Global Financial Crisis of 2008 are among the recent events that are thought to be results of contagion spread [6]. There are various methods that have been proposed to model the spread of financial contagion in networks; popular approaches include random graph models as well as those in which correlations (such as Pearson correlation) in financial data are used to build complex financial networks. In the former, the employed network structure may not reflect the real-world structure of financial networks since the latter are often characterized by a high degree of clustering and heterogeneity [7], which are not captured by many random graph models. The Pearson correlation coefficient, meanwhile, is unable to represent nonlinear dependencies between risky asset returns or tail dependence in the data. As an alternative, extreme value theory-based (EVT-based) approaches enable the measurement of nonlinear dependence in the tail of the distributions. One of the most common statistical concepts for computation of extreme risk in EVT is extremal dependence, for example through the tail dependence coefficient [8,9].

In their comprehensive review [10], the authors show that methodologies from disciplines such as physics and engineering can be employed to study various fields, including urban development, financial markets, cooperation and social networks. They provide an in-depth study of how econophysics employs the particle model from statistical physics to depict agent behaviour, demonstrating its effectiveness in modelling diverse financial interactions, including those found in financial markets and international trade. Another commonly employed method involves using epidemic models to study complex financial systems. For instance, Lazebnik *et al.* [11] develop a mathematical model integrating epidemiological, social and economic factors to assess policies like work-from-home and vaccination during pandemics. Unlike our focus on financial contagion within companies, this research examines social behaviour and supply chain networks. Results suggest vaccination significantly reduces output loss, especially in industries with close contacts. Other works employ susceptible-infected-recovered (SIR) epidemic models to study financial contagion in the banking sector [12] and between countries [1].

Many of these studies use monolayer networks (i.e. networks in which all the edges represent the same type of connection between the nodes). By contrast, multilayer networks can more

accurately represent interconnected structures [13,14], being able to describe separately different kinds of entities, connections and relationships in each network layer. In such networks, there are two types of edges: intra-layer edges, that connect nodes within the same layer, and inter-layer edges, which connect nodes in different layers. A multiplex network is a restriction of this more general class, where inter-layer links connect instances of the same node in each layer. Recently, multiplex networks have found many applications in finance. For example, the authors of [15] suggest that accounting for both intralayer and interlayer propagation of contagions in a multiplex structure of financial assets is important for understanding interconnected financial systems of countries. In addition, [16] introduces a multilayer network model to analyse systemic risk in China's financial system, examining liability and cross-shareholding among institutions to demonstrate how the network's nonlinear dynamics impact risk spreading and the connection to systemic risk. Other applications of multiplex network are reviewed in detail in [10,14,17], which delve deeper into the intricacies of multilayer network theory.

In this paper, we propose a novel framework for modelling financial contagion that is based on an SIR epidemic model defined on a multiplex network constructed from financial data. We employ a stochastic epidemic transmission mechanism in which financial crises can spread locally (to network neighbours) as well as globally (to any company). Then, by considering their local and global connectivity, we simulate how a financial shock spreads from the original infected companies to the others. To demonstrate our approach, we construct two multiplex networks, representing the financial dependence of 398 companies in the 2008 and the 2020 financial crises, where each node represents a company and each layer represents a different type of connection between the companies. Both networks consist of four layers: a tail dependence network layer, a continents layer, a sectors layer and a global layer. The tail dependence layer measures the strength of dependence between two companies using tail dependence coefficients, which are calculated using daily stock price data. This weighted (complete) network is filtered via the planar maximally filtered graph (PMFG) method [18], to remove weak and potentially spurious links. The continents and sectors layers, respectively, connect companies under the assumption that companies in the same continent or sector are more likely to be affected by a financial crisis simultaneously. Finally, the global layer is a complete network, in which each company is connected to every other company. This layer corresponds to the 'pure contagion' assumption that a crisis in any company may trigger a crisis in any other company. In addition, in our model, a company may experience a financial shock not just as a result of direct linkages to the initially infected company, but also as a result of indirect connections within the network of companies, amplifying the spread and impact of the financial shock. As a result, we allow for the so-called 'cascading effect', a phenomenon where the impacts of a financial crisis spread and intensify through interconnected channels, resulting in a broader and more severe contagion than initially anticipated [19], which is commonly overlooked in the literature.

We apply the model to the recent 2008 and 2020 financial crises and evaluate its utility in predicting the spread of financial shocks across the network. We first identify which companies have been 'infected' in each of the two crises using stock price volatility. We then study how using data from the previous n crisis days to parametrize the transmission model can be used to predict the infections in the future k days for different combinations of n and k. The results suggest that for each crisis a different combination of n and k gives the most accurate predictions. The proposed model outperforms the homogeneous mixing population approach in predicting the number of infected companies, the continents and economic sectors that will be most affected, and the sets of specific companies that will be infected during the future crisis days for both crises.

The remainder of the paper is organized as follows. Section 2 describes the dataset. Then, in §3, we describe the two parts of the modelling framework: the multiplex network construction procedure and the transmission mechanism. In §4, we apply the model to the 2008 and the 2020 financial crises. We first define the concept of infection in a financial context and then we study how the model can be used to predict future infections in each of the two crises, using past data. Finally, we assess the significance of each layer within our multiplex by conducting

a comparative analysis of its predictive accuracy on omission of various subsets of its layers. Section 5 concludes and discusses our findings, the limitations of our approach as well as avenues for further research. We defer additional explanation of our estimation approach and further descriptive and predictive statistical results to the electronic supplementary material.

2. Data

The analysis in this paper is based on the closing daily stock price of 398 companies from 17 January 2002 to 18 July 2022 (inclusive), representing $n = 5229$ trading days. The data are collected from https://finance.yahoo.com/ and the companies are selected such that for each company there are consistent data going back as far as 17 January 2002, covering a sufficient time period before the 2008 financial crisis. We separate the companies into groups, based on the Bureau van Dijk[1] company database. Firstly, we group the companies according to the geographical location of their headquarters, resulting in six groups: Africa (2), Asia (77), Europe (115), North America (194), Oceania (9) and South America (1). The disparity in geographical representation arises from the distribution of available data meeting our date span criteria. North America, for example, has a highly developed and mature financial market and hosts numerous publicly traded companies, many of which have extensive historical data available. This makes it easier to find companies with consistent data spanning back to 2002. By contrast, some regions, especially emerging markets in Africa or South America, may have fewer publicly traded companies or less robust historical financial data, making it more challenging to include a comparable number of companies from those regions in the dataset. Secondly, we separate the companies into 13 groups based on their primary economic sector, as defined by the Bureau van Dijk dataset: Finance (47), Oil and gas industry (36), Pharmaceutical industry (36), Automotive industry (35), Airline industry (17), Food industry (23), Mining activities (20), Electricity (17), Software industry (38), Electronics (58), Telecommunications (10), Chemicals (8) and Others (53).

The stock price returns for each company i at day t for $2 \leq t \leq 5229$ are calculated by taking the logarithmic difference of successive closing prices as follows:

$$x_{i,t} = \ln(p_{i,t}) - \ln(p_{i,t-1}), \tag{2.1}$$

where $p_{i,t}$ denotes the closing stock price of company i at day t for $1 \leq t \leq 5229$. Tables S1 and S2 in the electronic supplementary material show the characteristics of the studied data for the periods prior to, and after, the 2008 financial crisis.

Our objective is to model financial contagion using two different datasets: daily stock prices from the period 17 January 2002 to 30 June 2007, to model contagion during the 2008 financial crisis, and from 17 January 2002 to 29 February 2020 to do so in the 2020 crisis following the onset of the COVID-19 pandemic. We emphasize that here, and in all subsequent occurrences, datasets defined over stated date ranges are understood to be inclusive of start and end dates. The following section gives a thorough explanation of the network construction approach used in our analyses.

3. Model formulation

Our modelling framework comprises two main parts. We first build a multiplex financial network, where the nodes correspond to companies and the edges in each layer represent different types of connections between companies. By incorporating multiple network layers, we can capture the various ways in which financial contagion may spread between companies. We then employ an SIR epidemic model on each network layer. The model's key parameters are the transmission probabilities (i.e. the probability of an infected node transmitting the infection to a susceptible node on a given day) and the recovery probabilities (i.e. the probability that

[1]A significant business information publisher, Bureau van Dijk specializes in private corporate data together with software for searching and analysing businesses.

an infected node becomes recovered on a given day), which we estimate using a maximum-likelihood approach by fitting the model to past crisis data. Then, by simulating the spread of financial contagion using the SIR model with the estimated parameters, we can identify the companies, financial sectors and continents that are predicted to be most vulnerable to future contagion events.

(a) Network model

We construct a multiplex network, where each node represents a company and each layer represents a different type of connection between the companies. We construct four layers: a tail dependence network layer, a continents layer, a sectors layer and a global layer. The motivation for, and method for construction of, these networks are detailed in the following subsections.

(i) Tail dependence network layer

The relationship between tail dependence and the propagation of financial crisis risk is highlighted by a number of studies [6,20,21]. Tail dependence is used to study the likelihood of joint tail events, where the occurrence of extreme movements in one asset's return is associated with a higher likelihood of extreme movements in another. This phenomenon reflects the interconnectedness of financial markets, whereby shocks or disruptions in one asset class or market segment can trigger correlated movements in other assets. The tail dependence coefficient is a common measure of financial dependence between two companies. For example, the concept of marginal expected shortfall (MES), a widely recognized risk measure that evaluates the potential losses of a company given that another experiences an extreme loss, is intricately linked to tail dependence coefficients, thereby underscoring the relevance of tail dependence in capturing the tail behaviour of financial assets [22]. To study how likely it is that two companies experience extreme losses together we construct complex financial networks, via the following two-step process. Firstly, we calculate the tail dependence strength between each pair of companies' stock returns. Secondly, we filter the edge information required for network building using the PMFG approach.

Tail dependence estimation. Let $\{(-x_{i,t}, -x_{j,t}) : t = 1, 2, \ldots, N\}$ be the realizations of the bivariate negative stock return (X_i, X_j), where $x_{i,t}$ is as defined in (2.1). We assume throughout that X_i and X_j have continuous distribution functions. For each pair of negative stock returns (X_i, X_j) of companies i and j, the marginal aspects of the joint distribution can be removed by transforming the bivariate negative returns into unit Fréchet marginals (S_i, S_j) by using the following transformation:

$$S_i = -1/\ln F_i(X_i) \quad \text{and} \quad S_j = -1/\ln F_j(X_j), \tag{3.1}$$

where F_i and F_j are the marginal distribution functions of X_i and X_j, respectively. In practice, the functions F_i and F_j used in (3.1) are estimated by the empirical marginal distribution functions of the two random variables. This transformation does not affect the dependence structure of the bivariate joint distribution, so (S_i, S_j) possesses the same dependence structure as (X_i, X_j).

Since we are interested in the probability that one company experiences an extreme financial loss, given an extreme loss in another (the likelihood of crisis transmission), for each pair (S_i, S_j), we estimate the upper tail dependence coefficient (upper TDC) $\chi^U_{i,j}$, defined as

$$\chi^U_{i,j} = \lim_{q \to 1^-} \mathrm{P}(F_j(S_j) > q \mid F_i(S_i) > q).$$

Hence, the upper TDC corresponds to the likelihood that one margin will surpass a high threshold if the other margin also exceeds this threshold. The coefficient $\chi^U_{i,j}$ takes values in the range $[0, 1]$, describing the strength of the tail dependence between S_i and S_j: $\chi^U_{i,j} = 0$ means that the two variables S_i and S_j are upper tail independent and $\chi^U_{i,j} > 0$ indicates upper tail dependence.

The TDC can also be defined using the concept of a copula, introduced in [23]. A fundamental result shown by Sklar [23] states that $F_{i,j}$, the joint distribution function of (X_i, X_j), can be represented as $F_{i,j}(s_i, s_j) = C_{i,j}(F_i(s_i), F_j(s_j))$, where $C_{i,j}$ is a copula function (a bivariate distribution function with uniform margins). Then, as shown in [24],

$$\chi_{i,j}^U = \lim_{q \to 1^-} \frac{1 - 2q + C_{i,j}(q,q)}{1-q}.$$

In practice, we estimate the strength of tail dependence for each pair (S_i, S_j) and threshold $q \in (0,1)$ as follows:

$$\hat{\chi}_{i,j}^U = \hat{\chi}_{i,j}^U(q) = \frac{1 - 2q + \hat{C}_{i,j}(q,q)}{1-q}, \tag{3.2}$$

where $\hat{C}_{i,j}$, the empirical counterpart of $C_{i,j}$, is computed via

$$\hat{C}_{i,j}(u,v) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}(r_i^n \leq N - \lfloor N(1-u) \rfloor, r_j^n \leq N - \lfloor N(1-v) \rfloor).$$

Here, r_i^n and r_j^n are the ranks (the index of the element in an ascending list) of the nth observations of S_i and S_j, respectively. Note that the transformation in (3.1) is monotonically increasing, so that the rank of an observation from S_i is the same as that for the corresponding X_i.

The analysis in the remainder of the paper is based on the estimated upper tail dependence coefficients $\hat{\chi}_{i,j}^U(0.95)$; i.e. with threshold $q = 0.95$, this choice being consistent with the existing literature using tail dependence to build financial networks [25–27]. Moreover, in this paper, we construct separate networks employing all the data in our set prior to the 2008 and to the 2020 crises, respectively. The upper TDC values between each pair of companies i and j are used to measure the strength of dependence between the companies in our dataset, and are key to our construction of complex financial networks and our SIR model for financial contagion: the higher the TDC between two companies, the higher the probability of crisis transmission.

Planar maximally filtered graph. The PMFG method was first introduced in [18]. The primary goal is to filter complex networks by retaining only the most important links in a way that does not break planarity (i.e. the property of a graph being embeddable in a plane without any intersecting edges) [28]. By doing so PMFGs can assist with eliminating spurious (weak) connections, thereby emphasizing topological properties such as communities and easing computational burden. Planarity also permits more straightforward network visualization and, being maximally filtered, they are constructed in such a way that the number of connections between nodes is maximized while still maintaining planarity.

PMFGs constructed from financial datasets have been used to detect fundamental market changes and community structures [29], to study the spread of financial risk [30] and to analyse financial networks describing correlations (or other dependencies) between financial assets [31]. In addition, PMFGs can be used to reduce the complexity and dimensionality of financial networks, while retaining the clustering structure [32]. Prior to the study of [32], the two most popular tools for filtering the edge information in complex financial networks were the minimum spanning tree (MST) algorithm [33] and the correlation coefficient Threshold method [34]. However, the latter is extremely dependent on the threshold decision [35]; for the former, the key advantage of the PMFG algorithm is that it preserves more information: the MST has $n-1$ edges, while the PMFG has $3(n-2)$ edges (compared to $n(n-1)/2$ of the complete network with n nodes). Furthermore, the PMFG always contains the MST, so it is a connected network.

(ii) Additional layers

In addition to the tail dependence network layer (hereafter denoted PMFG layer for brevity), we include layers to incorporate other known relations between the companies and describe other possible crisis transmission channels.

Sector and continents layers. The 2008 financial crisis demonstrated the importance of interconnectedness as it quickly spread from the subprime mortgage market in the USA to the wider financial sector, causing significant losses for institutions, leading to a decline in consumer spending and demand for goods and services. On the other hand, sectors such as healthcare and technology did relatively well [36]. In addition, the 2008 global financial crisis impacted continents differently, with Europe, Asia and Latin America experiencing varying degrees of economic slowdown and challenges. During the 2020 financial crisis, initiated by the COVID-19 pandemic, the healthcare sector was the most directly affected, due to increased demand, while travel and tourism suffered from restrictions. As the pandemic continued to spread, other sectors such as retail, airline industry and manufacturing industries were impacted, facing declines in employment and demand [37]. The 2020 financial crisis had varying effects on different continents, unfolding over different time periods. The epidemic started in Asia, then it hit Europe and finally it spread to the Americas.

To account for these features, we add undirected 'sectors' and 'continents' layers in which companies are connected if they are in the same sector or continent, respectively. Hence, each connected component in the sector and continents layers is a complete network.

Global layer. In addition to the 'fundamentals-based contagion' embedded in the above network layers, we allow for 'pure contagion', whereby crises may spread due to global effects not explicitly accounted for so far.

(b) Contagion model

We employ a discrete-time SIR epidemic model defined on the network of n companies to simulate financial crisis propagation. At each time step, a company is either susceptible (S), infected (I) or recovered (R). Let the integer-valued functions S, I and R represent the number of companies that are in the states S, I and R, respectively, at time t.

The process starts at day $t=0$, with $m \geq 1$ initially infected companies ($I(0)=m$) and the remainder being susceptible ($S(0)=n-m$, $R(0)=0$). Then at each day $t=1,2,3,\ldots$, an infected company i infects each susceptible neighbour j on layer α independently with probability $w_{i,j}^{[\alpha]}$, after which each infected company i recovers independently with probability p. Once recovered, a company cannot be reinfected again. Infection or recovery of a node occurs simultaneously on all layers. The process continues until there are no more infected companies.

We model the transmission probabilities per edge (i,j) in each layer $\alpha \in \{1,2,3,4\}$, where the values $\alpha=1,2,3,4$ correspond to the PMFG, continents, sectors and global layers, respectively, as

$$w_{i,j}^{[\alpha]} = \begin{cases} \hat{\chi}_{i,j}^{U} \times \beta_1, & \alpha = 1, \\ \beta_\alpha, & \alpha \in \{2,3,4\}, \end{cases} \tag{3.3}$$

where $\hat{\chi}_{i,j}^{U}$ is defined in (3.2) and β_α for $1 \leq \alpha \leq 4$ are parameters to be estimated (see §b(ii)). The definition of 'infection' in a financial context is provided in §a.

4. Application to financial crises

In this section, we fit the model to the 2008 and the 2020 financial crises. In §a, we define what is meant for a company in the dataset to be 'financially infected'. We then build and compare two different networks representing the financial dependency between the companies in the periods prior to the 2008 and 2020 financial crises. We study how the model can be used to predict future infections in each case, using recent infection data. We finally assess the importance for predictive accuracy of each layer within our network.

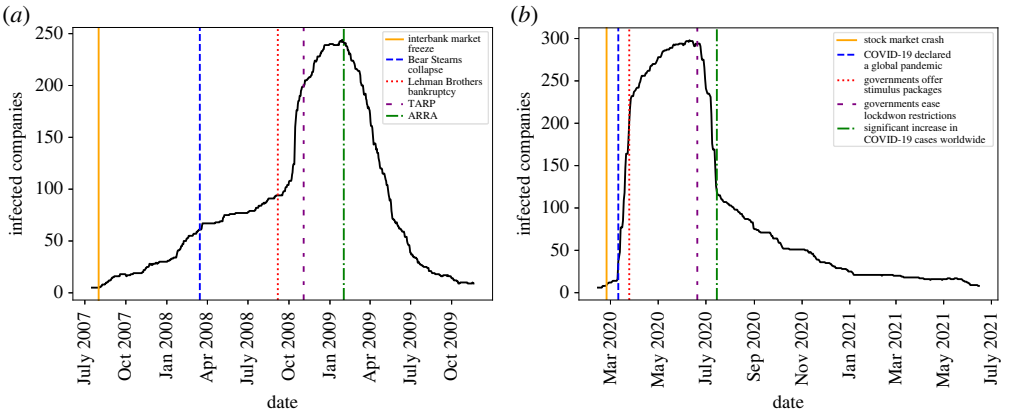

Figure 1. The total number of infected companies within our dataset, as defined in §a, during (*a*) the 2008 financial crisis and (*b*) the 2020 financial crisis. The vertical lines show the dates of significant events during each crisis.

(a) Infection

We define a company in the dataset to be infected whenever the volatility of its stock returns over a given period exceeds a predetermined threshold (meaning that the company's stock price is unstable) and its average stock return for the same period is negative.

The volatility for a time horizon $T > 1$ of company i at day t is defined as the standard deviation of the stock returns in the prior T trading days and is calculated as follows:

$$V_{i,t} = \sqrt{\frac{1}{T}\sum_{j=t-T}^{t-1}(x_{i,j} - \mu_{i,t})^2}, \tag{4.1}$$

where $\mu_{i,t}$ is the mean stock return over the same period and $x_{i,j}$ is defined in (2.1). Hence, company i is defined to be infected at day t whenever $V_{i,t} \geq \sigma_i$ and $\mu_{i,t} < 0$. In the following analysis, we use $T = 21$ trading days (one trading month) and the threshold σ_i to be the 90% quantile of the (empirical) volatility distribution for each company.

Using a rolling window of historical returns over the past 21 days is common in risk analysis [38–40] and suitable for estimating volatility for daily data because it strikes a balance between capturing recent changes in volatility and incorporating sufficient historical data to generate a stable estimate. This balance is especially important given our focus of identifying 'infection': longer periods could include stock price fluctuations whose effect on the market has passed, while short periods are likely to be sensitive to noise. Our choice of σ_i is determined by the 90% quantile of the (empirical) volatility distribution; however, it should be acknowledged that in practice, determining this threshold at a specific time without knowledge of future volatility values may not be feasible. Therefore, the quantile threshold is primarily used as a benchmarking tool to compare and analyse volatility levels across companies in a historical context.

Once we have determined at which day each company has been infected for those that become infected, we count the number of infected companies per day. Figure 1*a*,*b* illustrates the number of infected companies in the 2008 and 2020 crises, respectively, along with significant events that occurred during these periods. It can be seen that in the 2008 crisis after the Lehman Brothers bankruptcy in September 2008 (red vertical line on figure 1*a*), there is a substantial increase in the number of infected companies. Subsequently, after the Troubled Asset Relief Program (TARP) was implemented in October 2008, the rate at which the companies become infected decreases (purple vertical line on figure 1*a*) and after the American Recovery and Reinvestment Act (ARRA) was signed into law in February 2009, the companies start recovering (green vertical line on figure 1*a*). In the 2020 crisis, shortly after the WHO (World Health Organization) declared a global health

emergency in March 2020 (blue vertical line on figure 1*b*), accompanied by national lockdown measures in many countries,[2] the number of infected companies increases sharply in a short time period. When the US and UK governments started offering stimulus packages,[3] the rate at which the infections spread declined (around the red vertical line on figure 1*b*). Finally, in most of the countries, the lockdown restrictions were eased between June and July 2020 (the period around the purple vertical line on figure 1*b*), leading to recoveries. However, a month later (green vertical line on figure 1*b*), COVID-19 cases started increasing worldwide.[4] In summary, the analysis of figure 1*a*,*b* reveals the impact of these events on the spread and recovery of infected companies during the 2008 and 2020 crises, and indicates the suitability of our empirical definition of 'infection'.

(b) Predicting future infections

It is of key importance to be able to predict future infections, given past data, for risk prevention and mitigation purposes. This in turn is crucial for ensuring the stability and health of the global financial system as a way to protect investors and sustain economic growth. Here, we examine the accuracy of our model to estimate the number of infected companies in the future, given data from the past n days of each crisis. We evaluate the accuracy of predictions in terms of the total number of infected companies, the number per sector and continent and identifying the specific companies most likely to be affected.

We construct and compare two distinct networks, one for the 2008 and one for the 2020 financial crises, that represent the dependence structure preceding each crisis, as described in §b(i). These networks are then used to simulate future infections during the corresponding crises, and we analyse the empirical results in §b(ii), (b)(iii) and the electronic supplementary material.

(i) The 2008 and 2020 financial networks

The first network is constructed using all available data before the 2008 financial crisis, which includes the data from 17 January 2002 to 30 June 2007. The second network employs all available data before the 2020 financial crisis, i.e. from 17 January 2002 to 29 February 2020.

We now compare the community structure of the two PMFG networks. For each of the networks, we divide the nodes into communities by maximizing the modularity [41] of the network via the Louvain algorithm [42]. Then, for the two sets of communities, we estimate the similarity between them using the adjusted mutual information (AMI) score [43]. The AMI takes a value of 1 when the two partitions are identical (perfectly matched), while random partitions, having an expected AMI around 0 on average, can occasionally yield negative values (see electronic supplementary material). The AMI score between the clusterings of the two networks is 0.2568, suggesting that the community structures of the two graphs are substantially different.

We then perform a clique analysis by adopting the n-clique algorithm of [44] to analyse the community structures. A clique in a graph G is a complete subgraph of G. A clique, in other words, is a subset of a network in which the nodes are more intensively linked to one another than to

[2]National emergency was declared in the USA on 13 March 2020; the UK went into lockdown on 23 March 2020; a national lockdown in Italy was imposed on 9 March 2020; nationwide lockdown in France started on 17 March 2020; from 13 March 2020, German states mandated school and kindergarten closures and travel restrictions were put in place in Austria, Denmark, France, Luxembourg and Switzerland; Japan officially declared the COVID-19 outbreak as a national emergency on 19 March 2020.

[3]The Main Street Lending Program (9 April 2020), Primary Market Corporate Credit Facility (23 March 2020), CARES Act (27 March 2020) and Paycheck Protection Program Liquidity Facility (9 April 2020) were launched in the USA; the Coronavirus Job Retention Scheme (1 March 2020), Self-Employment Income Support Scheme (26 March 2020) and Coronavirus (Large) Business Interruption Loan Scheme (23 March 2020) were launched in the UK.

[4]The US confirms more than 50 000 new COVID-19 cases in one day for the first time, the Australian city of Melbourne goes back into lockdown for six weeks after a second outbreak, Florida reports a record 11 458 daily COVID-19 cases, Texas records more than 10 000 daily cases of COVID-19 for the first time, India becomes the third country to record one million cases of COVID-19, the WHO says the Middle East is at a 'critical threshold' with COVID-19 cases over one million.

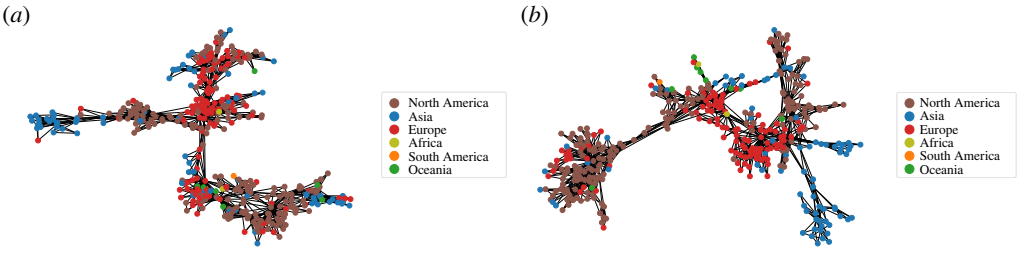

Figure 2. PMFG (tail dependence network) layers in the (*a*) 2008 and (*b*) 2020 financial networks, where the companies are coloured by continent. The two networks are constructed by the procedure described in §b(i) using stock price data for the periods from 17 January 2002 to 30 June 2007 and from 17 January 2002 to 29 February 2020, respectively.

Table 1. Clique analysis of the PMFG networks showing the cliques structure based on the sector or continent in which each company is based.

		continents		sectors	
clique type		2008	2020	2008	2020
3-cliques	total number of 3-cliques	25	4	25	4
	all nodes in same continent/sector	14	4	4	0
	two nodes in same continent/sector	11	0	10	2
	all nodes in different continent/sector	0	0	11	2
4-cliques	total number of 4-cliques	372	392	372	392
	all nodes in same continent/sector	174	252	24	58
	three nodes in same continent/sector	133	93	53	91
	two nodes in same continent/sector	63	47	194	162
	all nodes in different continent/sector	2	0	101	81

other members of the network. The maximal clique in the PMFG layer consists of four nodes, and is also called a 4-clique. By detecting cliques, we can uncover natural clusters or communities of companies that have strong connections or similarities. Table 1 shows the structure of the different 3- and 4-cliques in the two PMFGs based on companies' continents and sectors, respectively. The analysis shows that in both networks communities based on continents are more likely to form than communities based on sectors. In addition, the high number of 3- and 4-cliques in which all the companies are in the same continent indicates a strong tendency for continent-based communities. Figure 2*a*,*b* illustrates the PMFG networks for the 2008 and 2020 financial crises, respectively, where the companies are coloured by continent. On both figures, it can be seen that companies in the same continents tend to form clusters, indicating that the local level transmission is more likely to happen between companies in the same continent. The analysis of these data also suggests that communities based on sectors are more likely to form in the 2020 PMFG than in the 2008 PMFG, due to the higher occurrence of 4-cliques with all nodes in the same sector or three of the nodes in the same sector.

(ii) Prediction of the number of infected companies

We now evaluate the model's accuracy in predicting the number of companies that will be infected or recovered in the future k crisis days, based on the infection data from the past n days, utilizing a 'sliding window' technique. Firstly, we fit the model to the initial data window (data window 1), comprising data from day 1 to n, obtaining maximum-likelihood estimates $\hat{\beta}_i$ of the

layer transition probabilities β_i for $1 \leq i \leq 4$ and $\hat{p}$ for the recovery rate p (refer to the electronic supplementary material for further details). Next, we simulate $N = 10\,000$ realizations of the estimated SIR model for the upcoming k days (from day $n+1$ to day $n+k$, denoted prediction window 1), with the initial data being that from day n. After each simulation, we record the total number of infected companies, the number of newly infected companies, and the number of newly recovered companies, and calculate the mean of all simulations as the prediction. We then 'slide the window' forward by one day and refit the model to the period from day 2 to day $n+1$ of the crisis (data window 2), re-estimating $\hat{\beta}_i$ for $1 \leq i \leq 4$ and $\hat{p}$ for the new window. We repeat the above steps for each subsequent data window, with the final prediction window covering the period from day $L-k$ to L, where L is the length of the crisis in days.

Figure 3 displays the model predictions (coloured lines) alongside the observed infections (black lines) at selected time points. Predictions are computed from the mean of all $N = 10\,000$ simulations for the future $k \in \{10, 30\}$ days of each crisis, given infection data on the previous $n \in \{1, 30\}$ days. In both figures, it can be seen that fitting the model to the previous $n = 1$ crisis days gives the largest error between the actual and the predicted total number of infected individuals after k days for both values of k. For choices of $n > 1$, with greater prediction accuracy, we nevertheless observe large errors at those time points where significant changes in infection or recovery occur. This is natural since predictions are based on data prior to these change-points; it is important to note, however, that such events are often due to extrinsic factors, such as government intervention that could in principle be accommodated within the model. For example, substantial errors are observed in recovery prediction during periods associated with ARRA (2008; cf. figures 1*a* and 3) and stimulus packages and lockdown restrictions (2020; cf. figures 1*b* and 3). For all other periods in both crises, we obtain good prediction accuracy for suitable choices of k and n, as confirmed by further analysis. For completeness, the electronic supplementary material presents additional predictions for $k = 20$ in each crisis. The most precise forecasts were observed with $k = 10$, while the least precise predictions were observed with $k = 30$.

In order to compare the predicted and actual numbers of newly infected, total infected and newly recovered companies for each sliding window i (where $1 \leq i \leq L-k-n$), we calculate the absolute difference between the predicted and actual values in each simulation and then we take the average. To ensure accurate evaluation of the model's performance in predicting the number of newly infected (recovered, respectively) companies, we focus exclusively on suitable time periods. Specifically, we consider the period encompassing newly infected (recovered) companies, which corresponds to the time before (after) day 600 during the 2008 financial crisis, and before (after, respectively) day 120 in the case of the 2020 financial crisis. The results are shown in figure 4, which displays the distribution of the absolute difference between the predicted and actual number of total infected (row 1, *a*,*b*), newly infected (row 2, *c*,*d*) and newly recovered (row 3, *e*,*f*) companies when the model is fitted to the 2008 financial crisis (*a*,*c*,*e*) or the 2020 financial crisis (*b*,*d*,*f*). The white dots connected by white lines indicate the mean absolute difference over all sliding windows. For brevity, we present results for a prediction horizon of $k = 30$ days; those for $k = 10$ and $k = 20$ can be found in the electronic supplementary material and show that the trends in mean accuracy are similar for all choices of k.

The results indicate that the optimal window size for predicting future infections varies depending on the crisis being analysed. Specifically, for the 2008 financial crisis, the optimal window size is $n = 10$ for predicting both the future total number of infected and number of newly infected companies after k days, while for the 2020 financial crisis, the optimal window size is $n = 3$ for predicting the future total number of infected companies, and $n = 10$ for predicting the future number of newly infected companies. By contrast, a window size of $n = 1$ day for the 2008 crisis and a window size of $n = 30$ days for the 2020 crisis result in the worst predictions. Interestingly, when predicting the number of newly recovered companies in the future k days, for both crises, the worst predictions are obtained when the window size is the largest, i.e. $n = 30$, while the best predictions are obtained when the window size is the smallest, i.e. $n = 1$; we note, however, that the variation with n is not large.

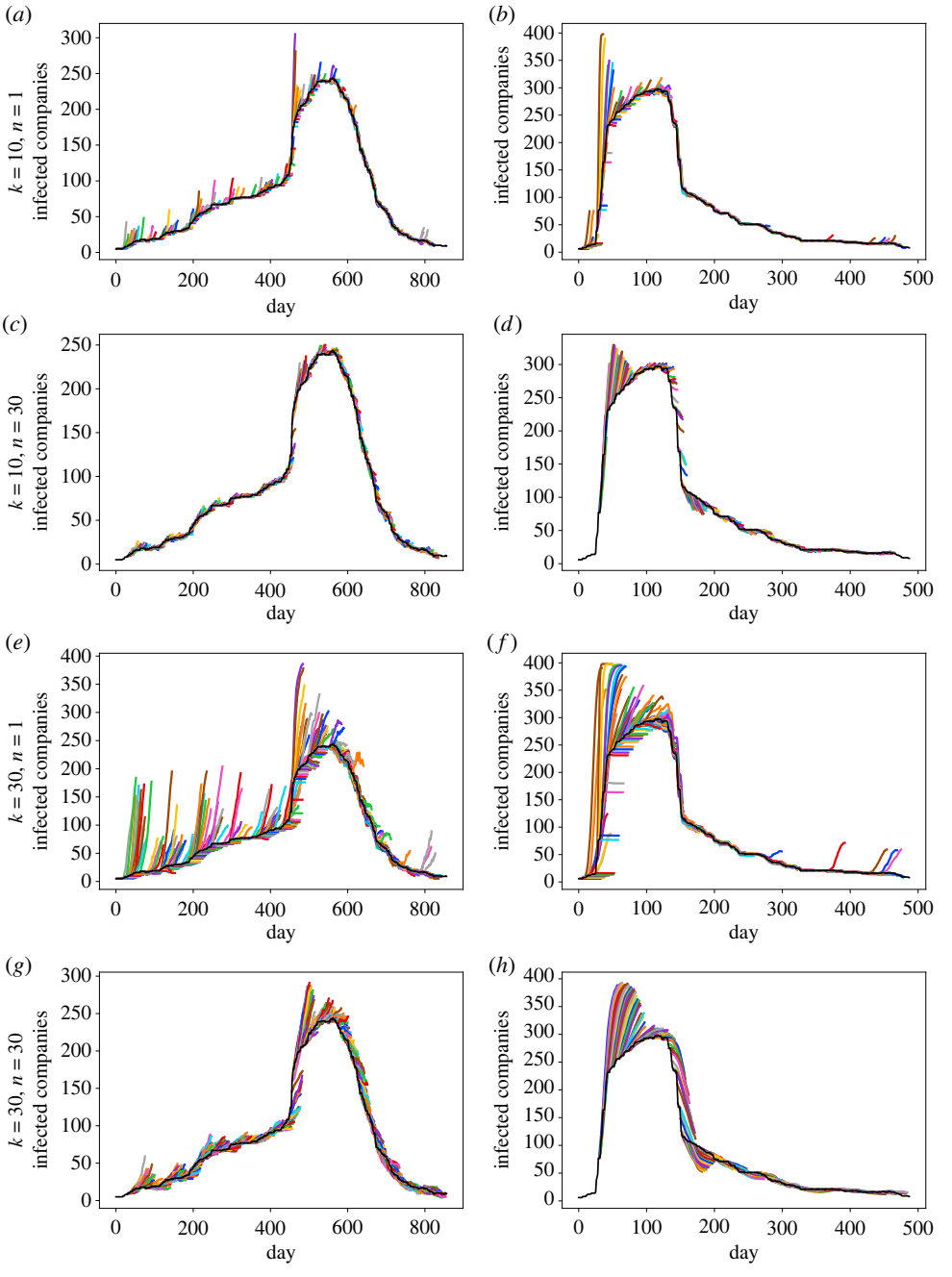

Figure 3. The curves (in colours) show the predicted mean total number of infected companies for each sliding window for the next k days over $N = 10\,000$ simulations, fitting the model to the previous n crisis days for the 2008 financial crisis (*a*,*c*,*e*,*g*) and 2020 financial crisis (*b*,*d*,*f*,*h*), respectively. The black lines show the observed number of infected companies as determined in §a, providing a reference for comparison with the model predictions.

(iii) Geographic- and sector-specific prediction

In this section, we investigate the model's ability to predict the geographical location and economic sector of the infected companies in the next k days, based on the previous n days' infection data. Rather than counting the number of infected companies, for each simulation, we

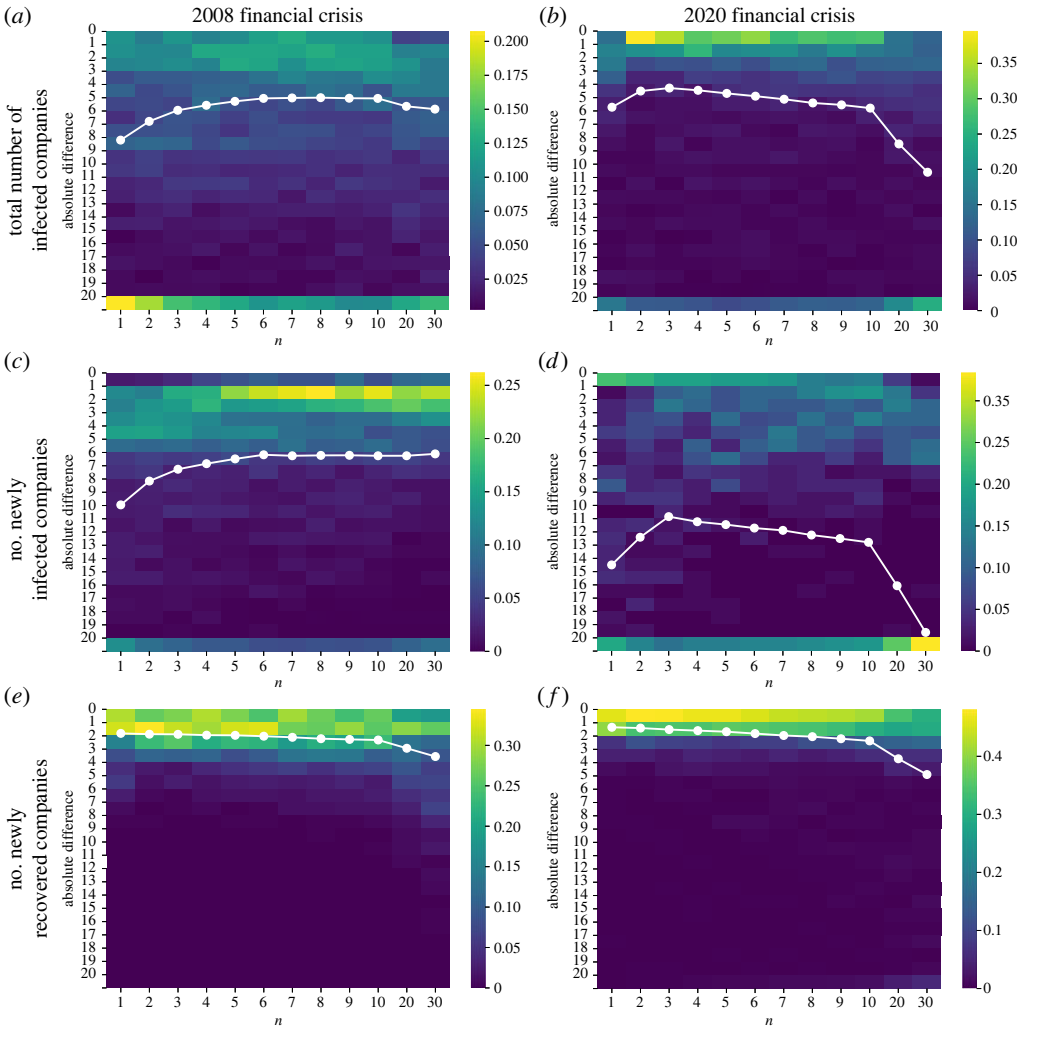

Figure 4. Heatmaps showing the distribution of the absolute difference between predicted and actual total number of infected companies (*a*,*b*), number of newly infected companies (*c*,*d*) and number of newly recovered companies (*e*,*f*) for the 2008 (*a*,*c*,*e*) and the 2020 (*b*,*d*,*f*) financial crises, using the infections data from the previous n days and at a prediction horizon of $k = 30$ days. The white dots indicate the mean absolute difference over all sliding window predictions.

construct a multiset (i.e. a set allowing for multiple instances of each of its elements) that includes the continents or sectors corresponding to the predicted infections in that simulation. We then compare each multiset to the observed continents or sectors multiset using the Sørensen–Dice similarity coefficient for multisets, defined as

$$D(A, B) = \frac{2|A \cap B|}{|A| + |B|}. \tag{4.2}$$

Here, A and B are multisets, not both empty, $|A|$ and $|B|$ denote the number of elements in A and B, respectively, and if an element appears in both A and B, it is included in the intersection $A \cap B$ with its minimal number of occurrences observed in A and B. The Sørensen–Dice coefficient takes values $D \in [0, 1]$ with $D = 1$ indicating identical multisets, and $D = 0$ complete dissimilarity.

By calculating the mean Sørensen–Dice coefficient from all simulations we obtain a measure of performance that reflects the overall effectiveness of the method for each prediction. Figure 5

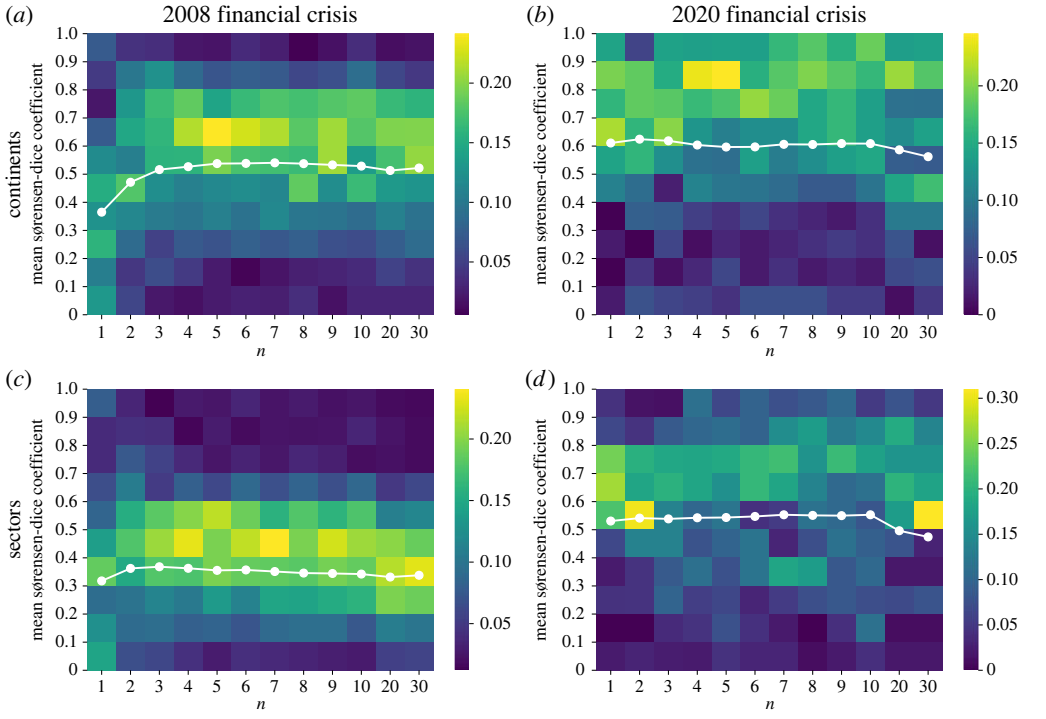

Figure 5. Heatmap for the distribution of the mean Sørensen–Dice coefficient between predicted and actual continents (*a*,*b*) and sectors (*c*,*d*) of newly infected companies in the future $k = 30$ days for the 2008 (*a*,*c*) and the 2020 (*b*,*d*) financial crises, using the infections data from the previous n days. The white dots indicate the mean over all sliding window predictions.

shows the distribution of Sørensen–Dice coefficients when comparing the predicted and actual continents and economic sectors of newly infected companies in the future $k = 30$ days, for different values of n, when the model is fitted to the 2008 (left column) and the 2020 (right column) financial crises. The results in each case indicate rather different optimal choices: for 2008, the least accurate predictions are obtained when using only the most recent data ($n = 1$), while for the 2020 crisis, smaller windows are in general preferable with $n = 30$ giving the worst predictions. However, apart from these worst cases, the dependence on n is not strong: for the 2008 data, very little variation in prediction accuracy as a function of n is observed, while for 2020, all choices $1 \leq n \leq 10$ give similar results. The results for the values of $k = 10$ and $k = 20$ are similar (see the electronic supplementary material), indicating that the model's ability to predict future infected continents and sectors remains stable for most values of the size n of the sliding window.

(c) Assessing the importance of the layers

In this section, we study the importance of each layer within our model, as defined by $w_{i,j}^{[\alpha]}$, $\alpha \in \{1, 2, 3, 4\}$ in (3.3). We compare the performance of six different networks: (i) the full network, comprising the PMFG, continents, sectors and global layer; (ii) the network without PMFG, i.e. $\alpha \in \{2, 3, 4\}$; (iii)–(v) duplex networks comprising the global layer and one other, i.e. $\alpha \in \{1, 4\}$, $\alpha \in \{2, 4\}$, or $\alpha \in \{3, 4\}$; (vi) the global layer only, i.e. $\alpha = 4$. We remark that the latter corresponds to the homogeneous mixing population case, which assumes that the probability of transmission is the same between all companies.

We first compare in figure 6 how the different multiplex networks perform, compared to the global layer only, in predicting the total number of infected companies in the future $k = 30$ days. For each of the six networks, we compute the average difference between the predicted and

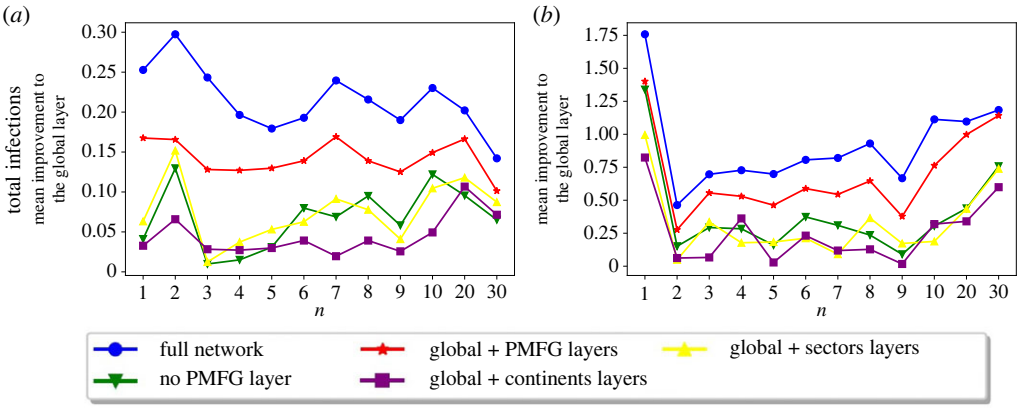

Figure 6. Comparison between the total number of infected companies at a prediction horizon of $k = 30$ days for the 2008 financial crisis (*a*) and the 2020 financial crisis (*b*), using the infections data from the previous n days, in comparison to the homogeneous mixing population model.

actual total number of infected companies after $k = 30$ days. The calculation is performed across different values of n. To assess the 'improvement' achieved by each network in comparison to the homogeneous mixing population model (which consists of only the global layer), we calculate the difference of the average differences between the predicted and actual total number of infected companies for each n between the global layer network and each of the other five networks. Our results demonstrate that for both financial crises the full network outperforms the other network structures and gives the highest improvement in predicting the total number of infected companies after $k = 30$ days, compared to the homogeneous mixing population model. Using the global layer alone (homogeneous mixing population model) gives the least accurate predictions since each other network produces positive improvements. Moreover, the second best results are achieved when using the network comprising the global and PMFG layers, indicating the importance of the PMFG layer. We remark that the six network models display similar accuracy of prediction in the case of new-recoveries, as is to be expected, since the recovery probability is independent of network structure (data not shown).

We then employ the same methodology as described in §b(iii) and compare accuracy of the results obtained from fitting all six models, according to the Sørensen–Dice coefficient. Figure 7 illustrates the comparison between the mean Sørensen–Dice coefficient over all sliding windows between predicted and actual continents (top row) and sectors (bottom row) of newly infected companies in the future $k = 30$ days for both the 2008 (*a*,*c*) and the 2020 (*b*,*d*) financial crises. The results demonstrate that for all studied values of k and n, the full model, containing all four layers, consistently yields the highest mean Sørensen–Dice coefficient for predicting both the continents and sectors in which newly infected companies will emerge. Conversely, employing only the global layer produces the lowest mean Sørensen–Dice coefficients across all combinations of n and k. Furthermore, the second-best results for all combinations of n and k are consistently observed when utilizing the network comprising only the global and PMFG layers. Adding each of the continents and sectors layers, in addition to the global layer, improves the quality of the predictions. This means that each of the layers within our model, and particularly the PMFG layer, includes information which improves the model's predictive power.

Section 5 of the electronic supplementary material offers a thorough examination of the model's predictive accuracy concerning the identification of specific companies likely to be affected in the upcoming k days. Here, we restrict attention to the performance for different values of n when $k = 30$, quantified by two metrics: Accuracy and F_1-score. In brief, the former describes the ratio of correct predictions to the total observations, while the latter is a commonly used measure that balances correct identification with minimising false positives. The results shown

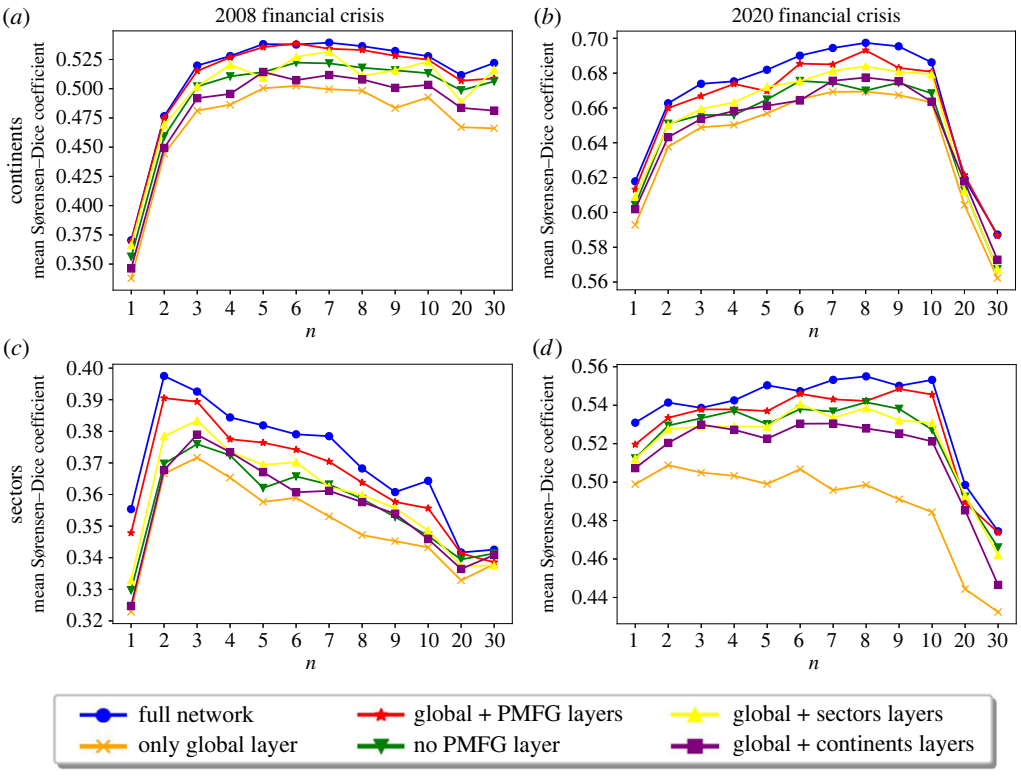

Figure 7. Comparison between the mean Sørensen–Dice coefficient, averaged over all prediction windows, between predicted and actual continents (*a*,*b*) and sectors (*c*,*d*) of newly infected companies in the future $k = 30$ days for the 2008 (*a*,*c*) and the 2020 (*b*,*d*) financial crises, using the infections data from the previous n days when using six different network models.

in figure 8 demonstrate that the full model consistently outperforms the homogeneous mixing population model in all scenarios. In particular, when examining the 2020 financial crisis, the full model's Accuracy surpasses that of the homogeneous mixing population model by nearly 10%.

Similar trends are observed when examining the mean F_1-score. Specifically, in the 2008 financial crisis the mean F_1-score of the full model improves that of the random model by around 5%, while an increase of nearly 10% is observed in the 2020 financial crisis. Furthermore, the network's performance is substantially improved when the model includes both the global and PMFG layers, resulting in the second highest scores. The results for $k = 10$ and $k = 20$, shown in the electronic supplementary material, are consistent with the ones for $k = 30$, which demonstrates the superiority of the full model over the homogeneous mixing population model, but also highlights the importance of incorporating the PMFG network for achieving more accurate predictions. The maximum F_1-score attained for the 2008 financial crisis is 0.08, whereas during the 2020 financial crisis, it reaches 0.24. These values are, of course, too low for practical prediction: our work constitutes a proof-of-concept rather than an immediately applicable method in this context. In addition, the observed improvements in Accuracy and F_1-score suggest that the additional information incorporated through the multilevel structure holds potential for enhancing predictive models in future research efforts.

5. Discussion and conclusion

This paper proposes a novel framework to analyse the spread of financial crises. We integrate stock price, geographical and economic sector data to provide a four-layer multiplex network on

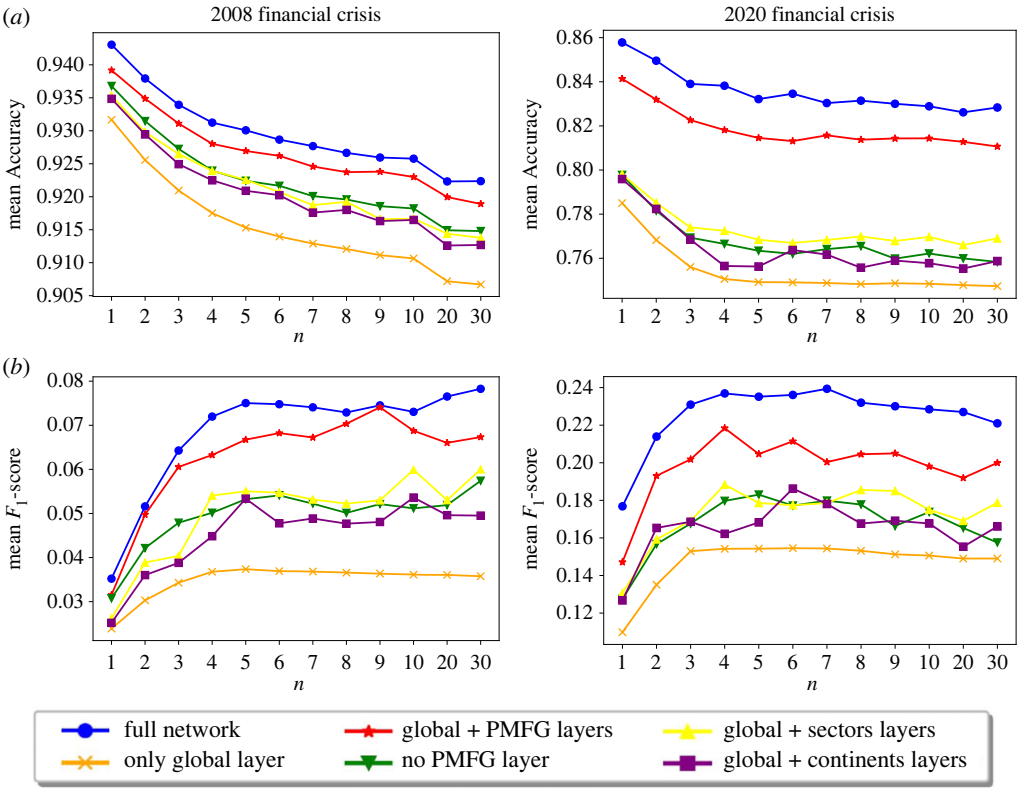

Figure 8. Comparison between the mean (*a*) Accuracy and (*b*) F_1-score in the future $k = 30$ days for the 2008 financial crisis (left column) and the 2020 financial crisis (right column), using the infections data from the previous n days to predict the set of individual infected companies when using six different network models. See the electronic supplementary material for more details.

which a discrete-time SIR model is simulated, so as to predict the spread of financial risk through interconnections between companies. Specifically, by fitting infection and recovery parameters on each layer of our network to historic stock data through a maximum-likelihood approach, we seek to predict future infection dynamics.

We investigate and evaluate the utility of our approach through application to two recent financial crises: the 2008 crisis, initiated by the subprime mortgage market and the 2020 crisis, associated with the COVID-19 pandemic. In each case, we examine the ability of our model to estimate dynamically future infection risk over a horizon of k days, given data from the prior n days. Using a range of accuracy measures, we analyse the dependence of prediction accuracy on k and n, in terms of total number of infections, as well as sector- and location-specificity. Thereby, we demonstrate that interactions among companies within and across sectors and continents in the financial network plays a substantial role in the spread of financial crises and their incorporation into the model improves the prediction of future outbreaks of financial distress. By comparison with a homogeneous mixing assumption in particular, we highlight the importance of understanding and accounting for the complex interdependencies between companies in financial systems for risk prediction.

While our model offers valuable insights into the spread of financial crises, it is essential to recognize its limitations and constraints. One significant limitation of our model is its reliance on historical stock price data, which only gives an incomplete view of the financial stability of a company. The accuracy and reliability of our predictions heavily depend on the availability and quality of the data, which may vary across different companies, sectors and regions. Moreover,

our model operates under several basic assumptions, such as the division into susceptible, infected, and recovered companies. While these assumptions simplify the complexity of financial contagion dynamics, they also impose constraints on the model's applicability and may not fully capture the nuances of real-world scenarios. Despite incorporating multiple layers representing stock prices, geographical locations, and economic sectors, our model overlooks other potentially important factors influencing financial contagion, such as macroeconomic indicators, regulatory policies, investor sentiment and systemic risk factors. We emphasize that the primary goal of this research is to present our novel framework combining extreme value theory, financial network construction and SIR modelling for the spread of financial risk in networks rather than to undertake comprehensive prediction. Despite its limitations, the incorporation of the multilevel network structure has shown potential in enhancing prediction power and capturing the interdependencies among companies driving financial contagion dynamics.

Overall, our results suggest that the proposed framework, which updates in real time as new data become available, is effective in predicting risk spread, this information potentially being useful in terms of risk prevention and mitigation. In addition, our results agree with the existing research which consistently shows the importance of including economic sector [45] and geographical location [46] information in predicting a company's future performance. However, [47] suggests that industry-level analysis may not always provide a quantity of information that results in substantial improvement of future profitability, therefore suggesting that efficiently incorporating further, more granular information might be difficult. In [48], the authors examine the macroeconomic consequences of firm- or industry-level shocks, and use a model that differentiates between supply side shocks and demand-side shocks. It would therefore be interesting to see how the results of our analysis change when transmission rates are set according to directed rather than undirected graphs. Moreover, since predictive accuracy suffers during periods of rapid change, natural future work includes dynamic updating of the multiplex connectivity structure and model infection parameters, though this is likely to result in a significant increase in computational complexity. We note, however, that such events are typically associated with extrinsic factors, such as government financial stimulus packages or lockdown periods. The incorporation of such additional information provides a route to suitable parameter updates to accommodate such change points.

Data accessibility. The datasets and codes used in the article are provided in electronic supplementary material [49].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. M.B.: conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing—original draft, writing—review and editing; F.B.: supervision, writing—review and editing; Y.v.G.: supervision, writing—review and editing; G.S.: supervision, writing—review and editing; R.D.O.: supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. M.B. acknowledges studentship support of EPSRC (grant reference no. EP/V520020/1); and Russell Group Limited. G.S. acknowledges financial support from the French National Research Agency under the grant nos. ANR-19-CE40-0013 (ExtremReg project) and ANR-11-LABX-0020-01 (Centre Henri Lebesgue), as well as from the TSE-HEC ACPR Chair 'Regulation and systemic risks', an AXA Research Fund Award on 'Mitigating risk in the wake of the COVID-19 pandemic' and the Chair Stress Test, RISK Management and Financial Steering of the Foundation Ecole Polytechnique. Y.v.G. has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant no. 777826.

References

1. Demiris N, Kypraios T, Smith LV. 2014 On the epidemic of financial crises. *J. R. Stat. Soc. A* (*Stat. Soc.*) **177**, 697–723. (doi:10.1111/rssa.12044)

2. Gai P, Kapadia S. 2010 Contagion in financial networks. *Proc. R. Soc. A* **466**, 2401–2423. (doi:10.1098/rspa.2009.0410)
3. Barja A, Martínez A, Arenas A, Fleurquin P, Nin J, Ramasco JJ, Tomás E. 2019 Assessing the risk of default propagation in interconnected sectoral financial networks. *EPJ Data Sci.* **8**, 32. (doi:10.1140/epjds/s13688-019-0211-y)
4. Murphy A. 2008 An analysis of the financial crisis of 2008: causes and solutions. Oakland University, School of Business Administration, see http://ssrn.com/abstract=1295344.
5. Allen F, Gale D. 2000 Financial contagion. *J. Pol. Econ.* **108**, 1–33. (doi:10.1086/262109)
6. Abduraimova K. 2022 Contagion and tail risk in complex financial networks. *J. Bank. Financ.* **143**, 106560. (doi:10.1016/j.jbankfin.2022.106560)
7. In't Veld D, Van der Leij M, Hommes C. 2020 The formation of a core-periphery structure in heterogeneous financial networks. *J. Econ. Dyn. Control* **119**, 103972. (doi:10.1016/j.jedc.2020.103972)
8. Ledford AW, Tawn JA. 1997 Modelling dependence within joint tail regions. *J. R. Stat. Soc.* **59**, 475–499. (doi:10.1111/1467-9868.00080)
9. Schmidt R. 2005 Tail dependence. In *Statistical tools for finance and insurance*, pp. 65–91. New York, NY: Springer.
10. Jusup M *et al.* 2022 Social physics. *Phys. Rep.* **948**, 1–148. (doi:10.1016/j.physrep.2021.10.005)
11. Lazebnik T, Shami L, Bunimovich-Mendrazitsky S. 2023 Intervention policy influence on the effect of epidemiological crisis on industry-level production through input–output networks. *Socio-Econ. Plann. Sci.* **87**, 101553. (doi:10.1016/j.seps.2023.101553)
12. Huaihu C, Jianming Z. 2012 Research on banking crisis contagion dynamics based on the complex network of system engineering. *Syst. Eng. Procedia* **5**, 156–161. (doi:10.1016/j.sepro.2012.04.025)
13. Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. 2014 Multilayer networks. *J. Complex Netw.* **2**, 203–271. (doi:10.1093/comnet/cnu016)
14. Aleta A, Moreno Y. 2019 Multilayer networks in a nutshell. *Annu. Rev. Condens. Matter Phys.* **10**, 45–62. (doi:10.1146/annurev-conmatphys-031218-013259)
15. del Rio-Chanona RM, Korniyenko Y, Patnam M, Porter MA. 2020 The multiplex nature of global financial contagions. *Appl. Netw. Sci.* **5**, 1–23. (doi:10.1007/s41109-020-00301-2)
16. Cao J, Wen F, Stanley HE, Wang X. 2021 Multilayer financial networks and systemic importance: evidence from China. *Int. Rev. Financ. Anal.* **78**, 101882. (doi:10.1016/j.irfa.2021.101882)
17. De Domenico M. 2023 More is different in real-world multilayer networks. *Nat. Phys.* **19**, 1247–1262. (doi:10.1038/s41567-023-02132-1)
18. Tumminello M, Aste T, Di Matteo T, Mantegna RN. 2005 A tool for filtering information in complex systems. *Proc. Natl Acad. Sci. USA* **102**, 10 421–10 426. (doi:10.1073/pnas.0500298102)
19. Glick R, Rose AK. 1999 Contagion and trade: why are currency crises regional? *J. Int. Money Fin.* **18**, 603–617. (doi:10.1016/S0261-5606(99)00023-6)
20. Tiwari AK, Abakah EJA, Yaya OS, Appiah KO. 2023 Tail risk dependence, co-movement and predictability between green bond and green stocks. *Appl. Econ.* **55**, 201–222. (doi:10.1080/00036846.2022.2085869)
21. Ahnert T, Georg CP. 2018 Information contagion and systemic risk. *J. Financ. Stability* **35**, 159–171. (doi:10.1016/j.jfs.2017.05.009)
22. Cai JJ, Einmahl JHJ, de Haan L, Zhou C. 2015 Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *J. R. Stat. Soc. B* **77**, 417–442. (doi:10.1111/rssb.12069)
23. Sklar M. 1959 Fonctions de répartition à *n* dimensions et leurs marges. *Publ. de l'Institut Statistique de l'Université de Paris* **8**, 229–231.
24. Coles S, Heffernan J, Tawn J. 1999 Dependence measures for extreme value analyses. *Extremes* **2**, 339–365. (doi:10.1023/A:1009963131610)
25. Le TH, Do HX, Nguyen DK, Sensoy A. 2021 Covid-19 pandemic and tail-dependency networks of financial assets. *Financ. Res. Lett.* **38**, 101800. (doi:10.1016/j.frl.2020.101800)
26. Chen Q, Giles DE, Feng H. 2012 The extreme-value dependence between the Chinese and other international stock markets. *Appl. Financ. Econ.* **22**, 1147–1160. (doi:10.1080/09603107.2011.631890)
27. Singh AK, Allen DE, Powell RJ. 2017 Tail dependence analysis of stock markets using extreme value theory. *Appl. Econ.* **49**, 4588–4599. (doi:10.1080/00036846.2017.1287858)
28. Nishizeki T, Chiba N. 1988 *Planar graphs: theory and algorithms*. Amsterdam, The Netherlands: Elsevier.

29. Musmeci N, Aste T, Di Matteo T. 2015 Relation between financial market structure and the real economy: comparison between clustering methods. *PLoS ONE* **10**, e0116201. (doi:10.1371/journal.pone.0116201)
30. Pozzi F, Di Matteo T, Aste T. 2013 Spread of risk across financial markets: better to invest in the peripheries. *Sci. Rep.* **3**, 1–7. (doi:10.1038/srep01665)
31. Fiedor P. 2014 Networks in financial markets based on the mutual information rate. *Phys. Rev. E* **89**, 052801. (doi:10.1103/PhysRevE.89.052801)
32. Song WM, Di Matteo T, Aste T. 2012 Hierarchical information clustering by means of topologically embedded graphs. *PLoS ONE* **7**, e31929. (doi:10.1371/journal.pone.0031929)
33. Mantegna RN. 1999 Hierarchical structure in financial markets. *Eur. Phys. J. B* **11**, 193–197. (doi:10.1007/s100510050929)
34. Bonanno G, Caldarelli G, Lillo F, Mantegna RN. 2003 Topology of correlation-based minimal spanning trees in real and model markets. *Phys. Rev. E* **68**, 046130. (doi:10.1103/PhysRevE.68.046130)
35. Yan X, Jeub LG, Flammini A, Radicchi F, Fortunato S. 2018 Weight thresholding on complex networks. *Phys. Rev. E* **98**, 042304. (doi:10.1103/PhysRevE.98.042304)
36. Taylor JB. 2009 The financial crisis and the policy responses: an empirical analysis of what went wrong. Technical report National Bureau of Economic Research, Working Paper 14631.
37. Pak A, Adegboye OA, Adekunle AI, Rahman KM, McBryde ES, Eisen DP. 2020 Economic consequences of the COVID-19 outbreak: the need for epidemic preparedness. *Front. Public Health* **8**, 241. (doi:10.3389/fpubh.2020.00241)
38. Alexander C. 2008 *Market risk analysis, pricing, hedging and trading financial instruments*. Hoboken, NJ: John Wiley & Sons.
39. Schwert GW. 2002 Stock volatility in the new millennium: how wacky is Nasdaq? *J. Monet. Econ.* **49**, 3–26. (doi:10.1016/S0304-3932(01)00099-X)
40. Kristjanpoller W, Fadic A, Minutolo MC. 2014 Volatility forecast using hybrid neural network models. *Expert Syst. Appl.* **41**, 2437–2442. (doi:10.1016/j.eswa.2013.09.043)
41. Newman ME. 2006 Modularity and community structure in networks. *Proc. Natl Acad. Sci. USA* **103**, 8577–8582. (doi:10.1073/pnas.0601602103)
42. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP. 2010 Community structure in time-dependent, multiscale, and multiplex networks. *Science* **328**, 876–878. (doi:10.1126/science.1184819)
43. Xuan N, Julien V, Wales S, Bailey J. 2010 Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. *J. Mach. Learn. Res.* **11**, 2837–2854.
44. Palla G, Derényi I, Farkas I, Vicsek T. 2005 Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814–818. (doi:10.1038/nature03607)
45. Marfatia H. 2023 The financial market's ability to forecast economic growth: information from sectoral movements. *J. Econ. Stud.* **50**, 1467–1484. (doi:10.1108/JES-08-2022-0466)
46. Chang X, Li J. 2019 Business performance prediction in location-based social commerce. *Expert Syst. Appl.* **126**, 112–123. (doi:10.1016/j.eswa.2019.01.086)
47. Fairfield PM, Ramnath S, Yohn TL. 2009 Does industry-level analysis improve profitability and growth forecasts? *J. Account. Res.* **47**, 147–178. (doi:10.1111/j.1475-679X.2008.00313.x)
48. Acemoglu D, Akcigit U, Kerr W. 2016 Networks and the macroeconomy: an empirical exploration. *NBER Macroecon. Annu.* **30**, 273–335. (doi:10.1086/685961)
49. Bozhidarova M, Ball F, van Gennip Y, O'Dea RD, Stupfler G. 2024 Describing financial crisis propagation through epidemic modelling on multiplex networks. Figshare. (doi:10.6084/m9.figshare.c.7131861)