Mortgage Loan Flow Networks and Financial Norms

Richard Stanton
Haas School of Business, U. C. Berkeley

Johan Walden
Haas School of Business, U. C. Berkeley

Nancy Wallace
Haas School of Business, U. C. Berkeley

We develop a theoretical model of a network of intermediaries whose optimal behavior is jointly determined, leading to heterogeneous financial norms and systemic vulnerabilities. We apply the model to the network of U.S. mortgage intermediaries from 2005 to 2007, using a data set containing all private-label, fixed-rate mortgages, with loan flows defining links. Default risk was closely related to network position, evolving predictably among linked nodes, and loan quality estimated from the model was related to independent quality measures, altogether pointing to the vital importance of network effects in this market. (*JEL* G21, G01, L14)

Received April 20, 2016; editorial decision July 11, 2017 by Editor Stijn Van Nieuwerburgh.

In a recent paper, Stanton, Walden, and Wallace (2014) establish the existence of strong empirical network effects in the U.S. residential mortgage market, with clearly demarcated segments of connected firms subject to differing amounts of default risk. A recent theoretical literature also highlights the potential

---

A previous version of this paper was circulated under the title "Securitization Networks and Endogenous Financial Norms in U.S. Mortgage Markets." We are grateful for financial support from the Fisher Center for Real Estate and Urban Economics. We thank Stijn Van Nieuwerburgh (the editor); two anonymous referees; and seminar participants at the 2015 meeting of the Western Finance Association, the 2015 NBER Summer Institute conference on Risks of Financial Institutions, the Consortium for Systemic Risk Analytics (MIT), the Institute for Pure and Applied Mathematics (IPAM, UCLA), the Fall 2014 Journal of Investment Management Conference, FIRS 2016, the 2016 ITAM Finance, the 2016 meeting of the European Finance Association, BI Norwegian Business School, Carnegie Mellon, the Federal Reserve Bank of San Francisco, the Federal Reserve Board, London Business School, London School of Economics, NYU Stern, Swedish Institute for Financial Research (SIFR), University of Alberta, University of Oregon, University of Lausanne, University of Wisconsin-Madison, the Norwegian School of Economics, and Stanford GSB. We are grateful to Daron Acemoglu, Maryam Farboodi, Xavier Gabaix, Michael Gofman, Burton Hollifield, George Papanicolaou, and Roger Stein for helpful comments and suggestions. Walden thanks the 2015 IPAM program on financial mathematics for hosting his visit, during which part of this research was carried out, and the Swedish House of Finance for valuable support. Send correspondence to Richard Stanton, Haas School of Business, U. C. Berkeley, 545 Student Services Building, Berkeley, CA 94720-1900; telephone: (510) 642-7382. E-mail: stanton@haas.berkeley.edu.

doi:10.1093/rfs/hhx097

importance of network connections between intermediaries and institutions in explaining financial market risk.[1] These studies show that financial networks may create resilience against shocks in a market via diversification and insurance, but the network interconnections also expose market participants to additional sources of risk. Network structure can thus be a crucial determinant of the riskiness of a financial market.

These theoretical studies of networks and risk in financial markets typically focus on ex post effects: how the network redistributes risk between participants, and the consequences for the system's solvency and liquidity after a shock. Particularly in the mortgage market, ex ante effects also should be important: the presence and structure of a financial network should affect—and be affected by—the quality choices (e.g., the resources devoted to screening and monitoring loans) and other actions of individual intermediaries and financial institutions, even before shocks are realized. The focus of our study is to build a model that can both explain the empirical findings of Stanton, Walden, and Wallace (2014) and help us to understand the equilibrium interaction between network structure, the quality choices made by market participants, and the market's riskiness.

We introduce a model with multiple agents—representing financial intermediaries—who are connected in a network. Network structure in our model, in addition to determining the ex post riskiness of the financial system, also affects—and is affected by—what we call the *financial norms* in the network, inspired by the literature on influence and endogenous evolution of opinions and social norms in networks (see, e.g., Friedkin and Johnsen 1999; Jackson and López-Pintado 2013; López-Pintado 2012). Financial norms in our model are defined as the quality and risk decisions agents make, which are, in turn, influenced by the actions of other agents in the network.

Our model is parsimonious, in that the strategic action space of agents and the contract space are limited. Links in the network represent risk-sharing agreements, like in Allen, Babus, and Carletti (2012). Agents may add and sever links, in line with the concept of pairwise stability in games on networks (see Jackson and Wolinsky 1996), and also have the binary decision of whether to invest in a costly screening technology that improves the quality of the projects they undertake.

The equilibrium concept used is subgame perfect Nash. In an equilibrium network, each agent optimally chooses whether to accept the network structure, as well as whether to invest in the screening technology, given (correct) beliefs about all other agents' actions and risk. Shocks are then realized and distributed among market participants according to a clearing mechanism

[1] See, for example, Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015), Allen, Babus, and Carletti (2012), Allen and Gale (2000), Babus (2016), Cabrales, Gottardi, and Vega-Redondo (2014), Chang and Zhang (2015), Di Maggio and Tahbaz-Salehi (2014), Elliott, Golub, and Jackson (2014), and Glasserman and Young (2015). A growing literature deals with trading networks in OTC markets, for example, Babus and Kondor (2013), Gofman (2011), and Zhong (2014), but risk exposure is not the focus of this literature.

similar to that defined in Eisenberg and Noe (2001). Like in Elliott, Golub, and Jackson (2014), we assume that costs are associated with the insolvency of an intermediary, potentially creating propagation of shocks through the clearing mechanism, and thereby making the market systemically vulnerable. The model is simple enough to allow us to analyze the equilibrium properties of large-scale networks computationally.

Our model has several general implications. First, network structure is related to financial norms. Given that an agent's actions influence and are influenced by the actions of those with whom the agent interacts, this result is natural and intuitive. Importantly, an agent's actions affect not only his direct counterparties but also those who are indirectly connected through a sequence of links. As a consequence, a rich relationship between equilibrium financial norms and network structure suggests a further relationship, between the network and the financial strength of the market, beyond the mechanical relationship generated by shock propagation.

Second, heterogeneous financial norms may coexist in the network in equilibrium. Thus, two intermediaries that are ex ante identical may be very different when their network positions are taken into account, not just in how they are affected by the other nodes in the network but also in their own actions. Empirically, this suggests that network structure is an important determinant not only of the aggregate properties of the economy but also of the actions and performance of individual intermediaries.

Third, proximity in the network is related to financial norms: nodes that are close tend to develop similar norms, just like in the literature on social norms in networks. This result suggests the possibility of decomposing the market's financial network into "good" and "bad" parts, and addressing vulnerabilities generated by the latter.

Fourth, the behavior of a significant majority of nodes can typically be analyzed in isolation, while a small proportion of nodes affect the whole network through their actions. Such *systemically pivotal* nodes are especially important, suggesting a "too pivotal to fail" characterization of the systemically most important intermediaries in the market, rather than a "too big to fail" focus.

Fifth, the financial norms of a node are related to the size and connectedness of its own connections. All else equal, nodes that are connected to other nodes that are larger and/or less connected tend to be of lower quality.

We analyze the mortgage origination and securitization network of financial intermediaries in the United States, using a data set containing all fixed-rate, private-label mortgages (i.e., mortgages not securitized by either Fannie Mae or Freddie Mac) originated and securitized between 2005 and 2007. We use loan flows to identify the network structure of this market and ex post default rates to measure performance, and use the model to estimate the evolution of risk and financial norms in the network.

We document a strong relationship between network position and performance, in line with the predictions of our model, which is present even after controlling for other observable characteristics like type and geographical position of the lender. We also find a positive relationship between predicted and actual out-of-sample default rates of the firms in the market. Finally, we compare estimated firm quality with labor intensity, which we argue is a proxy for how much firms invested in monitoring of loans, and again find an economically and statistically significant positive relationship.

1. The U.S. Residential Mortgage Market

The precrisis residential mortgage origination market comprised thousands of firms and subsidiaries, including commercial banks, savings banks, investment banks, savings and loan institutions (S&Ls), mortgage companies, real estate investment trusts (REITs), mortgage brokers, and credit unions. These had various roles in handling the loan flows from origination to ultimate securitization. Except for the closure of the Office of Thrift Supervision in 2011, the types of firms in the market are largely the same today.

A major driver of the subprime mortgage crisis was increased credit supply, as shown by Mian and Sufi (2009, 2011, 2014), who study heterogeneous loan performance at the ZIP-code level and show that performance was closely related to credit availability. Other authors (see, e.g., Bernanke 2007; Di Maggio and Kermani 2014; Kermani 2014; Rajan 2010) have argued that the cumulative effect of low interest rates over the decade leading up to the financial crisis lowered user costs and increased the demand for credit to purchase housing services. Others have argued that the rapid expansions in the precrisis mortgage market arose due to widely held beliefs concerning continued house price growth (see, e.g., Cheng, Raina, and Xiong 2014; Glaeser and Nathanson 2017; Shiller 2014) Alternative explanations have focused on how mortgage securitization led to the expansion of mortgage credit to risky or marginal borrowers (see, e.g., Demyanyk and Van Hemert 2011; Loutskina and Strahan 2009; Nadauld and Sherlund 2013). Palmer (2015) suggests that vintage effects were important in explaining heterogeneous loan performance: mortgages originated earlier had more time for house price appreciation in the booming market, which created a cushion against default. Our focus complements this literature, since as we shall see, precrisis network effects were also important in explaining heterogeneous loan performance in the securitized residential mortgage market.

Figure 1 presents the market structure for loan sales, decomposed into three levels. Loans flow from the mortgage originator of record to the aggregator of the loans, and then to the securitization shelf and to the holding company that owns it.[2] Figure 1 portrays two possible holding-company types. The left-hand side of the schematic, shown in yellow, represents bank or thrift holding

[2] When private-label issuers file a registration statement to register an issuance of a REMIC security, they typically use a "shelf registration," which is owned by a shelf holding company. The shelf registration includes a "core" or "base" prospectus, which outlines the parameters of the various types of REMIC

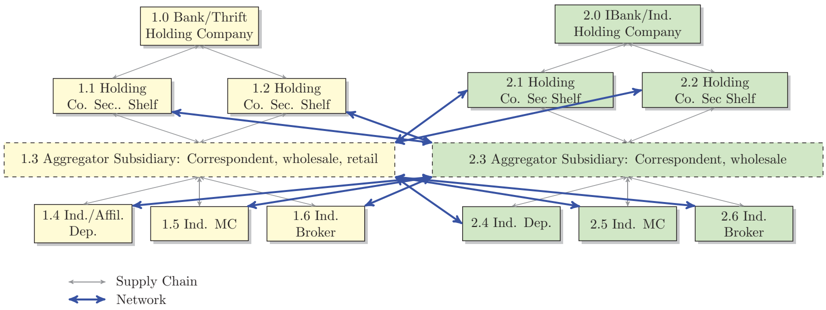

Figure 1
The residential mortgage market structure
Loan sales within affiliated firms (gray arrows) versus loan sales between unaffiliated firms (blue arrows).

company operations, while the right-hand side of the schematic, shown in green, represents investment bank or large independent mortgage company operations in the precrisis period.

The gray arrows represent loan sales between entities that are subsidiaries (they may or may not be fully consolidated) of the same holding company and the blue arrows represent loan sales between two unaffiliated firms and holding companies.[3] The graph of the gray arrows represents two trees. However, since loans in the market also flow between these entities, represented by the additional blue arrows in the figure, the market is really a network.[4] The network structure of the market is important, because it suggests complex interactions between intermediaries, potentially affecting their behavior and providing a channel through which shocks can spread.

Note that the arrows in Figure 1 are double-headed, representing bidirectional links in the network. In 2006, the preponderance of loan sales into the pools, typically organized as real estate mortgage investment conduits (REMICs), occurred within 60 days of the origination date of the loan due to the contractual structure of the wholesale lending mechanisms used to fund mortgage origination.[5] These contractual funding structures assign the cash

---

securities offerings that will be conducted in the future through the shelf registration. The rules governing shelf issuance are part of the Secondary Mortgage Market Enhancement Act (SMMEA) (see Simplification of Registration Procedures for Primary Securities Offerings, Release No. 33-6964, Oct. 22, 1992, and SEC Staff Report: Enhancing Disclosure in the Mortgage-Backed Securities Markets, January, 2003, http://www.sec.gov/news/studies/mortgagebacked.htm#secii).

3 For visual clarity, we do not show sales between entities within the investment bank or independent mortgage holding company (the green entity on the right-hand side of Figure 1) to the bank and thrift entities, but these sales would also exist. There also would be sales between bank/thrift entities and investment bank/mortgage company entities.

4 Technically, a network is a general graph of nodes and links, with no restriction with respect to the existence of cycles or connectivity, whereas a tree is a connected graph with no cycle.

5 The two most important of these funding mechanisms are (1) the master repurchase agreement, a form of repo, which received safe-harbor protections under the Bankruptcy Abuse Prevention and Consumer Protection

Table 1
Examples of network loan sales

Loan number	County	State	Lender of record	Aggregator	Shelf name	Holding company
1	Virginia Beach	VA	Accubanc	B of A	333-133985-08	Lehman Brothers
2	Queens	NY	Ameriquest	B of A	333-129480-11	Lehman Brothers
3	Baltimore	MD	B of A	B of A	333-131374-59	Bear Stearns
4	Hillsborough	FL	GMAC	B of A	333-129480-11	Lehman Brothers
5	Union	GA	TB&W Mtg	B of A	333-133985-74	Lehman Brothers
6	Ventura	CA	World Savings	B of A	333-129480-11	Lehman Brothers
7	Salt Lake	UT	Wells Fargo	B of A	333-132232-17	Bear Stearns

Examples of network loan sales for a sample of seven loans originated and securitized in 2006. The geographic location of the loans varied, as did the lender of record, the shelf registration of the loan, and the holding company that owned the shelf.

flows of the originated mortgages forward to each purchaser. However, the lender of record, or in some cases the aggregator, retains a contractual put-back option on the loan, which makes the risk structure of the loan flows bidirectional. In line with this observation, Stanton, Walden, and Wallace (2014) find that the mortgage market is well represented by a bidirectional network, and that the performance of an individual node is closely related to the performance of the node's neighbors (i.e., the nodes with which it shares a link) in the network. We therefore follow Stanton, Walden, and Wallace (2014) and use a bidirectional network representation.

Table 1 shows an example of actual loan flows for a sample of seven loans originated and securitized in 2006. Each of the seven loans was aggregated by the same subsidiary of Bank of America. As shown in the table, Bank of America (BofA) aggregated loans from independent mortgage companies (Accubanc, Ameriquest, GMAC, Taylor, and Bean & Whittaker Mortgage (TB&W Mtg)), from a S&L (World Savings), from a Bank of America branch, and from the subsidiary of another bank depository (Wells Fargo Bank). Bank of America then sold the mortgages to REMICs created within five different shelf-registration facilities. Three of these shelves were owned by Lehman Brothers and two by Bear Stearns. Thus, this Bank of America aggregator had a one-to-many relation both with lenders (the bottom level in Figure 1) and with holding companies (the top level in Figure 1), in line with the network description in Figure 1. We stress that since links in the mortgage network represent loan flows, a neighboring relationship between two nodes has no geographical interpretation.

Figure 2 shows the network structure for a small subset of the market containing 21 lenders, eight aggregators and four holding companies, which are connected via loans originated and securitized in 2006. The network has 61 links, compared with 28 links in a tree representation, with 4 trees "rooted" at the holding company level. Thus, in a tree representation, more than half of the

Act of 2005 (BAPCPA) (Pub.L. 109-8, 119 Stat. 23) enacted April 20, 2005, and (2) extendable asset-backed commercial paper programs.

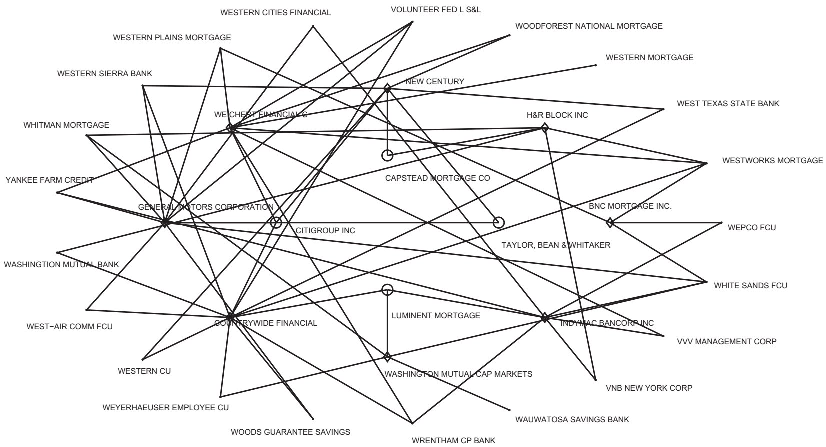

Figure 2
Between-firm networks
Between-firm networks for a subsample of private-label mortgage originations in 2006. There are twenty-one lenders (outer level), eight aggregators (middle level), and four holding companies (inner level). Links represent loan flows.

links would be unaccounted for. In Appendix A, a so-called "minimum spanning tree" (MST) representation of the subset is presented, which, by including the links with the highest loan flow volumes, provides the most complete tree representation of the market. For the full market of intermediaries, consisting in 2006 of 11,095 lenders, 2,030 aggregators, and 56 holding companies, the MST representation only accounts for about 20% of the links. The fraction of loans accounted for in the MST is higher—about 50%—but still a significant fraction of the market is excluded when using a tree representation rather than a network representation of the market.

Stanton, Walden, and Wallace (2014) find a positive relation between loan performance and the intermediary's network position. For example, they document a correlation of 0.23 between the average default rates of loans handled by an aggregator and those of other aggregators indirectly connected to that aggregator via a common node.[6]

Why is this network effect present? One potential reason is that a shock that affects an intermediary spills over to its neighbors through contractual and business interactions. This effect, which plays out once a shock occurs and which we therefore call an ex post effect, is in line with the propagation mechanism described, for example, in Elliott, Golub, and Jackson (2014). Another reason is that the very presence of ex post effects influences the behavior of intermediaries ex ante, before shocks occur, in their decisions regarding whom to interact with and which standards to choose. Indeed, a

[6] We verify and extend these results in Section 4.2.

common explanation for the observed heterogeneous performance of banks and other financial institutions is that they vary in their standards, through their ability to monitor and evaluate the projects they undertake (see Berger and Humphrey 1997; Billett, Flannery, and Garfinkel 1995, and references therein). Heterogeneous and decreasing lending standards were documented in the mortgage market in the years before the crisis (see Demyanyk and Van Hemert 2011; MacKenzie 2011; Poon 2008). In what follows, we develop a model in which both ex ante and ex post network effects are important for explaining default rates in the U.S. mortgage market.

2. A Strategic Network Model of Intermediaries

We introduce the model in several steps. There are N intermediaries with limited liability, each owned by a different risk-neutral agent. Each intermediary initially has full ownership of a project that generates risky cash flows at $t=1$, $\tilde{P}^n$, and may moreover incur some costs at $t=0$. The 1-period discount rate is normalized to 0. The intermediary's cash flow at $t=1$ is denoted by $\tilde{F}^n$, which may differ from $\tilde{P}^n$, because of insolvency costs and because intermediaries enter into risk-sharing agreements with each other. Agent n's objective is to maximize expected cash flows, net of any costs incurred at time 0,

$$V^n = E_0\left[\widetilde{F}^n\right] - C_0^n.$$

The risky project has scale $s^n > 0$, with two possible returns represented by the Bernoulli-distributed random variables ξ^n, so that $R^n = R_H$ if $\xi^n = 1$, and $R^n = R_L$ if $\xi^n = 0$. The probability, p, that $\xi^n = 0$ is exogenous, with $0 < p \ll 1$. We assume complete symmetry of risks in that the probability is the same for each of the N projects.[7] Realized cash flows to all agents are summarized by the random vector $\tilde{F} = \left(\tilde{F}^1, \dots, \tilde{F}^N\right)'$, and the realized project cash-flows are summarized by the vector $\tilde{P} = \left(\tilde{P}^1, \dots, \tilde{P}^N\right)'$.

Each agent has the option to invest a fixed amount, $C_0^n = cs^n$ at $t=0$, where $c > 0$, to increase the quality of the project. This cost is raised externally at $t=0$. If the agent invests, then in case of the low realization, $\xi^n = 0$, the return on the project is increased by $\Delta R > c$, to $R_L + \Delta R$. This investment cost could, for example, represent an investment in a screening procedure that allows the agent to filter out the parts of the project that are most vulnerable to shocks.

We represent the quality investment choice by the variable $q^n \in \{0, 1\}$, where $q^n = 1$ denotes that the intermediary invests in quality improvement. Intermediaries who choose $q = 1$ are said to be high quality, whereas those who choose $q = 0$ are said to be low quality. For the time being, we assume that c

[7] Note that each project may be viewed as a representative project for a portfolio of a large number of small projects with idiosyncratic risks that cancel out, and an aggregate risk component measured by ξ^n.

is the same for all intermediaries. We will subsequently allow c to vary across intermediaries, representing exogenous variation in costs.[8]

There is a threshold, $d>0$, such that if the return on the investment for an intermediary falls below d, additional costs are imposed, and no cash flows can be recovered by the agent.[9] For simplicity, we call these costs "insolvency costs," in line with Nier et al. (2008), who note that systemic events typically originate from insolvency shocks, although they are often also associated with liquidity constraints. We stress, however, that d would not necessarily represent the point at which a firm becomes insolvent, but more broadly a region below which additional costs are incurred. Also, although only important from a policy viewpoint and not for the actual predictions of the model, we take the view that these additional costs are wasteful in that they impose real costs on society rather than representing transfers.

It will be convenient to define the functions

$$X(z)=\begin{cases} 1, & z>d, \\ 0, & z\leq d, \end{cases} \qquad \text{and} \qquad Y(z)=X(z)z.$$

We also make the following parameter restrictions:

$$0\leq R_L<d, \qquad \text{and} \tag{1}$$

$$R_L+\Delta R<R_H. \tag{2}$$

The first restriction implies that there will be insolvency costs for a low-quality intermediary after a low realization. The second restriction states that the outcome in the high realization is always higher than in the low realization, even for a high-quality intermediary.

2.1 Isolated intermediaries

We first focus on the setting in which intermediaries do not interact and study the choice of whether to be high or low quality. The cash flows generated by project n are then

$$\tilde{P}^n=\begin{cases} s^n R_H, & \xi=1, \\ s^n(R_L+q\Delta R), & \xi=0, \end{cases} \tag{3}$$

the total cash flows generated to the owner after insolvency costs are accounted for are

$$\tilde{F}^n=s^n Y\left(\frac{\tilde{P}^n}{s^n}\right)=\begin{cases} s^n Y(R_H), & \xi=1, \\ s^n Y(R_L+q\Delta R), & \xi=0, \end{cases} \tag{4}$$

[8] Several variations of the model are possible, for example, assuming that screening costs, c, are part of the $t=1$ cash flows; fixing the scale of all intermediaries to $s=1$; and imposing a cost function concave in the size of the node, $c=c(s_n)$. All lead to qualitatively similar results. The version presented here was chosen for its tractability in combination with empirical relevance.

[9] This assumption, similar to assumptions made, for example, in Elliott, Golub, and Jackson (2014) and Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015), is a stylized way of modeling the additional costs related to risk for insolvency, for example, direct costs of bankruptcy, costs of fire sales, loss of human capital, and customer and supplier relationship capital. It could also more generally represent other types of convex costs of capital faced by a firm with low capitalization, along the lines described in Froot, Scharfstein, and Stein (1993).

and the $t=0$ value of the intermediary is

$$V^n(q)=E_0\left[\tilde{F}^n\right]-C_0^n=s^n((1-p)Y(R_H)+pY(R_L+q\Delta R)-qc). \quad (5)$$

Given the parameter restrictions (1), we have $V^n(0)=s^n(1-p)R_H$. We make the technical assumption that if an intermediary is indifferent between being high and low quality, it chooses to become low quality. Therefore, $q=1$ if and only if $V^n(1)>s^n(1-p)R_H$, immediately leading to the following result:

Proposition 1. An isolated intermediary chooses to be high quality, $q=1$, if and only if

$$R_L+\Delta R>\max\left(d,\frac{c}{p}\right). \quad (6)$$

Proposition 1 is very intuitive, implying that increases in the probability of a high outcome, the insolvency threshold, or the costs of quality investments, make it less attractive for an intermediary to be high quality. The first argument in the maximum function on the RHS ensures that a high-quality firm avoids insolvency in the low state. If the condition is not satisfied, there is no benefit to being high quality even in the low state. The second argument ensures that the expected increase of cash flows in the low state outweighs the cost of investing in quality. The value of the isolated intermediary when following the rule (6) is then $V_I^n=s^n V_I$, where

$$V_I=\begin{cases}(1-p)R_H, & q=0,\\ (1-p)R_H+p(R_L+\Delta R)-c, & q=1.\end{cases}$$

Note that the intermediary's optimal choice in (6) is decreasing in the solvency threshold, d, that is, for low thresholds it is potentially optimal to invest in quality, whereas for high thresholds it is not.

Also note that the objective functions of the agents coincide with that of society in this case with isolated intermediaries. Specifically, given that society has the social welfare function, $\mathcal{V}=\sum_n V^n$, under the constraint that intermediaries must act in isolation, the efficient outcome is realized by the intermediaries' joint actions. We obviously do not expect this to be the case in general, when agents interact.

2.2 Networks of intermediaries

Intermediaries may enter into bilateral contracts that transfer risk, and we represent such relations with links in a market network. Associated with the market network is the $N\times N$ sharing matrix Γ, where $\Gamma_{nn'}$ represents the fraction of risk that intermediary n' transfers to n.

The model's key ingredients are intuitive: The network allows intermediaries to diversify risks and thereby avoid insolvency in some states. However, the network also creates incentive problems when intermediaries pay the full cost of quality investments that they then share as benefits from with other intermediaries, and it also exposes intermediaries to systemic risk. Intermediaries are free to add or sever links if it makes them better off. An equilibrium network must be robust against such strategic changes in network structure, which significantly restricts the possible equilibrium outcomes.[10]

The market's network is represented by the graph $\mathcal{G}=(\mathcal{N},E)$, $\mathcal{N}=\{1,\dots,N\}$. The relation $E\subset\mathcal{N}\times\mathcal{N}$ describes which intermediaries (nodes) are connected. Specifically, the edge $e=(n,n')\in E$, if and only if there is a connection (edge, link) between intermediary n and n'. No intermediary is connected to itself, $(n,n)\notin E$ for all n, that is, E is irreflexive. The transpose of the link (n,n') is $(n,n')^T=(n',n)$. Connections are bidirectional, $e\in E\Leftrightarrow e^T\in E$. The operation $E+e=E\cup\{e,e^T\}$, augments the link e (and its transpose) to the network, whereas $E-e=E\backslash\{e,e^T\}$ severs the link if it exists. The number of neighbors of node n is $Z_n(E)=|\{(n,n')\in E\}|$.

Because of the high dimensionality of the problem when agents act strategically, we necessarily need a limited contract space. Specifically, we assume that the contracts available are such that intermediaries swap claims on the aggregate cash flows generated by their projects in a one-to-one fashion, similar to what is assumed in Allen, Babus, and Carletti (2012). Although obviously a simplification, this assumption qualitatively captures the bidirectional risk structure between intermediaries, discussed in the previous section. The contracts are settled according to a market-clearing system along the lines of that in Eisenberg and Noe (2001).

The risk-sharing rule is represented by the sharing matrix $\widehat{\Pi}\in\mathbb{R}_+^{N\times N}$, where $0\le\widehat{\Pi}_{nn'}\le\min\{s^n,s^{n'}\}$ is the amount of project risk that intermediary n receives from (and in turn gives to) intermediary n'. The residual risk intermediary keeps is $\widehat{\Pi}_{nn}$, leading to the summing up constraint, $\widehat{\Pi}\mathbf{1}=s$, where $s=(s^1,\dots,s^N)'$, and $\mathbf{1}=(1,1,\dots,1)'$ is a vector of N ones. For a general vector, $v\in\mathbb{R}^N$, we define the diagonal matrix $\Lambda_v=diag(v)\in\mathbb{R}^{N\times N}$. It will be convenient to characterize the *fraction* of the risk that intermediary n' provides to agent n, which is represented by element $\Pi_{nn'}$ of the matrix $\Pi=\widehat{\Pi}\Lambda_s^{-1}$, implying that $s=\Pi s$, and also the fraction of intermediary n's risk that it receives from n', represented by $\Gamma_{nn'}$ of the normalized sharing matrix

$$\Gamma=\Lambda_s^{-1}\widehat{\Pi},\tag{7}$$

which, in turn, satisfies $\mathbf{1}=\Gamma\mathbf{1}$.

The network represents a restriction on which sharing rules are feasible. This restriction may be self-imposed by intermediaries in equilibrium, who could choose not to interact even if they may, or it could be exogenous. Specifically, for a sharing rule to be feasible it must be that every off-diagonal element in the sharing matrix that is strictly positive is associated with a pair of agents who are linked, $\widehat{\Pi}_{nn'}>0\Rightarrow(n,n')\in E$.

---

[10] The reader who wishes to skip the formal details of the model may, with this intuition, directly move to Section 3.

We focus on a simple class of sharing rules that ensure that all weights are nonnegative and that each intermediary keeps some of its own project risk, namely

$$\widehat{\Pi}_{nn'} = \min\left\{\frac{s^n}{1+Z_n(E)}, \frac{s^{n'}}{1+Z_{n'}(E)}\right\}, \qquad n \neq n', \tag{8}$$

and

$$\widehat{\Pi}_{nn} = s^n - \sum_{n' \neq n} \widehat{\Pi}_{nn'}. \tag{9}$$

We write $\widehat{\Pi}(E)$ when stressing the underlying network from which the sharing rules is constructed.

The joint quality decision of all agents is represented by the vector $q = (q^1, \ldots, q^N)' \in \{0,1\}^N$. In the general case, the cost of investing in quality may vary across intermediaries, represented by the vector $c = (c^1, \ldots, c^N)' \in \mathbb{R}^N_{++}$. In our analysis, however, we focus on the case in which $c^n = c_0/N$ is the same for all intermediaries. The state realization is represented by the vector $\xi = (\xi^1, \ldots, \xi^N) \in \{0,1\}^N$. We will work with a limited state space, assuming that $\xi \in \Omega$, where $\Omega = \left\{\xi \in \{0,1\}^N : \xi'\mathbf{1} \geq N-1\right\} \subset \{0,1\}^N$, $\mathbb{P}(\xi = \mathbf{1} - \delta_n) = p = p_0/N$, $1 \leq n \leq N$. Here, δ_n represents a vector of zeros, except for the nth element which is 1. Thus, either zero or one low realization occurs, and the probability for a low realization is the same for all intermediaries.

2.3 Clearing mechanism

The clearing mechanism for contract settlements is similar to that of Eisenberg and Noe (2001) (see also Acemoglu, Ozdaglar, and Tahbaz-Salehi 2015; Elliott, Golub, and Jackson 2014), providing a mapping from project cash flows, $\tilde{P}$, to realized net settlement cash flows, $\tilde{F} = \mathcal{CM}[\tilde{P}]$. The outcome is chosen so that it minimizes the number of defaults (which all agents agree are costly), by solving

$$\tilde{F} = \Lambda_f \tilde{G}, \qquad \text{where} \tag{10}$$

$$\tilde{G} = \max_{f \in \{0,1\}^N} (\Pi \Lambda_f) \times \tilde{P}, \qquad \text{s.t.} \tag{11}$$

$$f = X\left(\Lambda_s^{-1} \tilde{F}\right). \tag{12}$$

Here, f, represents the *solvency vector* $f \in \{0,1\}^N$, with $f^n = 1$ if intermediary n is solvent in equilibrium and $f^n = 0$ if intermediary n is insolvent. Moreover, $\tilde{G}$ represents the cash flows generated by intermediaries, taking into account the default of all other intermediaries but not the intermediaries' own defaults, whereas $\tilde{F}$ represents the cash flows to the owners (which will be zero below the insolvency threshold), taking into account the intermediaries' own defaults. Condition (11) ensures that the number of solvent intermediaries is maximized, whereas condition (12)—in which the function X is applied element-wise on the input vector—is a constraint that ensures that only intermediaries above the solvency threshold remain solvent.

As discussed in the appendix, (10)–(12) can be solved by initially assuming that all intermediaries remain solvent, and then iteratively updating those that violate (12) until no further insolvencies occur. This leads to a dynamic interpretation of how defaults propagate: after m steps of the iteration, $\tilde{F}_m$ and $\tilde{G}_m$ represent the cash flows, given the insolvencies that have occurred so far.

The net cash flows to the intermediaries are given by the vector

$$w(\xi|q,E)=\tilde{F}(\xi,q)-\Lambda_c\Lambda_s q, \tag{13}$$

and the $t=0$ value-vector of intermediaries, given quality investments, q, and network E is then given by

$$V(q|E)=\sum_{\xi\in\Omega} w(\xi|q,E)\mathbb{P}(\xi). \tag{14}$$

2.4 Equilibrium

We build upon the pairwise stability concept of Jackson and Wolinsky (1996), in addition requiring equilibrium in the multistage game to be subgame perfect. Starting with a proposed equilibrium network, E, each agent may unilaterally decide to sever a link to another agent or to sever links with all other agents. Bilaterally two agents can decide to add a link between themselves. In equilibrium we require agents to have correct beliefs about the quality decisions made by their counterparties. Appendix C provides an extended treatment.

Let E^* denote the complete network in which all nodes are connected and $E_0=\emptyset$ the empty network in which all nodes are isolated. There is a maximum possible network, $\bar{E}\subset E^*$, such that only links that belong to $\bar{E}$ may exist in the equilibrium network. This restriction on feasible networks could, for example, represent environments in which it is impossible for some agents to credibly commit to deliver on a contract written with some other agents, because of low relationship capital or limited contract enforcement across jurisdictions. A network, E, is feasible if $E\subset\bar{E}$. If all agents who may be linked actually choose to be linked in equilibrium, that is, if $E=\bar{E}$, we say that the equilibrium network is *maximal*. It also may be the case that E is a strict subnetwork of $\bar{E}$. The primitives of the economy are then summarized in the tuple $\mathcal{E}=(N,\bar{E},s,p_0,c_0,R_L,R_H,\Delta R,d)$.

The sequence of events in the game is described in Figure 3. At $t=-2$, each agent may unilaterally decide to sever a link to another agent, or to all other agents. Each agent can also propose to add a link, in which case the other agent has the option to accept or decline at $t=-1$. Then, after the resultant network is determined, agents choose quality and outcomes are realized. Note that we implicitly assume that the actual quality decision is not contractible.[11]

[11] In our stylized model, the quality decision can be inferred from the realization of project cash flows. This issue would be avoided by assuming a small positive probability for $\Delta R=0$ in case of a low realization with quality investments.

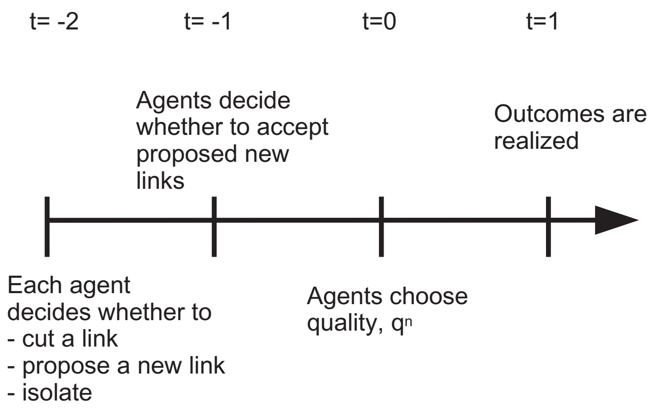

Figure 3
Sequence of events in network formation game with endogenous financial norms

Consider a candidate equilibrium, represented by a network E and quality choices q. Each agent, n, has the opportunity to accept the sharing rule, $\hat{\Pi}(E)$, as is, by neither severing nor proposing new links at $t=-2$. But, in line with the pairwise stability concept, any agent n can also unilaterally decide to sever a link with one neighbor, n', leading to the sharing network $E' = E-(n,n')$, and corresponding sharing rule $\hat{\Pi}(E')$. Also, any agent can propose an augmentation of another link $(n,n') \in \bar{E} \setminus E$, which if agent n' accepts leads to the sharing network $E'' = E+(n,n')$ with sharing rule $\hat{\Pi}(E'')$. Finally, we assume that each agent can unilaterally choose the isolated outcome, V_I^n, by severing its links to all other agents.

The possibility to unilaterally sever all links in a sharing network—although not technically part of the standard definition of pairwise stability—is natural, in line with there being a participation constraint that no intermediary can be forced to violate. It provides a minor extension of the strategy space.

The severance, proposal, and acceptance/rejection of links occur at $t=-2$ and $t=-1$. The agents then decide whether to invest in quality or not at $t=0$, each agent choosing $q^n \in \{0,1\}$. A pair (q,E), where $E \subset \bar{E}$, is defined to be an equilibrium, if agents given network structure E choose investment strategy q, if no agent given beliefs about other agent's actions—under the current network structure as well as under all other feasible network structures in $\bar{E}$—has an incentive to either propose new links or sever links, and if every agent's beliefs about other agents actions under network E as well as under all feasible alternative network structures are correct.

3. Analysis of Equilibrium

We first analyze some specific types of equilibrium networks, shown in Figure 4, then use simulations to study the relationship between quality choice and

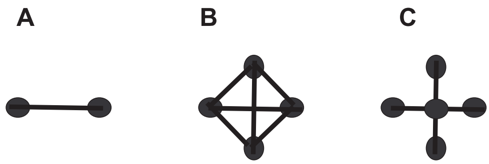

Figure 4
Network examples
(A) Two-node network, (B) complete network, E^* , and (C) star network, E^S.

network position, and finally provide a partial characterization of node behavior in general networks. The key takeaways from our analysis, relating local network position, global network structure, and quality, is that heterogeneous financial norms tend to be clustered in the network, that the equilibrium quality decision of a node is related to the size and connectedness of that nodes' neighbors, and that the quality decisions of a *systemically pivotal* subset of nodes are disproportionately important for the systemic vulnerability of the market.

3.1 Two intermediaries

Consider the case with two intermediaries in the network with risk sharing shown in Figure 4A, with scales $s^1 = s^2 = 1$. The risk-sharing contract is such that intermediary 1 agrees to deliver $\tilde{P}^1/2$ to intermediary 2 at $t = 1$, and, in turn, receive $\tilde{P}^2/2$ from intermediary 2. The sharing matrix in such a risk-sharing equilibrium is

$$\Gamma = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

There is also another possible equilibrium in this case, the isolated one E_0, with sharing matrix

$$\Gamma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

For the action $q \in \{0,1\}$, we let $\neg q$ denote the complementary action ($\neg q = 1 - q$). The conditions for risk sharing to be an equilibrium are then the following

$$V^1\left(q^1, q^2 | E^*\right) \geq V^1\left(\neg q^1, q^2 | E^*\right), \qquad V^2\left(q^1, q^2 | E^*\right) \geq V^2\left(q^1, \neg q^2 | E^*\right), \tag{15}$$

$$V^1\left(q^1, q^2 | E^*\right) \geq V_I^1, \qquad V^2\left(q^1, q^2 | E^*\right) \geq V_I^2. \tag{16}$$

Condition (15) ensures that it is incentive compatible for both agents to choose the suggested investment strategies given that they share risks, while condition (16) presents participation constraints that state that risk sharing dominates acting in isolation for both agents.

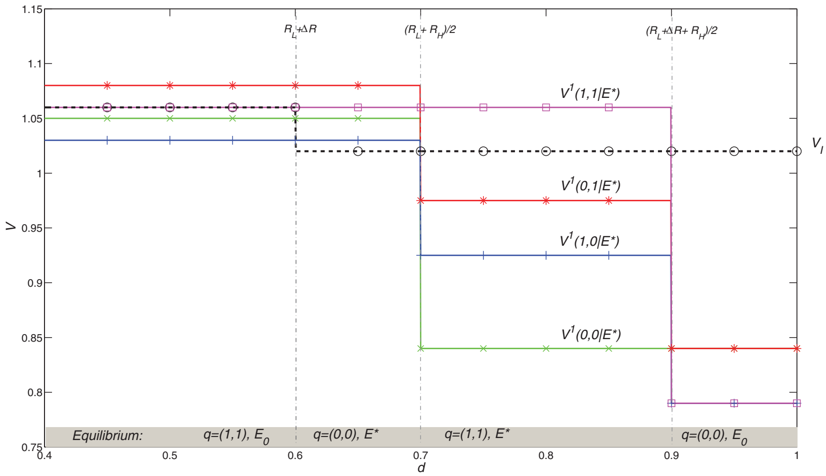

Figure 5
Value functions
Value functions in economy with two intermediaries, as a function of insolvency threshold, d. The following value functions are shown: V_I (circles, dotted black line), $V^1(1,1|E^*)$ (squares, magenta), $V^1(0,1|E^*)$ (stars, red), $V^1(1,0|E^*)$ (pluses, blue), and $V^1(0,0|E^*)$ (crosses, green).

We choose parameter values $R_H = 1.2$, $R_L = 0.2$, $\Delta R = 0.4$, $c = 0.05$, $p = 0.15$, and vary the insolvency threshold, d. Figure 5 shows the resultant value functions.

There are four different regions with qualitatively different equilibrium behavior. In the first (leftmost) region, where $0.6 = R_L + \Delta R > d$, the unique equilibrium is the one where both agents invest in quality ($q = (1,1)$) and there is no risk-sharing (network E_0), leading to values V_I for both intermediaries. No intermediary ever becomes insolvent in this region (since $d < R_L + \Delta R$). The outcome where both agents invest and share risk would lead to the same values, but cannot be an equilibrium because each agent would deviate and choose to avoid investments in this case, given that the other agent invests. This can be seen by observing that the red line with stars corresponding to $V^1(0,1|E^*)$, that is, risk sharing with low quality for agent 1 and high quality for agent 2, is above the magenta line with squares, which corresponds to $V^1(1,1|E^*)$, that is, risk sharing and high quality for both agents. The deviation strategy is profitable because the deviating agent shares the benefits of quality investments, and therefore the benefit of investing is only $p \times \Delta R/2 = 0.15 \times 0.4/2 = 0.03$ at $t = 1$, but he incurs the whole investment cost of $c = 0.05$ at $t = 0$. The agents therefore cannot coordinate on the high-quality outcome.

In the second region, where $0.6 = R_L + \Delta R \leq d < 0.7 = (R_L + R_H)/2$, the equilibrium changes so that both agents choose low quality, $q = (0,0)$ and share risks, that is, the equilibrium network is E^*. The outcome corresponds to the green line with crosses, and default never occurs. The reason why isolation is no longer an equilibrium is that quality investments in isolation are no longer sufficient to avoid default after a low shock realization ($R_L + \Delta R \leq d$). Therefore, V^I drops substantially. A risk-sharing outcome with quality investments would increase the value to the magenta line with squares, but coordination on this outcome is not possible for the same reason in the first region.

In the third region, where $0.7 = (R_L + R_H)/2 \leq d < 0.9 = (R_L + \Delta R + R_H)/2$, the equilibrium is one of risk sharing, E^*, with quality investments $q = (1,1)$ (magenta line with squares). The reason why it is possible to coordinate on quality investments in this region is that without quality investments, default occurs after a bad realization ($(R_L + R_H)/2 \leq d$), so the expected benefit for an agent of investing in quality rather than deviating is $p((R_L + \Delta R + R_H)/2) = 0.15 \times 0.9 = 0.135$, which readily outweighs the cost of 0.05.

Interestingly, although insolvency costs are higher in the third region than in the second region, so is equilibrium welfare. This suggests that if the insolvency cost friction can be affected, for example, by a regulator, the policy implications of decreasing d may be nontrivial: decreasing the friction may actually decrease welfare.

Finally, in the fourth (rightmost) region, where $0.9 = (R_L + \Delta R + R_H)/2 \leq d$, default always occurs after a bad realization regardless of the risk-sharing agreement and quality investments. Investment in quality is therefore useless and the agents also choose to be isolated since defaults will propagate if they share risk (increasing the individual probability of default from p to $2p$). Thus, $q = (0,0)$, and the network is E_0 in this region.

We note that $q^1 = q^2$ for all equilibrium outcomes. This suggests that the "financial norm" of an intermediary—defined as its quality—depends on the financial norm of the intermediary with which it interacts, in line with the intuition that norms are jointly determined among interacting agents. In the two-agent case, it is straightforward to show that this property is generic:

Proposition 2. In any equilibrium of the network model with two agents, both agents make the same quality choice, that is, $q^1 = q^2$. No insolvencies occur in equilibria with risk sharing.

In our terminology, intermediaries share financial norms in equilibrium.[12] The possibility for agents to cut links is crucial for the result. As shown in the proof

[12] Note that our notion of financial norms follows from game theoretic principles. Each intermediary maximizes its own "utility" (defined as its total expected payoff), and the common quality is an equilibrium outcome. Our approach is in line with much of the economics literature on networks, but has been criticized in the sociology and organizations literature as being too narrow to capture the richness of network interactions in a business environment. Granovetter (1985) and Uzzi (1997), for example, suggest that economic actions are embedded in structures of social interaction, and therefore must be understood within such a context. Such considerations would likely amplify the importance of financial norms if incorporated into our model.

of Proposition 2, if the risk-sharing network was exogenously given, it would be possible to get a risk-sharing outcome with heterogeneous financial norms. In other words, common equilibrium financial norms are a consequence of the possibility for agents to affect network structure.

3.2 Complete network

The complete network E^* is shown in Figure 4B with $N=4$ intermediaries. For large N, we can completely characterize when a complete network is an equilibrium, as shown in the following proposition:

Proposition 3. Consider the economy $\mathcal{E}$ with maximal network $\bar{E}=E^*$. For large N:

(a) If $R_L+\Delta R \leq d$ or $\Delta R \leq \frac{c_0}{p_0}$, then the complete network, E^*, with no quality investments, $q=\mathbf{0}$, is an equilibrium. No intermediary ever becomes insolvent in this equilibrium, and there is no other complete-network equilibrium.

(b) If both conditions in (a) fail, then there is no complete-network equilibrium.

The result suggests that the complete network efficiently facilitates complete risk sharing—no insolvencies occur—but may not achieve efficient quality investments. Indeed, intermediaries may well be better off if they could coordinate to invest in quality, but the free-rider problem becomes more severe the larger the complete network, since each intermediary shares less of the benefits of quality investments the larger the network. Therefore, only low-quality equilibria may survive in large complete networks. When the conditions in (a) fail, it is more efficient for an intermediary to isolate and invest in quality, and the low-quality complete network no longer constitutes an equilibrium.

Also note that, like in the two-node network (which is a special case of a complete network), all agents share the same (low) financial norms in large complete network equilibria.

3.3 Star network

The star network E^S is shown with $N=5$ intermediaries in Figure 4C. Note that the network is asymmetric in that the central "hub" has 4 neighbors whereas the peripheral nodes, $n=2,\ldots,5$ have one neighbor each. We first study the special case where the hub is exactly as large as all the other nodes together, $s_1=N-1$, $s_2,\ldots,s_N=1$, so that each peripheral node shares exactly half of its risk with the hub, which, in turn, shares half of its risk with the other nodes altogether. In this case, we have

Proposition 4. Consider the economy $\mathcal{E}$ with parameters, $s_1=N-1$, $s_2,\ldots,s_N=1$, and maximal network $\bar{E}=E^*$. For large N:

(i) If $(R_L + R_H)/2 > d$ and $\Delta R/2 > c_0/p_0$, or if $(R_L + R_H)/2 \leq d$ and $R_L + \Delta R > c_0/p_0$, then the star network E^S in which all nodes are high quality, $q = \mathbf{1}$, is an equilibrium.

(ii) If $(R_L + R_H)/2 > d$ and $\Delta R/2 \leq c_0/p_0$, and either $R + \Delta R \leq d$ or $\Delta R < c_0/p_0$, then the star network E^S in which all nodes are low quality, $q = \mathbf{0}$, is an equilibrium.

(iii) There is no equilibrium star network in which there are nodes of both high and low quality.

Thus, for large star network equilibria, the financial norms are again homogeneous in this case. Like in previous cases, high-quality nodes will not keep connections with low-quality nodes, either preferring to isolate or to switch quality.

If we allow other size relations between the hub and peripheral nodes, then there exist equilibria for which the financial norms of the nodes differ. We have:

Proposition 5. Consider the economy $\mathcal{E}$ with parameters $s_1 = \alpha(N-1), \alpha > 0$, $s_2, \ldots, s_N = 1$, and maximal network $\bar{E} = E^*$. For large N:

(a) For all α, there exist star network equilibria in which all nodes have high quality, as well as star network equilibria in which all nodes have low quality.

(b) If $\alpha < 1$, there are star network equilibria for which the hub is low quality and all the peripheral nodes are high quality, whereas if $\alpha \geq 1$, there are no such star network equilibria.

(c) For no α is there a star network equilibrium for which the hub is high quality and all the peripheral nodes are low quality.

The result provides further intuition about the relationship between size, number of connections, and quality choice. Specifically, within the class of star networks, all else equal, it is more likely for a node to be high quality if it is larger and if it has fewer connections. In both cases the node keeps a larger fraction of its own project, and therefore has a higher incentive to invest in quality.

Note that there is an asymmetry in Proposition 5 in that it is possible to have a hub with low norm and peripheral nodes with high norms, whereas the reverse equilibrium outcome is not possible. The reason is that in a hypothetical equilibrium in which the hub is high quality, and the peripheral nodes are low quality, the hub can always increase its value by severing one link. Indeed, such a move is preferred by the hub, since it does not change the states in which insolvency occurs because each peripheral node has a very marginal impact on the hub, while allowing it to keep more of its high-quality project.

Table 2
Summary statistics of high- and low-quality intermediaries

Average	Number in network	$q(neighbors)$	Number of neighbors	cost
$q=1$	6.34	0.79	4.32	0.011
$q=0$	2.66	0.4	3.12	0.016

Summary statistics of high- and low-quality intermediaries. Number of simulations: 1,000. Parameters: $R_H=1.1$, $R_L=0.2$, $\Delta R=0.3$, $d=0.75$, $p=4/90$, $c\sim U(0,0.025)$. Scale $s=\mathbf{1}$.

The situation is quite different in an equilibrium with a low-quality hub and high-quality peripheral nodes. In this case, the deviation strategy is for two peripheral nodes to form a link, but because of the more severe incentive problems they face when they share more risk, the outcome may be that the peripheral nodes switch to low quality in this case, making the outcome worse. Therefore, the star network with a low-quality hub and high-quality peripheral nodes survives as an equilibrium. We stress that the aforementioned argument is only valid for large networks.

3.4 Random networks

We explore whether the common financial norms documented so far extend to general, less structured networks. Specifically, using simulations, we explore the relationship between network structure and quality choice for a class of randomly generated networks. We simulate 1,000 networks, each with $N=9$ nodes. We use the classical Erdös-Renyi random-graph-generation model, in which the probability that there is a link between any two nodes is i.i.d., with the probability 0.25 for a link between any two nodes, and we also randomly vary c across intermediaries, $c^n \sim U(0,0.025)$.[13] For computational reasons, we focus on networks in which equilibria are maximal, $E=\bar{E}$.

Table 2 shows summary statistics for high-quality nodes compared with low-quality nodes equilibrium. We see that there are on average more high-quality nodes. Also, not surprisingly, the average cost of investing in quality for high-quality nodes is lower than for low-quality nodes. Most interestingly, the average quality of neighbors of high-quality nodes is higher than of low-quality nodes. These differences are statistically significant. The last result is especially important, since it shows that the financial norms are indeed closely related to network position in that they tend to be clustered, in line with our results for two intermediaries.

Another way of measuring the presence of such clusters is to partition each network into a high-quality and a low-quality component, and study whether the number of links between these two clusters is lower than it would be if quality were randomly generated across nodes. Specifically,

13 Other possible network generation processes also would be possible, for example, the preferential attachment model introduced in Barabasi and Albert (1999). Given the small size of our simulated networks ($N=9$), the choice of network generation process is not crucial for the results.

consider a network with a total of $K = |\{(n,n') \in E\}|$ links, and a partition of the nodes into two clusters: $\mathcal{N} = \mathcal{N}^A \cup \mathcal{N}^B$, of size $N^A = |\mathcal{N}^A|$ and $N^B = |\mathcal{N}^B| = N - N^A$, respectively, and the number of links between the two components: $M = |\{(n,n') \in E : n \in \mathcal{N}^A, n' \in \mathcal{N}^B\}|$. In the terminology of graphs, M is the size of the cut-set, and is lower the more disjoint the two clusters are. The number of links one would expect between the two clusters, if links were randomly generated, would be $W = \frac{1}{N(N-1)} N^A N^B K$, so if the average M in the simulations is significantly lower than the average W, this provides further evidence that financial norms are clustered. Indeed, the average M in our simulations is 12.2, substantially lower than the average W which is 14.2, corroborating that network position is related to financial norms.

The intuition for why norms are clustered is that, in general, connected nodes need to coordinate on high quality. When such coordination fails, nodes jointly choose low quality instead. A few nodes may "bridge" high- and low-quality clusters. For such nodes, it is optimal to be high quality, even though they are connected to some low-quality nodes, because of the diversification benefits these low-quality connections provide.

3.5 Equilibrium behavior in general networks

A complete characterization of the equilibrium behavior of all nodes in a general network is out of reach, but we are able to characterize the behavior of a usually significant majority of the nodes, namely those whose quality choices do not affect the cash flows they receive from other nodes. In Section 2.3, we introduced the solvency vector, f, where f^n represents whether intermediary n is solvent in equilibrium.

Definition 1. Node n is said to be *systemically pivotal* in equilibrium if its quality choice influences the set of nodes that become insolvent in some state, that is, if for some state $\xi \in \Omega$, and intermediary $m \neq n$,

$$f(\xi, q)^m \neq f(\xi, (\neg q^n, q^{-n}))^m.$$

Here, $(\neg q^n, q^{-n})$ denotes the action vector for which agent n switches to the complementary action, and all other agents choose the same actions used in equilibrium (see Appendix C).

A node that is merely a channel through which shocks propagate will not be classified as systemically pivotal. In other words, our definition focuses on how financial norms generate systemic risk. We partition the network into the set of nodes that are systemically pivotal, $\mathcal{N}^S$, and those that are not, $\mathcal{N}^U$. Nodes that belong to $\mathcal{N}^S$ are of especially high interest, since their quality decisions impact the solvency of other nodes in the network, creating a systemic externality.

The behavior of nodes in $\mathcal{N}^U$ is straightforward to characterize. We define $\Theta^{n,z} = R_L \Gamma_{nn} + R_H \sum_{j \neq n} \Gamma_{jn} f^j(\mathbf{1} - \delta_n, (q^n = z, q^{-n})), z \in \{0,1\}$,[14] and then have

Proposition 6. Any node $n \in \mathcal{N}^U$ is high quality in equilibrium, $q^n = 1$, if and only if one of the following two conditions holds:

(1) $d - \Theta^{n,1} < 0$ (1a) and $\frac{c^n}{p} < \Gamma_{nn} \Delta R$ (1b),

(2) $0 \leq d - \Theta^{n,1} < \Gamma_{nn} \Delta R$ (2a) and $\frac{c^n}{p} - \Theta^{n,1} < \Gamma_{nn} \Delta R$ (2b).

Note that Proposition 1 for the two-node network is a special case of Proposition 6, because when node n is not connected to any other node it follows that $\Gamma_{nn} = 1$, $\Theta^n = R_L$, and that condition (1) is never satisfied (because of the assumption that $R_L < d$). Moreover, condition (2) then collapses to $d - R_L < \Delta R$, and $\frac{c^n}{p} - R_L < \Delta R$, which are equivalent to the condition for high quality given in Proposition 1. Note that (1b) imposes a stronger condition on a connected node for being high quality than for an isolated node, since $\Gamma_{nn} < 1$ for a connected node. The reason is the incentive problem that arises because a connected node shares the benefits of quality investments with other nodes, in contrast to an isolated node.

For nodes in $\mathcal{N}^S$, the above restrictions on equilibrium behavior do not apply. The following proposition states that systemically pivotal nodes always have at least as high incentive to be high quality as nodes that are not systemically pivotal, and it is straightforward to show:

Proposition 7. For any node $n \in \mathcal{N}^S$, each of conditions (1) and (2) in Proposition 6 is sufficient for n to be high quality in equilibrium, $q^n = 1$.

The intuition behind the result is straightforward: systemically pivotal nodes are not only directly affected by a negative shock if they decide to be low quality but also potentially by negative feedback from other nodes that they affect. The feedback effect always increases the value of being high quality compared with nodes that are not systemically pivotal. Nonpivotal nodes only affect the network locally, as shown by the following proposition.

Proposition 8. Only systemically pivotal nodes may through their quality choices affect the cash flows of nodes beyond their direct neighbors.

In practice, systemically pivotal nodes tend to make up a small fraction of the nodes in the network, that is, there will be many nonpivotal nodes that behave

[14] For nodes that are not systemically pivotal, $n \in \mathcal{N}^U$, the insolvencies f^j, $j \neq n$ per definition do not depend on the quality choice of node n, that is, $\Theta^{n,z}$ does not depend on z, so we can drop the z dependence for such nodes and write Θ^n.

in a manner that is straightforward to characterize, in line with Proposition 6, and a smaller set of pivotal nodes with richer behavior. We note that although pivotal nodes are more motivated to invest in quality than nonpivotal nodes, they still do not internalize the full effect of their quality choices, in contrast to the market with isolated nodes in Section 2.1.[15]

We study how local network characteristics are related to whether a node is systemically pivotal. Define

$$\kappa = \frac{R_H - d}{R_H - R_L}, \tag{17}$$

which is a normalized measure of the slack between the insolvency threshold and the high outcome. We say that node n is *well diversified* if $\Gamma_{nn} < \kappa$. It is easy to see that well-diversified nodes survive after a low shock realization, even if they are low quality. It also follows that:

Proposition 9. Consider an economy with normalized slack κ defined by (17). If node n's neighbors are all well diversified, then n is not systemically pivotal, $n \in \mathcal{N}^U$.

Node n's prospects of being well diversified depends on it own connectivity, but also on the connections and sizes of its neighbors. All else equal, node n's likelihood to be well diversified (because of a lower Γ_{nn}) increases the larger the sizes of its connections, and the fewer connections they have (Γ_{nn} is nonincreasing in Γ_{nj}, and Γ_{nj} is, in turn, nondecreasing in s^j and nonincreasing in Z_j). Moreover, the lower Γ_{nn} is for a well-diversified node, the lower the incentive of the node to invest in quality, because of the free-rider problem, altogether leading to

Corollary 1. For a well-diversified node, n, all else equal,

(i) The more connected node n is, the more likely it is that n is low quality.
(ii) The less connected the neighbors of node n are, the more likely it is that n is low quality.
(iii) The larger the neighbors of node n are, the more likely it is that n is low quality.

These argument are based on local properties, that is, on the network properties in a neighborhood of a given node. Similar local arguments also holds for an extension of the model in which d varies across the network, for example, because of different regulatory regimes, although this extension is not part of our base model. A higher d implies a lower κ and therefore a stricter hurdle

[15] Because of the potential nonlocal effect of the actions of pivotal nodes, a regulator who is concerned about contagion should therefore focus on these nodes, potentially being "too pivotal to fail."

for a node to be well diversified. Moreover, a well-diversified node has, all else equal, less of an incentive to be high quality than a node that is not well diversified, since quality investments may save the latter node from default, leading to:

Corollary 2. In an economy in which the solvency threshold d varies, nodes are more likely to be low quality in regions with lower d.

The insolvency threshold for a node also may be related to other node characteristics, for example, the number of neighbors, Z_n.[16] In case of a negative relationship between d_n and Z_n, the two effects from Corollary 1(i) and Corollary 2 reinforce each other, whereas they go in opposite directions when there is a positive relationship. In unreported numerical simulations we have verified that the insolvency threshold effect dominates, so that high-quality nodes are associated with higher insolvency thresholds both when the relationships between size and insolvency threshold are positive and negative.

4. Application to the Mortgage Market

This section investigates the extent to which the model can help to explain the empirical findings of Stanton, Walden, and Wallace (2014), who documented a strong relationship between loan performance and empirical mortgage-network characteristics. We also test several new empirical predictions. Specifically (more detail is provided below):

1. There should be a positive relation between the loan performance of a firm and that of its neighbors.
2. There should be a positive relation between the financial distress of a firm and that of its neighbors.
3. The estimated quality measure in the model should be related to other, independently observable, measures of quality.
4. There should be a positive relation between predicted and realized loan performance.
5. There should be lower predicted performance of firms that defaulted than of those that remained in the sample.
6. Ex ante effects, that is, financial norms, should be important for the model's ability to predict future defaults.

4.1 Data

To evaluate the model, we use data assembled from Dataquick, ABSNet, and Dealrate.com. Table 3 provides summary statistics.

[16] We thank a referee for pointing this out.

Table 3
Summary statistics

Variable name	Mean	SD	Minimum	Maximum
	2006			
Original loan balance ($)	277,425.00	224,182.00	50,000.00	989,560.00
Original cumulative loan to value ratio	0.79	0.16	0.14	1.65
Original mortgage contract rate	0.07	0.03	0.04	0.13
Original mortgage amortization term (years)	31.38	4.41	10.00	40.00
Conventional conforming loan indicator	0.33		0	1
FICO score	670	67	370	850
No documentation	0.68		0	1
Loan origination cost	0.15	0.04	0.07	0.28
Loan-level default	0.37		0	1
Geographic effects (excluding subject loan)				
Loan ZIP code average loan-level default	0.37	0.13	0	1
Weighted average loan-level default in contiguous ZIP codes	0.32	0.12	0	1
Network effects (excluding subject node)				
Neighboring lenders average loan-level default rate	0.44	0.01	0.0000	1.0000
Neighboring aggregators average loan-level default rate	0.38	0.01	0.24	0.53
Neighboring holding company average loan-level default rate	0.38	0.02	0.03	0.49
Local market economy (ZIP code averages, 2006–2013)				
Average adjusted gross income ($)	53,115.00	39,308.00	14,635.00	2,157,816.00
Average median sales price ($)	302,545.00	323,080.00	17,666.00	16,143,015.00
Lender Type				
Bank indicator	0.30		0	1
Savings and loan institution indicator	0.08		0	1
Mortgage company indicator	0.59		0	1
Credit union indicator	0.03		0	1
Number of loans	1,152,312			
	2007			
Original loan balance ($)	369,627.00	264,452.00	51,000.00	2,000,000.00
Original cumulative loan to value ratio	0.83	0.12	0.60	1.18
Original mortgage contract rate	0.07	0.01	0.05	0.12
Original mortgage amortization term (years)	31.30	4.42	10.00	40.17
Conventional conforming loan indicator	0.47		0	1
Loan origination cost	0.14	0.04	0.07	0.240
Loan-level default	0.3337		0	1
Number of loans	522,875			
Lender performance 2007				
Average share of 2006 network linked lenders that disappeared (closure or bankruptcy) from the 2007 network, excluding subject lender	0.2670	0.0297	0.0093	1
Average loan-level default in 2006 network neighbors (January 2006–December 2006)	0.0115	0.0002	0.0092	0.0142
Average loan-level default in contiguous ZIP codes (January 2000–December 2006)	0.0558	0.0325	0	0.3343
Number of lenders (2006)	11,095			
Number of lenders (2007)	5,957			

Summary statistics for loan origination characteristics, default performance (60 days delinquent), the performance of loans in the loan's ZIP code (excluding subject loan), the performance over firms' network paths, and aggregate lender performance.

4.1.1 Individual loan details, 2006. The first eight rows of the upper panel of Table 3 provide loan contract and origination summary statistics for the first-lien, fixed-rate mortgages originated and securitized in private-label mortgage-backed securities in 2006. The average loan balance was $277,425 and the average cumulative loan-to-value ratio was 79%. The average coupon on the mortgages was 7% and the average maturity was 33 years. About 33% of the loans were conventional, conforming loans and the average FICO score was 670. Loan costs averaged about 15% of the balance, with a standard deviation of 4%. In total, 37% of the loans defaulted during the period.

4.1.2 Geographic effects. We construct two ZIP-code-level measures for local mortgage-market conditions and the localized incidence of loan defaults. The first of these measures, *Loan ZIP code average loan-level default*, is the average mortgage default incidence through 2013, excluding the subject loan, where default is defined as real estate owned, foreclosure, bankruptcy, 60, 90, 120, and 150+ day delinquencies for 2006 mortgage originations within that ZIP code. Our second measure, *Weighted average loan-level default in contiguous ZIP codes* is constructed by geoprocessing the U.S. Census shape files for ZIP code tabulation areas (ZCTAs) using polygon neighbor functions to identify the continguous neighbors for each ZCTA and then using the ZIP code for each loan to map loans onto their appropriate ZCTA and contiguous neighbors. We then use ZCTA centroid distances and the 2006 mortgage origination balances to construct a weighted average loan-level default rate for all of the loans originated in ZIP codes contiguous to the subject loan's ZIP code. Our definition of loan default uses the same performance outcome variables as those defined above, and the evaluation period is again through 2013. As shown in Table 3, the average ZIP-code-level default was about 37% for 2006 mortgage originations over the period, and the average default rate among 2006 loans originated in contiguous ZIP codes was about 32%.

4.1.3 Network neighbor default performance. We construct the network, E, from the loan flows, as discussed in Section 1. The performance within the network neighborhoods is measured as the average (excluding the subject node) of the default rates of nodes that are indirectly connected to a node in the network (i.e., at distance 2). We have three measures of performance within network neighborhoods: (1) *Neighboring Lenders Average Default Rate*, the average default rate of lenders connected to a lender via an aggregator; (2) *Neighboring Aggregator Average Default Rate*, the average default rate of aggregators connected to an aggregator, either via a lender or via a holding company; and (3) *Neighboring Holding Company Average Default Rate*, the average default rate of holding companies connected to a holding company via an aggregator.

The next section of Table 3 presents the default performance within the three network neighborhoods of the node, again excluding the subject node.

As shown, the average lender default level for neighboring network lenders was 44%. The average default level for neighboring aggregators to the subject firm's aggregator was 38%, and the average default level for the neighboring holding company was 38%.

4.1.4 Local economy. The local market economy is measured using ZIP-code-level averages from 2006 to 2013 for adjusted gross income obtained from the Internal Revenue Service and for median sales prices for homes obtained from Zillow.[17] As shown in Table 3, the average ZIP-code-adjusted gross income over the period was $53,115 and the average median house price was $302,545.

4.1.5 Lender type. To control for possible regulatory effects (see, e.g., Agarwal et al. 2014), we introduce indicator variables for the type of lending institution that originated the loan. We have five lender types: banks, savings and loan institutions, credit unions, mortgage companies, and other. The banks are primarily regulated by the Office of the Controller of the Currency, the savings and loan institutions were regulated at the time by the Office of Thrift Supervision, the credit unions by the National Credit Union Administration, and the mortgage companies were primarily regulated by the Department of Housing and Urban Development. The composition of lender types in the data is 30% banks, 8% savings and loans, 59% mortgage companies, and 3% credit unions. The remaining mortgages were originated by finance companies and home builders.

4.1.6 Individual loan details, 2007. The second-lowest panel of Table 3 presents summary statistics for the first-lien, fixed-rate mortgages originated and securitized in private-label mortgage-backed securities in 2007. As shown, the loan balances at origination were larger, the cumulative loan-to-value ratios were larger, and the contract rates were higher than in the 2006 originations. A larger share of the fixed rate mortgages were conventional conforming, and the ex post default rates are similar.

4.1.7 Firms that left sample in 2007. The lowest panel of Table 3 shows that overall, of the 11,095 firms in the sample in 2006, 5,138 (46.3%) dropped out in 2007 due either to closure or to bankruptcy. Of the indirect lender-to-lender connections in the 2006 network, 26.7% disappeared during 2007.

---

[17] The IRS data were obtained from https://www.irs.gov/uac/SOI-Tax-Stats-Individual-Income-Tax-Statistics-ZIP-Code-Data-(SOI), and the Zillow median house sales price data were obtained from http://www.zillow.com/research/data/#median-home-value.

4.2 Network position versus loan performance

Prediction 1. There is a positive relation between the loan performance of a firm and that of its neighbors.

Table 4 presents the results of regressing loan defaults, defined as loans that are 60 or more days delinquent at any time between their origination date and December 2013, on the physical geography and macroeconomic environment of the loan collateral, the performance of the originators, aggregators, and holding companies within the loan's network, the contractual characteristics of the loan, and the type of originator. The regression is intended to highlight the empirical relationship between network position and loan performance, controlling for a wide range of contractual, local economy, and regulatory effects that are typically found in default-estimation strategies (see, e.g., Adelino, Schoar, and Severino 2016; Mian and Sufi 2009).

As shown in Table 4, the default rate of a loan is significantly affected by the average mortgage default rate within the same ZIP code (excluding the sample loan), the default rates of contiguous ZIP codes and the local macroeconomy, as well as by the average default rate for the lenders, aggregators and holding companies within the loan's network. To see the importance of network effects relative to purely geographical effects, a 1% increase in the default rates of neighboring lenders within the network leads the subject loan's default rate to increase by 2.28%; a 1% increase in purely geographically determined local-market default rates leads to an increase in the subject loan's default rate of only 0.54%. Similarly, an increase of 1% in the default rates of the neighboring aggregators within the network leads the subject loan's default rate to increase by 1.15%; a 1% increase in the average default rates of contiguous ZIP codes leads to an increase of only 0.37%. Even for the more network-distant holding companies, a 1% increase in default rates leads to an increase of 0.84% in the subject loan's default rate.

All the other coefficients reported in Table 4 have the expected association with default. Interestingly, we find that ZIP codes with higher gross income levels and higher median sales prices are associated with elevated default levels.[18] Moreover, surprisingly, mortgage companies, banks and savings and loan institutions are associated with lower loan-level default rates relative to credit unions and finance companies, despite the presence of loans from several bankrupt firms that were mortgage companies and savings and loan institutions.

Although we include geographic controls, we cannot completely disentangle whether these effects are driven by network proximity or by some omitted factor that is correlated with network position. By studying the change of the network over time, we can rule out that such a factor captures static node characteristics. Specifically, we study the nodes that were neighbors of a node in 2005 but not in

[18] This result is consistent with those of Adelino, Schoar, and Severino (2016).

Table 4
Default regressions

	Coefficient estimate	SE	Coefficient estimate	SE	Coefficient estimate	SE
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	−1.8431***	0.199	0.0133	0.041	−2.2550***	0.164
Geographic effects						
Loan ZIP code average default (excluding subject loan)	0.5204***	0.0065	0.5193***	0.0006	0.5171***	0.0065
Weighted average default in contiguous ZIP codes	0.3591***	0.0059	0.3656***	0.006	0.3632***	0.0059
Average house prices in ZIP code (0000)	0.0988***	0.0034	0.1014***	0.003	0.1012***	0.0034
log of average IRS median gross income in ZIP code	0.0013***	0.0001	0.0014***	0.0003	0.0014***	0.0001
Network effects (excluding subject loan)						
2006 Neighboring network lenders average default rate	2.3958***	0.0635			1.7859***	0.0649
2006 Neighboring network aggregators average default rate	1.2717***	0.0497			1.1181**	0.2776
2006 neighboring network holding companies' average default rate	1.0441***	0.09			2.3367**	0.5777
2005 neighboring network lenders average default rate			0.3783	0.198	0.3335	0.276
2005 neighboring network aggregators average default rate			0.2978	0.157	0.0526	0.2444
2005 neighboring network holding companies' average default rate			−0.5128***	0.049	0.5164	0.5895
Contract structure						
Loan origination cost	0.0855***	0.013	0.0878***	0.0131	0.0988***	0.013
Original loan balance	−0.0007**	0.0002	−0.0006*	0.0003	−0.0006*	0.0002
Original cumulative loan-to-value ratio	0.6379***	0.0031	0.6369***	0.003	0.6369***	0.0031
Original mortgage contract rate	−0.0497*	0.0247	−0.0658**	0.025	−0.0562*	0.0247
Original mortgage amortization term	0.0047***	0.0001	0.0047***	0.0001	0.0047***	0.0001
FICO score	−0.0012***	0.0000	−0.0012***	0.0000	−0.0012***	0.0000
Conventional conforming loan indicator	−0.0059***	0.0016	−0.0060**	0.002	−0.0060***	0.0016
No document loan	0.0866***	0.001	0.0867***	0.001	0.0864***	0.0011
Lender type						
Bank indicator	−0.0425**	0.0161	−0.0236	0.0162	−0.0278	0.0163
Savings and loan indicator	−0.1021***	0.0161	−0.1008***	0.0161	−0.1009***	0.0161
Mortgage company indicator	−0.1019***	0.016	−0.0896***	0.0162	−0.0932***	0.016
Credit union indicator	0.0224	0.0163	0.0246	0.0163	0.0286	0.0163
Firm fixed effects	Yes		Yes		Yes	
Month fixed effects	Yes		Yes		Yes	
N =	969,317		969,317		969,317	

*** (≤.001)
** (≤.01)
* (≤.05)

Linear regression of loan-level default on the average default rate within the loan's ZIP code (excluding the subject loan) and the average default rate within the loan's network neighborhood (excluding the subject loan).

2006 (see Columns 4–7 of Table 4). The results indicate little relation between these 2005 neighbors and 2006 node performance, both when included alone (Columns 4 and 5) and when included jointly with the 2006 nodes (Columns 6 and 7).

Overall, these results suggest that the mortgage loan flow network captures an important factor through which heterogeneous risk exposure and underwriting quality affected loan performance. These network effects appear to be over and above the well known effects of local house price and income dynamics, loan contract characteristics and other local market conditions, and are broadly consistent with the implications of our model.

4.3 Network position versus firm performance

Prediction 2. There is a positive relation between the financial distress of a firm and that of its neighbors.

Above we consider how network position affects the performance of individual loans. Here we consider network position's effect on *firm* performance, in the process verifying that loan-level default performance is a good proxy for firm performance. To make sure that any effects we identify are not merely due to, say, geographical differences that are correlated with network structure, we include additional variables to control for the subject firm's location, local macro factors such as house prices and household income, and the type of firm. We estimate the following specification:

$$C_{i,2007} = \beta_0 + \beta_1 Zip_{LDi} + \beta_2 Network_{LDi} + \beta_3 Network_{FFi} + \sum_k \beta_k Macro\, factors_{ki,2006} + \sum_l \beta_l Firm\ types_{li,2006} + \epsilon_i, \qquad (18)$$

where $C_{i,2007}$ is the 2007 disappearance, due to closure or bankruptcy, of lenders who appeared in the 2006 network of mortgage originators, Zip_{LDi} is the weighted average contiguous ZIP code percentage of loan defaults over all the lender's geographic locations from 2000 to 2006, $Network_{LDi}$ is the weighted average network linked percentage of loan defaults over all the 2006 network locations of the lender during 2006, $Network_{FFi}$ is the percentage of firms among the lenders that were network linked to the subject lender in 2006 and were closed or went bankrupt in 2007 (excluding the subject firm). The macro factors include the level of house prices and the average IRS gross income across all the states where the firm is located. Firm-type dummies are included for commercial banks, savings and loan institutions, and independent mortgage companies.

Results are presented in Table 5. The primary finding is that the percentage of firm failures in 2007 among neighboring network lenders, $Network_{FFi}$, has a large effect on the failure of the subject lender in 2007, even after controlling for geography, loan performance, and macro factors. A one-percentage-point

Table 5
Linear probability closure estimates

	Coefficient estimate	SE
Intercept	−4.9829***	0.3305
Average loan defaults across contiguous ZIP codes and subject firm locations, 2000–2006 (Zip_{LDi})	0.1138***	0.01804
Average loan defaults across all 2006 network neighboring lenders to subject firm ($Network_{LDi}$)	0.2461***	0.0123
Percentage of firm failures in 2007 among 2006 neighboring network firms ($Network_{FFi}$)	0.4330***	0.0523
Log of average IRS household gross income across all subject firm state locations in 2006	−0.1149**	0.0462
Average house price across all subject firm state locations in 2006 (000)	−0.0023	0.0056
Subject firm is a bank	−0.1086***	0.0135
Subject firm is a savings and loan institution	0.0572	0.0446
Subject firms is an independent mortgage company	0.0400***	0.0122
Number of firms	11,095	
R^2	0.0700	

*** ($\leq$.001)
** ($\leq$.01)
* ($\leq$.05)

Linear probability estimates of the 2007 disappearance, due to closure or bankruptcy, of lenders who appeared in the 2006 network of mortgage originators. The covariates follow those outlined in Equation (18).

increase in the 2007 failure rate of neighboring network lenders would lead to a 0.43% increase in the subject firm's failure rate.

Geographically, both the local mortgage defaults, Zip_{LDi}, and the network-local defaults, $Network_{LDi}$, are positively associated with firm failure in 2007 at better than the .0001 level of statistical significance. This result suggests that our loan-level default measure of firm performance is a reasonable proxy for firm level performance even with direct controls for firm failures among network linked firms. Among the macro factors, higher average IRS household gross income leads to a statistically significant lower failure rate. Higher house prices in the state in which the subject firm is located do not have a statistically significant effect on firm failures in 2007. Finally, commercial banks were significantly less likely to be closed or become bankrupt in 2007, while independent mortgage companies were more likely. Being a savings and loan institution did not have a statistically significant effect on the likelihood of 2006 lenders disappearing from the sample.

4.4 Financial norms versus labor intensity

Prediction 3. The estimated quality measure in the model should be related to other, independently observable, measures of quality.

We estimate the network's quality vector and study how it relates to independent quality measures.[19] We use the number of links a node has as a proxy for its size, s^n. For each node, we compute the average of all costs of loans flowing through that node in a given year, and use this as a proxy for the cost c^n. We also use the measures of $\tilde{G}$ and E described in Section 4.1.

We use the method described in Appendix D to estimate the quality vector, q, and the parameters R_L, R_H, ΔR, d, and p, under the assumption that observations occur after $\bar{m}=4$ steps of the clearing mechanism. We adopt the visualization approach in Stanton, Walden, and Wallace (2014), in which the high- and low-default segments of the network are shown separately. Figure 6 presents the network with a low-default part (left panel) and a high-default part (right panel). As is clear from Figure 6, the high-default segment of the network constitutes a concentrated part.

In total, 2,521 estimated high-quality nodes make up 19% of the network. The aggregators and HCs are overrepresented among the high-quality nodes: 21 of 54 HCs are high quality, as are 1,071 out of 2,030 aggregators, whereas only 1,429 out of 11,095 lenders are high quality. There are 2,170 systemically pivotal nodes. A disproportionately high fraction of these, 66%, were high quality, compared with only 9.9% of the remaining 11,012 nodes. Thus, the additional incentives that pivotal nodes had to invest in quality were important for the outcome.

[19] We focus on financial norms. An alternative approach would be to study predictions on network structure, as is done in Kelly, Lustig, and Van Nieuwerburgh (2013). This provides an interesting avenue for future research.

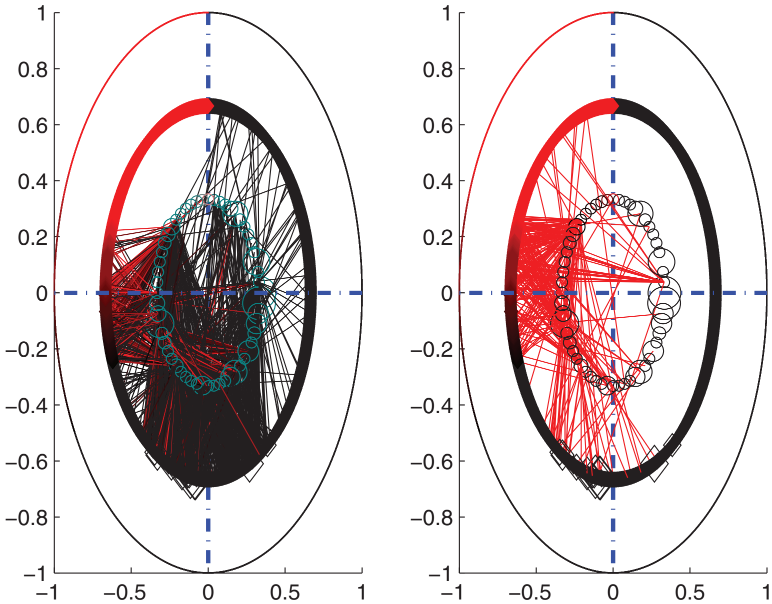

Figure 6
High- versus low-default parts of 2006 network
High- (left) and low-default (right) part of network, between originators and aggregators (above) and between aggregators and holding companies (below). The outer circle of nodes represents originators of record (11,095 in total), sorted clockwise in increasing order of quality of loans. The middle circle represents aggregators (2,030 in total) with the same ordering. The inner circle represents the holding companies (56 in total). The cutoff is made so that links with a higher than 80% default rate are shown in the right panel, whereas links with a lower default rate are shown in the left panel.

Within the mortgage lending industry, we argue that the main lever for firms to ensure high quality is to invest sufficient time on loan applications to filter out lemons. In our loan flow network we do not observe the time spent on loans, but we can see firms' employee profiles and therefore calculate a measure of labor intensity. We obtain our labor intensity measure for each firm in the sample using SAS and Python to merge by the firm name of each mortgage originator, aggregator, and holding company in our sample with firm names in another data set, called the National Establishment Time-Series (NETS) data. The NETS data provides a nearly-complete census of all U.S. establishments and includes information on the firm names, geographic location, number of employment positions at the location, and the eight digit NAICS number among other firm-level identifiers.[20] Our merge success rate between the NETS data and our loan-level mortgage data, using multiple passes of fuzzy merging technology, was quite high and we were able to accurately match employment data to

[20] The data are were obtained from Walls and Associates and they are constructed as a panel of annual snapshots of Dun and Bradstreet (D&B) credit rating data. Walls and Associates compiles the raw D&B snapshots into a continuous longitudinal data series. In our application, we use the 2005, 2006, and 2007 data panels.

98% of the mortgage holding companies, to 94% of the aggregators, and to 74% of the originators.[21] Each establishment in NETS contains information on the number of employment positions that exist at that establishment. An important advantage of the NETS data is that it provides a linkage between each establishment and the firm's headquarters establishment. Using these firm-family linkages, we are able to compute the total employment base for each of the firms in our sample of originators, aggregators, and holding companies. We use this labor intensity measure as an independent observation of quality and explore how it is related to our estimated quality vector, which we stress is not based on employment data.

Panel A of Table 6 shows how many firms are within the top 50–90 percentiles with respect to number of employees, which are also estimated to be high quality. We do this both for the aggregator group and for all firms (the results for the HQ group go in the same direction but the group is too small to obtain statistical significance). For example, there were 169 aggregators that were both within the 90th percentile with respect to number of employees and estimated to be high quality. This was 66% higher than 102, which would be the expected number if the two variables were independent, so the two variables are strongly positively related both economically and statistically, as are the results for the other percentiles and when calculated for all firms.

As previously discussed, quality is positively correlated with node size, and this is also the case for number of employees. To disentangle what is driving the positive relation with quality, we regress labor intensity (number of employees per loan) on quality and log-loan size in Panel B of Table 6. The coefficient is positive at the 5% significance level in the aggregator group, and strongly positively significant (t-statistic of 7.2) when all firms are used. The results are very similar when loan origination costs and default rates are included in the regressions. Thus, the documented relation between our estimated quality vector and the independently observed employment variables is indeed positive.

4.5 Predicted and realized default rates

Prediction 4. There is a positive relation between predicted and realized loan performance.

As mentioned above, our estimation of the market in 2006 is based on $\bar{m}=4$ steps in the clearing mechanism. By allowing shocks to propagate for one more period, we can generate (out-of-sample) predictions for the change in preinsolvency cash flows between 2006 and 2007,[22]

$$\Delta\tilde{G}=\tilde{G}_{\bar{m}=5}-\tilde{G}_{\bar{m}=4}.$$

[21] For some firms in our sample that were not in the NETS data, we also used public filings and internet searches to determine their number of employees.

[22] This assumes that the time period for such a step is one year, an assumption that we vary.

Table 6
Labor intensity versus quality

A.

	Aggregators realized overlap	Expected overlap	Excess overlap	All firms realized	Expected	Excess overlap
90%	169	102	66.0%***	450	225	100.0%***
80%	298	203	49.5%***	758	448	69.2%***
70%	412	304	35.6%***	1,023	702	45.8%***
60%	517	396	30.5***	1,215	936	29.8%***
50%	622	505	23.2***	1,359	1,129	20.4%***
N =	1,914			10,294		

B.

	Aggregators			All firms		
Quality	7.60*	7.75*	7.91*	2.28***	2.28***	2.27***
log(1+# loans)	−4.52***	−4.49***	−4.49***	−3.64***	−3.64***	−3.64***
Default rate		−3.32	−3.04		−0.85	−0.83
Loan orig. cost			−46.6			−2.15
N =	1,914			10,294		

Panel A compares the realized overlap of nodes (Columns 2 and 5) that are both high quality and in the top employment percentile shown in the first column, with the benchmark expected overlap if these variables were independent (Columns 3 and 6). The excess fraction of firms compared with the benchmark is shown in Columns 4 and 7. Columns 2–4 present for nodes among the aggregator group, and Columns 5–7 present for all nodes. Panel B shows regressions of labor intensity on quality, size, default rate and origination cost. Columns 2–4 are for aggregators, and Columns 5–7 are for all firms. Employment data are available for 94% of aggregators (1,917 out of 2,030) and 78% of all firms (10,294 out of 13,182). Labor intensity is winsorized at the 1% level, to control for outliers. (*), (**), and (***) represent significance at the 5%, 1%, and 0.1% level, respectively.

Given the assumptions of our model, this predicted change ought to be negatively related to realized changes in defaults over the period,

$$\Delta r^n = r^n_{2007} - r^n_{2006}.$$

We calculate these values for the 6,876 nodes that were present in the 2007 sample: 45 HCs, 687 aggregators, and 6,144 lenders (altogether slightly more than half of the 2006 sample of 13,182 nodes).

Results are shown in Table 7. Panel A shows OLS regressions of actual performance on predicted performance, quality choice, whether a node was systemically pivotal, and also on a node's size measured as the logarithm of the number of loans it handled. Columns 1–4 show the results for the full sample, and Columns 5–8 focus on lenders and Columns 9–12 on aggregators.[23] In line with our prediction, the coefficients are strongly significantly negative for the predicted performance coefficient, both in the univariate and multivariate regression. The coefficients are also significant for the aggregator subsample, whereas the coefficients have the right sign but are not significant for the lender subsample.

[23] Since only 46 HCs remained in 2007, the sample was too small for meaningful regressions at this level.

Table 7
Predicted and actual performance

A. Ordinary least squares

	1	2	3	4	5	6	7	8	9	10
	All									
Performance, ΔG	−0.550**				−0.508**					
Quality, q		0.018			0.010					
Syst. piv., $\mathcal{N}^S$			0.011		0.006					
Size, $\log(s)$				0.0006	0.0006					
	Lenders					Aggregators				
Performance, ΔG	−0.354				−0.402	−0.574*				−0.613*
Quality, q		0.0006			−0.006		0.064			0.079*
Syst. piv., $\mathcal{N}^S$			0.005		0.007			−0.003		−0.016
Size, $\log(s)$				−0.0007	−0.0007				0.006	0.006

B. Rank correlation

	1	2	3	4	5	6	7	8	9	10
	All									
Performance, ΔG	−0.044***				−0.034**					
Quality, q		131.4			52.9					
Syst. piv., $\mathcal{N}^S$			80.6		11.0					
Size, $\log(s)$				−0.012	−0.009					
	Lenders					Aggregators				
Performance, ΔG	−0.019				−0.020	−0.024				−0.033
Quality, q		11.1			−32.1		35.4			40.5
Syst. piv., $\mathcal{N}^S$			41.8		31.3			−0.202		−8.59
Size, $\log(s)$				−0.014	−0.015				0.042	0.049

The table displays realized changes of default rates of nodes between 2006 and 2007, regressed on predicted changes, estimated quality vector, whether nodes are systemically pivotal, and log-size measured as number of loans. Panel A shows ordinary least-squares regressions, and panel B shows rank correlation regressions. Predicted and actual performance are negatively related in all regressions, and the results are strongly significant in the full sample. Significance at the 5% (*), 1% (**), and 0.1% level (***) is shown. Estimation based on actual mortgage network in 2006.

Prediction 5. There is lower predicted performance of firms that defaulted than of those that remained in the sample.

The companies that dropped out of our sample between 2006 and 2007 mainly did so because they defaulted.[24] They were thus the companies that performed the worst during the time period, which ought to be correlated with the model's predicted performance.

Comparing the mean $\Delta\tilde{G}$ for the nodes that remained in the sample and those that dropped out, using a two-sample t-test, the mean is -0.0036 for the remaining nodes, and -0.025 for those that dropped out. The difference is strongly statistically significant, with a t-statistic of 23.3.

For robustness, we carry out several variations of the tests. Panel B of Table 7 shows that the results are similar when using univariate and multivariate rank correlation regressions, which are more robust to nonlinearities. The choices of using $\bar{m}=4$ steps for the clearing mechanism in the estimation, and 5 steps for the predicted 2007 performance, are also somewhat arbitrary. The results are very similar when we vary the steps between 3 and 6, and also under multivariate regressions on an arbitrary subset of the variables in the table (not reported).

Prediction 6. Ex ante effects, that is, financial norms, should be important for the model's ability to predict future defaults.

The model without financial norms is nested, in that when $\Delta R=0$ there is no benefit to firms from investing in quality, and they will therefore all choose to be low quality. This allows us to study the importance of allowing for heterogeneous financial norms for our results. Specifically, by forcing $\Delta R=0$, we turn off the financial norm component of the model. We estimate the network and predicted performance for this special case. Regardless of the number of steps in the clearing mechanism, $\bar{m}\in\{3,4,5,6\}$, we find no relationship between predicted and realized default rates. Thus, the financial norms are indeed important for the results.

5. Conclusions

In our network model of intermediaries, heterogeneous financial norms and performance, as well as systemic vulnerabilities, arise as equilibrium outcomes. The optimal behavior of each intermediary, in terms of its attitude toward risk, the quality of the projects it undertakes, and the intermediaries it

[24] Of the eleven holding companies that no longer appear in the 2007 sample, 82.82% of those firms went bankrupt, were closed, or were acquired. Of the twenty two aggregators that no longer appear in the 2007 sample, 72.72% went bankrupt, were closed, or were acquired. Among the top 50 lenders that no longer appear in the 2007 sample, 25% went bankrupt, were closed or were acquired. Among all of these firms, bankruptcy was the primary cause of exit from the sample.

chooses to interact with, are influenced by the behavior of its prospective counterparties. These network effects, together with the intrinsic differences between intermediaries, jointly determine financial strength, quality, and systemic vulnerability, at the aggregate level of the market, as well as for individual intermediaries.

We apply the model to the mortgage origination and securitization network of financial intermediaries, using a large data set of more than three million mortgages originated and securitized through the private-label market between 2005 and 2007. As predicted, the network position of an intermediary is strongly related to the default rates of its loans, above and beyond geographical and other observable factors. In addition, firm performance is related to the performance of neighboring firms in the network, even after controlling for the performance of loans issued by those neighboring lenders. Moreover, an independent quality observation—labor intensity—is positively related to our estimated quality. Finally, the model predicts future default rates, though this predictive ability goes away if we turn off the financial norms component of the model, emphasizing the joint importance of both ex ante and ex post components of the model.

Altogether, our results show the importance of network effects in the U.S. mortgage market, suggesting that our equilibrium framework for jointly studying network structure and financial norms is fruitful in this market.

Appendix A. Minimum Spanning Tree

Figure A1 shows the minimum spanning trees representation of the subsample of the 2006 loan flow mortgage network discussed in Section 1.

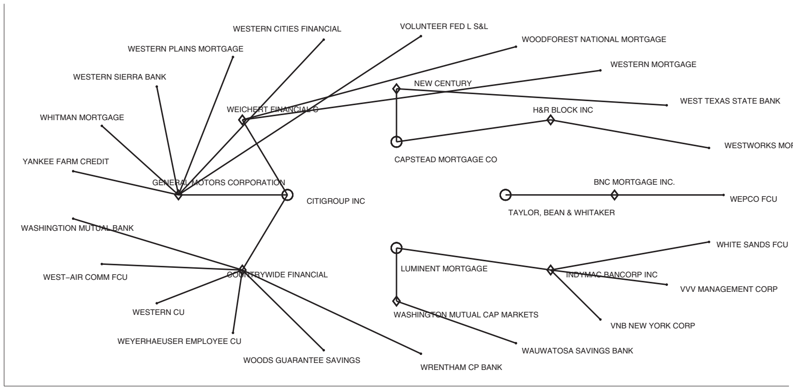

Figure A1
Trees representation of a subsample of private-label mortgage originations in 2006

Appendix B. Clearing Algorithm for General Network Model

As explained in the body of the paper, the initial assumption in the clearing mechanism is that all nodes remain solvent. If this is a feasible outcome, then it is the outcome of the clearing mechanism. If, however, some nodes fall below the insolvency threshold, then the new cash flows are calculated, given the insolvencies in the first step and new insolvencies are checked for. The algorithm terminates when no new insolvencies occurs, which takes at most N steps.

Algorithm 1 (Clearing mechanism).

1. Set iteration $m=0$ and the initial solvency vector $f_1=\mathbf{1}$.
2. Repeat:
   - Set $m=m+1$.
   - Calculate $z_m=\Lambda_{f_m}\Pi\Lambda_{f_m}\tilde{P}$.
   - Calculate $f_{m+1}^n=X\left(\frac{z_M^n}{s^n}\right)$, $n=1,\ldots,N$.
3. Until $f_{m+1}=f_m$.
4. Calculate the $t=1$ cash flow as $\tilde{F}=z_m$.

The iteration over M can be viewed as showing the gradual propagation of insolvencies, where f_M-f_{M-1} shows the insolvencies triggered in step M by the insolvencies that occurred in step $M-1$.

We can write the algorithm using returns, defining $R=\Lambda_s^{-1}\tilde{C}F$, and $\Gamma=\Lambda_s^{-1}\hat{\Pi}$. The algorithm then becomes even simpler:

Algorithm 2 (Clearing mechanism on return form).

1. Set iteration $m=0$ and the initial solvency vector $f_1=\mathbf{1}$.
2. Repeat:
   - Set $m=m+1$.
   - Calculate $z_m=\Lambda_{f_m}\Gamma\Lambda_{f_m}R$.
   - Calculate $f_{m+1}=X(z_m)$.
3. Until $f_{m+1}=f_m$.
4. Calculate the $t=1$ cash flow as $\tilde{F}^n=\Lambda_s z_m$.

It is straightforward to show that the two algorithms are equivalent, since $\hat{\Pi}=\Lambda_s\Gamma=\Pi\Lambda_s$.

The $\bar{m}$ step version of the clearing mechanism is obtained by replacing condition in step 3 of Algorithm 2 by the terminal condition *"Until $m=\bar{m}$,"* leading to $\tilde{F}_{\bar{m}}=\mathcal{CM}\left[\tilde{P}|\bar{m}\right]$.

The preinsolvency m-step cash flows are defined as $\tilde{G}_{\bar{m}}=\mathcal{CM}_{Pre}\left[\tilde{P}|\bar{m}\right]=(\Pi\Lambda_{f_{\bar{m}}})\times\tilde{P}$, which via (12) leads to the relation $\tilde{F}_{\bar{m}}=\Lambda_{f_{\bar{m}}}\mathcal{CM}_{Pre}\left[\tilde{P}|\bar{m}\right]$.

Appendix C. Network Formation Game

The sequence of events is the same as that in Figure 3. We use subgame perfect, pairwise-stable Nash as the equilibrium concept. We make one extension of the pairwise stability concept in the definition of agents' action space. Specifically, agents are allowed to unilaterally decide to become completely isolated by severing links to all agents they are connected to. In contrast, with the standard definition of pairwise stability agents are only allowed to sever exactly one link or propose the addition of one link. The assumption that agents can choose to become isolated can thus be viewed as a network participation constraint.

The sequence of events is as follows: at $t = -2$ (the proposal/severance stage of the game) there is a given initial network, $E \subset \bar{E}$. Recall that $\bar{E}$ here is the maximal network in the economy, which arises if all possible links are present. Each agent, $n = 1, \ldots, N$, simultaneously chooses from the following mutually exclusive set of actions:

1. Sever links to all other agents and become completely isolated.
2. Sever exactly one existing link to another agent, $(n, n') \in E$.
3. Propose the formation of a new link to (exactly) one other agent, $(n, n') \in \bar{E} \backslash E$.
4. Do nothing.

In contrast to actions 1 and 2, which are unilateral, agent n' needs to agree for the links to actually be added to the network under action 3.

The set of networks that can potentially arise from this process is denoted by $\mathcal{E}$. We note that $E' \subset \bar{E}$ for all $E' \in \mathcal{E}$. The set of actual proposals for addition of links, generated by action 3, is denoted by L^A. The set of links that are actually severed, generated by actions 1 and 2, is L^S. The total set of potential link modifications is $\mathcal{L} = \{(L^A, L^S)\}$.

At $t = -1$ (the acceptance/decline stage) L^A and L^S are revealed to all agents, who then simultaneously choose whether to accept or decline proposed links. Formally, for each proposed link, $\ell = (n, n') \in L^A$, agent n' chooses an action $a_\ell \in \{D, A\}$ (representing the actions of Declining or Accepting the proposed link). The total set of actions is then $A = \{a_\ell : \ell \in L_A\} \in \{D, A\}^{L_A}$, which for each n' can be decomposed into $A^{n'} \cup A^{-n'}$, where $A^{n'} = \{(n, n') \in L_A\}$ represents the actions taken by agent n', and $A^{-n'} = A \backslash A^{n'}$ the actions taken by all other agents. Altogether, E, L_A, L_D, and A then determines the resultant network, E', after the first two stages of the game, at $t = 0$.

At $t = 0$ (the quality choice stage) each agent, $n = 1, \ldots, N$, simultaneously chooses the quality $q^n \in \{0, 1\}$. The joint quality actions of all agents are summarized in the action vector $q \in \{0, 1\}^N$.

At $t = 1$, shocks, ξ, are realized, leading to realized net cash flows $w^n(\xi | q, E')$ as defined by equations (11–12). The value of intermediary n at $t = 0$ is thus

$$V^n(q | E') = \sum_{\xi \in \Omega} w^n(\xi | E', q) \mathbb{P}(\xi), \qquad n = 1, \ldots, N. \tag{C1}$$

C.1 Equilibrium

A stable equilibrium to the network formation game is an initial network and quality strategies, together with a set of beliefs about agent actions for other feasible network structures, such that no agent has an incentive to add or sever links, given that no other agents do so, and agents' have consistent beliefs about each others' behavior on and off the equilibrium path.

Specifically, the action-network pair (q, E), together with $t = -1$ acceptance strategies: $A : \mathcal{L} \to \{D, A\}^{L_A}$, and $t = 0$ quality strategies $Q : \mathcal{E} \to \{0, 1\}^N$ constitute an equilibrium, if

1. At $t = 0$, strategies are consistent in that for each $E' \in \mathcal{E}$, and $q = Q(E')$,

   $$q^n \in \operatorname*{arg\,max}_{x \in \{0,1\}} V^n((x, q^{-n}) | E),$$

   for all $n = 1, \ldots, N$, that is, it is optimal for each agent, n, to choose strategy q^n, given that the other agents choose q^{-n}.
2. At $t = -1$, strategies are consistent in that for each $(L_A, L_D) \in \mathcal{L}$, $A^n(L_A, L_D)$ is the optimal action for each agent, n, given that the other agents choose $A^{-n}(L_A, L_D)$. Optimal here, means that the action maximizes V^n at $t = 0$.
3. At $t = -2$, the strategy $L = \emptyset$ is consistent, that is, for each agent n, given that no other agent severs links or proposes additional links, it is optimal for agent n not to do so either, given the value such actions would lead to at $t = -1$.

A stable equilibrium is said to be maximal if $E = \bar{E}$. We note that since we focus on pure strategies, the existence of stable equilibrium is not guaranteed, which has not been an issue in the examples we have studied. Neither is uniqueness of stable equilibrium guaranteed.

It follows that the following conditions are necessary and sufficient for there to exist acceptance and quality strategies such that (q, E) is an equilibrium:

1. For all n, $V^n(q|E) \geq V^n((\neg q^n, q^{-n})|E)$.
2. For all n, $V^n(q|E) \geq V_I^n$.
3. For all $(n,n') \in E$, $\exists q' \in \{L,H\}^N$ such that
   - $V^n(q|E) \geq V^n(q'|E-(n,n'))$,
   - For all n'', $V^{n''}(q'|E-(n,n')) \geq V^{n''}((\neg(q')^{n''}, (q')^{-n''})|E-(n,n'))$.
4. For all $(n,n') \in \bar{E} \backslash E$, $\exists q' \in \{L,H\}^N$ such that
   - $V^n(q|E) \geq V^n(q'|E+(n,n'))$ or $V^{n'}(q|E) \geq V^{n'}(q'|E+(n,n'))$,
   - For all n'', $V^{n''}(q'|E+(n,n')) \geq V^{n''}((\neg(q')^{n''}, (q')^{-n''})|E+(n,n'))$.

The first condition ensures that each agent makes the optimal quality choice at $t=0$, given that no change to the network is made. The second condition ensures that it is not optimal for any agent to sever all links and become isolated. The third condition ensures that there are consistent beliefs about future actions, such that no agent has an incentive to sever a link at $t=-2$. The fourth condition ensures that there are consistent beliefs about future actions, so that no two agents can be made jointly better off by adding a link.

Appendix D. Computation of Equilibrium

An important property of our approach is that it can be applied to large scale real-world networks. Assume that the network, E, size vector, s, and the cost vector, c are observable, but that the parameter values $\phi = (R_L, R_H, \Delta R, d, p)$, and the quality vector $q \in \{0,1\}^N$ are not. The project cash-flows (3), given a state realization ξ, are in vector form

$$\tilde{P}(\xi, q) = \Lambda_s(R_H \Lambda_\xi \mathbf{1} + \Lambda_{\mathbf{1}-\xi}(R_L \mathbf{1} + \Delta R q)).$$

We assume that preinsolvency cash flows, $\tilde{G}$, including noise, are observed after $\bar{m}$ steps of the clearing mechanism,

$$\hat{w} = \mathcal{CM}_{Pre}\left[\widetilde{CF}_P|\bar{m}\right] - \Lambda_c \Lambda_s q + \epsilon, \tag{D1}$$

where ϵ is a vector of independent, identically normally distributed noise.

Given an equilibrium quality vector, q, we define the best response for agents (at time 0), given the actions of other agents,

$$\mathcal{F}(q) = \mathcal{F}(q|\phi) = \{\hat{q} : \hat{q}^n \in \operatorname*{argmax}_{x \in \{0,1\}} V^n(x, \hat{q}^{-n})\},$$

where we as before assume that indifferent agents choose low quality, so that $\mathcal{F}$ is a single-valued function. For q to be consistent with equilibrium, it must be that $q = \mathcal{F}(q)$. We define the equilibrium set $\mathbb{Q} \subset \{0,1\}^N$, as

$$\mathbb{Q}(\phi) = \{q : q = \mathcal{F}(q)\}.$$

An estimate of the unobservable quality vector, q, and thereby of all other properties of the equilibrium network, is given by solving the problem:

$$\min_{\phi} \min_{q \in \mathbb{Q}(\phi)} \min_{\xi \in \Omega^1} \left|\left| \hat{w} - \left(\mathcal{CM}_{Pre}\left[\widetilde{CF}_P(\xi, q)\middle|\bar{m}\right] - \Lambda_s \Lambda_c q\right)\right|\right|. \tag{D2}$$

Here, the mean squared norm is used, and it then follows that (D2) is the maximum likelihood estimator of the equilibrium, conditioned on there being a shock, that is, conditioned on $\xi \neq \mathbf{1}$.

The equilibrium behavior of nodes in $\mathcal{N}^U$ is characterized by Proposition 6. For nodes in $\mathcal{N}^S$, which in our application to the U.S. mortgage market make up a small fraction (less than 20%) of the network, the characterization is not as simple. For each node in $\mathcal{N}^S$, the value of being high quality needs to be compared with the value of being low quality, leading to two full evaluations of the clearing mechanism for each such node. Importantly, however, it follows easily that the optimal quality choice of a node does not depend on the quality choices made by other nodes. Thus, rather than iterating over all possible combinations of quality choices of the nodes in $\mathcal{N}^S$, each node's behavior can be determined on its own, allowing for efficient calculation of the equilibrium set $\mathbb{Q}(\phi)$.

Also note that although the observed network structure, size vector and cost vector are high-dimensional in a market like the U.S. residential mortgage market with thousands of intermediaries, the model is actually strongly specified in that only the five parameters in ϕ are free in the estimation (D2).

Appendix E. Proofs

Proof of Proposition 1. The value if the agent invests in quality is

$$V_A = s^n((1-p)R_H + pY(R_L+\Delta R)-c),$$

versus

$$V_B = s^n(1-p)R_H$$

if not investing. It follows that $V_A > V_B$ if and only if

$$pY(R_L+\Delta R)-c>0,$$

for which $R_L+\Delta R>d$ and $p(R_L+\Delta R)>c$ are, in turn, necessary and sufficient conditions. *QED*.

Proof of Proposition 2. It immediately follows from Proposition 1 that in any isolated equilibrium, agents will choose the same quality $q^1 = q^2$. We therefore focus on the risk-sharing equilibrium. Without loss of generality, assume that intermediary 1 is high quality and intermediary 2 is low quality in equilibrium. The value of intermediary 1 is then

$$\begin{aligned} V_A = &(1-2p-p_2)R_H + pY((1-\pi)(R_L+\Delta R)+\pi R_H) \\ &+pY((1-\pi)R_H+\pi R_L)+p_2Y((1-\pi)(R_L+\Delta R)+\pi R_L)-c. \end{aligned} \tag{E1}$$

For it to be an equilibrium strategy for agent 1 to be high quality, it must be the case that this value weakly dominates that of being isolated and choosing high quality, which is

$$V_B = (1-2p-p_2)R_H + pY(R_L+\Delta R)+pY(R_H)+p_2Y(R_L+\Delta R)-c, \tag{E2}$$

and also that it strictly dominates the value of risk sharing and choosing low quality, which is

$$\begin{aligned} V_C = &(1-2p-p_2)R_H + pY((1-\pi)R_L+\pi R_H) \\ &+pY((1-\pi)R_H+\pi R_L)+p_2Y(R_L). \end{aligned} \tag{E3}$$

that is, $V_A \geq V_B$ and $V_A > V_C$. In words, these conditions ensure that the diversification benefits of sharing risk for the high-quality agent outweighs his cost of giving away the payoffs of a high-quality project for the payoffs of a low-quality project, and also the cost of investing into being high quality. It is easy to show that the first inequality is satisfied if and only if

$$R_L+\Delta R \leq d < \min((1-\pi)R_H+\pi R_L, (1-\pi)(R_L+\Delta R)+\pi R_H), \tag{E4}$$

and that the second inequality requires that $p(1-\pi)\Delta R > c$ in this case.

The value for the low-quality agent if switching to high quality while still sharing risk is

$$V_D = (1-2p-p_2)R_H + pY(\pi(R_L+\Delta R)+(1-\pi)R_H)$$
$$+pY(\pi R_H+(1-\pi)(R_L+\Delta R))+p_2Y(R_L+\Delta R)-c, \tag{E5}$$

which, given the restriction (E4), is equal to

$$V_D = (1-2p-p_2)R_H + p(\pi(R_L+\Delta R)+(1-\pi)R_H)$$
$$+p(\pi R_H+(1-\pi)(R_L+\Delta R))-c, \tag{E6}$$

versus

$$V_E = (1-2p-p_2)R_H + p(\pi(R_L+\Delta R)+(1-\pi)R_H)+pY(\pi R_H+(1-\pi)R_L), \tag{E7}$$

if remaining of low quality while still sharing risk. For the case in which $\pi R_H+(1-\pi)R_L > d$, it immediately follows that $p(1-\pi)\Delta R > c$ is necessary and sufficient for agent 2 to switch to high quality, that is, for $V_D > V_E$. For the case in which $\pi R_H+(1-\pi)R_L \le d$, it immediately follows that $p(1-\pi)\Delta R > c$ is sufficient for agent 2 to switch.

Thus, $q^1=1$, $q^2=0$ is ruled out as a risk-sharing equilibrium, since regardless of parameter values it is optimal for at least one of agent 1 and 2 to deviate. Note that if we relax the constraint $V_A > V_B$, corresponding to not allowing agent 1 to cut the link to agent 2, it is easy to construct examples where $q^1=1$ and $q^2=0$. The proposition therefore does not in general hold if the network is exogenously given. *QED*

Proof of Proposition 3. Given a maximal network with no quality investments, the value of each node is

$$V = (1-p_0)R_H + p_0\left(\frac{N-1}{N}R_H+\frac{1}{N}R_L\right) = R_H - \frac{p_0}{N}(R_H-R_L).$$

We first note that, given that the network E^* is sustainable, no node has an incentive to be high quality since the cost of investing in quality is $\frac{c_0}{N}$, whereas the expected benefit is only $\frac{p_0}{N}\times\frac{\Delta R}{N}$, which is dominated for large N. Thus, only $\mathbf{q}=\mathbf{0}$ may be an equilibrium quality vector for the maximal network.

It follows that no node in the low-quality maximal network has an incentive to sever a link, since the resultant network will again be one of low quality (N is large), with the resultant value for the node that severed the link being the same as in the maximal network:

$$V = (1-p_0)R_H + \frac{p_0}{N}R_H + \frac{p_0}{N}\left(\frac{N-2}{N}R_H+\frac{2}{N}R_L\right) + (N-2)\frac{p_0}{N}\left(\frac{N-1}{N}R_H+\frac{1}{N}R_L\right)$$
$$= R_H - \frac{p_0}{N}(R_H-R_L).$$

The only incentive a node may have to deviate from the low-quality complete network is thus to completely isolate.

If either of the conditions in (a) is satisfied, it follows from Proposition 1 that the isolated outcome is low quality, with value $V^I = \left(1-\frac{p_0}{N}\right)R_H < V$, and the isolated outcome is therefore dominated by the low-quality maximal network for all nodes. Thus, in this case, the low-quality maximal network is indeed an equilibrium.

If both conditions in (a) fail, it follows from Proposition 1 that the isolated outcome is one high quality, with resultant value

$$V^I = \left(1-\frac{p_0}{N}\right)R_H + \frac{p_0}{N}(R_L+\Delta R) - \frac{c_0}{N} = V + \frac{1}{N}(p_0\Delta R - c_0) > V,$$

so the low-quality maximal network is not an equilibrium network in this case. We are done.

Proof of Proposition 4. (i) First note that for large N it will never be the case that a low shock realization for a peripheral node causes the hub to default, since peripheral nodes are relatively small. Peripheral nodes are thus insulated from shocks to other peripheral nodes in the star network, regardless of the quality decisions made by nodes.

We first study the case in which $(R_L+R_H)/2>d$ and $\Delta R/2>c_0/p_0$. Given a high-quality star network, the value for a peripheral node is:

$$V=\left(1-2\frac{p_0}{N}\right)R_H+2\frac{p_0}{N}\times\frac{R_L+\Delta R+R_H}{2}-\frac{c_0}{N}=R_H-\frac{p_0}{N}(R_H-R_L-\Delta R)-\frac{c_0}{N}.$$

A peripheral node's payoff, if switching to low quality, is

$$\begin{aligned}V'&=\left(1-2\frac{p_0}{N}\right)R_H+\frac{p_0}{N}\times\frac{R_L+\Delta R+R_H}{2}+\frac{p_0}{N}\times\frac{R_L+R_H}{2}\\&=V-\frac{1}{N}(p_0\Delta R/2-c_0)\\&<V,\end{aligned}$$

so no peripheral node has an incentive to switch to low quality. Moreover, no peripheral node has an incentive to isolate, since isolation with low quality gives a value of

$$V''=\left(1-\frac{p_0}{N}\right)R_H<V$$

and isolation with high quality gives a value of

$$V'''=\left(1-\frac{p_0}{N}\right)R_H+\frac{p_0}{N}X(R_L+\Delta R)-\frac{c_0}{N}\leq V.$$

Finally, it is easy to verify that the value of two peripheral nodes that form a link and stay high quality remains V, whereas at least one of them has strictly lower value if either of them switches to low quality, so no two nodes can be found that have an incentive to form such a link. Thus, no peripheral node has an incentive to deviate from in a high-quality star equilibrium network.

The situation is very similar for the hub-node, which with the high-quality star network strategy has a value of $(N-1)\times V$. The values of the alternative strategies—switching quality, isolating and being of low and high quality—are also $N-1$ times higher, $(N-1)\times V'$, $(N-1)\times V''$, and $(N-1)\times V'''$, respectively. leading to the same conclusion as for the peripheral nodes. Finally, the value remains $(N-1)\times V$ if the hub severs one link, so it has no incentive to do so.

The argument is also very similar in the case in which $(R_L+R_H)/2\leq d<(R_L+\Delta R+R_H)/2$ and $R_L+\Delta R>c_0/p_0$. In this case, the value for a peripheral node of being high quality remains V, whereas it is $V-\frac{1}{N}(p_0(R_L+\Delta R)-c_0)<V$ if the node switches to low quality. Also, the values of isolation remain the same, $V''<V$ and $V'''\leq V$, respectively, and the node therefore has is no incentive to isolate. An identical argument as before implies that neither is there an incentive for two peripheral nodes to form a link. Finally, the argument for the hub-node is also identical to the previous case. We have shown (a).

For (b), an argument similar to that made in (a) shows that neither peripheral nodes, nor the hub, have incentives to switch from low to high quality. It also follows from Proposition 1 that $V^I=\left(1-\frac{p_0}{N}\right)R_H<V$, so the nodes have no incentives to isolate either. Moreover, it is easy to check that the hub is no better off after severing a link to a peripheral node. Finally, the same argument as before implies that two peripheral nodes have no incentive to form a link. Therefore, (ii) follows.

For (c), we denote the peripheral node n's value of being high quality in a star network by V^H, and of low quality by V^L, given the quality decisions of the other nodes, $\mathbf{q}^{-n}$. It is easy to verify that regardless of $\mathbf{q}^{-n}$,

$$\Delta V\stackrel{\text{def}}{=}V^H-V^L=\frac{p_0}{N}\left(X\left(\frac{R_L+R_H+\Delta R}{2}\right)-X\left(\frac{R_L+R_H}{2}\right)\right)-\frac{c_0}{N}.$$

If $\Delta V > 0$, all peripheral nodes therefore choose to be high quality, and otherwise all peripheral nodes choose to be low quality. Finally, the value difference between high and low quality for the hub is $(N-1)\times\Delta V$, and thus the hub chooses the same quality as the peripheral nodes. We are done.

Proof of Proposition 5. Note that Proposition 4 covers the special case in which $\alpha=1$. The main difference between Proposition 4 and Proposition 5 is that when $\alpha \neq 1$ the risk-sharing between the hub and peripheral nodes is no longer balanced. If $\alpha < 1$, the hub still shares half its risk in total, but each peripheral nodes shares less than half of its risk. If $\alpha > 1$, the situation is reversed and the hub shares less than half of its risk in total whereas the peripheral nodes each share half of their risk.

For (a), similar parameter restrictions as for (a) and (b) of Proposition 4 allows the construction of high-quality and low-quality equilibrium star networks. Specifically, regardless of α, if c_0 is sufficiently high and $(1-x)R_L+xR_H > d$, where $x=\frac{1}{2}\min(\alpha,1/\alpha)$, an argument similar to that made in Proposition 4 (a) implies that E^S with $q=\mathbf{1}$ is an equilibrium. Similarly, if c_0 is sufficiently low and $(1-x)R_L+xR_H > d$, an argument similar to that made in Proposition 4 (b) implies that E^S with $q=\mathbf{0}$ is an equilibrium.

For (b), note that when $\alpha < 1$, the hub shares half of its risk, whereas the peripheral nodes share less than half of their risk. The hub therefore has less of an incentive to choose high quality than a peripheral node, given the star network, E^S. Specifically, the peripheral node chooses high quality if $(1-\alpha/2)\Delta R > c_0/p_0$, whereas the hub chooses high quality if $\Delta R/2 > c_0/p_0$. It is therefore possible to construct an equilibrium with high-quality peripheral nodes and a low-quality hub, by choosing $\frac{c_0}{2p_0\Delta R}\in(1,2-\alpha)$. For example, it is straightforward to verify that this is an equilibrium outcome for the economy with $c_0=1/2$, $p_0=1/2$, $\Delta R=\frac{4}{3-\alpha}$, $R_L=1/2$, $R_H=6$, and $d=3$. Now, for $\alpha \geq 1$, the hub shares less than half of its risk whereas peripheral nodes share half of theirs, so the hub is high quality whenever the peripheral nodes are, due to a similar incentive argument as before.

For (c), an argument similar to that made in (b) implies that only if $\alpha > 1$ may it be possible to construct an equilibrium star network such that the hub is high quality and the peripheral nodes of low quality. Indeed, given the star network E^S, $(1-1/(2\alpha))\Delta R > c_0/p_0 > \Delta R/2$ is necessary and sufficient for such quality choices. However, in contrast to peripheral nodes, the hub may choose to sever a link to one of its neighbors. It is easy to check that for large N, this will always be optimal—and thus E^S will not be an equilibrium network—since severing a link allows the hub to still avoid default in case of a low shock, while keeping a higher fraction of its (high-quality) project than under E^S. Thus, a high-quality hub, low-quality peripheral node, star network equilibrium outcome is not feasible. *QED*

Proof of Proposition 6. If node n chooses high quality, its value is

$$s^n(pY(\Gamma_{nn}(R_L+\Delta R)+R_H\Theta^n)+(1-p)V)-c^n,$$

whereas if it chooses low quality, its value

$$s^n(pY(\Gamma_{nn}R_L+R_H\Theta^n)+(1-p)V).$$

Here V is the expected value of cash flows conditioned on $\xi^n=1$, which is the same regardless of q^n, since node n's project cash flows are the same (R_H) regardless of q^n if $\xi^n=1$. So, node n will choose to be high quality, if and only if

$$pY(\Gamma_{nn}(R_L+\Delta R)+R_H\Theta^n)-c^n > pY(\Gamma_{nn}R_L+R_H\Theta^n), \tag{E8}$$

and it is easy to verify that the requirement that either condition (1) or (2) in the proposition holds is necessary and sufficient for (E8). *QED*

Proof of Proposition 7. If node n chooses high quality, its value is

$$s^n\left(pY(\Gamma_{nn}(R_L+\Delta R)+R_H\Theta^{n,1})+(1-p)V\right)-c^n,$$

whereas if it chooses low quality, its value

$$s^n\left(pY(\Gamma_{nn}R_L+R_H\Theta^{n,1})+(1-p)V\right).$$

Here, $\Theta^{n,1}>\Theta^{n,0}$ since $n\in\mathcal{N}^S$, and an identical argument as that in the proof of Proposition 6 shows that conditions (1) or (2) are sufficient for the value of being high quality to outweigh that of being of low quality. *QED*

Proof of Proposition 8. The result is immediate, since any node that is not systemically pivotal never affects whether any of its neighbors becomes insolvent, and therefore never affects the cash flows to any of its neighbors' neighbors. *QED*

Proof of Proposition 9. By the definition of the sharing rule (8,9), it follows that $\Gamma_{jj}\geq\Gamma_{nj}$ for all n. Therefore, a well- diversified connection of node n, j, will have $\Gamma_{nj}\leq\Gamma_{jj}<\kappa$, which it is easy to verify implies that it survives a shock initiated by node n, regardless of node n's quality choice. Since all of node n's other connections are also well diversified, there will be no negative cash flow spillovers from other nodes, so neither of n's connections therefore default. Thus, $n\in\mathcal{N}^U$. *QED*

References

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi. 2015. Systemic risk and stability in financial networks. *American Economic Review* 105:564–608.

Adelino, M., A. Schoar, and F. Severino. 2016. Loan origination and defaults in the mortgage crisis: The role of the middle class. *Review of Financial Studies* 29:1635–70.

Agarwal, S., D. Lucca, A. Seru, and F. Trebbi. 2014. Inconsistent regulators: Evidence from banking. *Quarterly Journal of Economics* 129:889–938.

Allen, F., A. Babus, and E. Carletti. 2012. Asset commonality, debt maturity and systemic risk. *Journal of Financial Economics* 104:519–34.

Allen, F., and D. Gale. 2000. Financial contagion. *Journal of Political Economy* 108:1–33.

Babus, A. 2016. The formation of financial networks. *RAND Journal of Economics* 47:239–72.

Babus, A., and P. Kondor. 2013. Trading and information diffusion in over-the-counter markets. Working Paper, Federal Reserve Bank of Chicago.

Barabasi, A.-L., and R. Albert. 1999. Emergence of scaling in random networks. *Science* 286:509–512.

Berger, A. N., and D. B. Humphrey. 1997. Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research* 98:175–212.

Bernanke, B. 2007. Global imbalances: Recent developments and prospects. Bundesbank Lecture, Berlin.

Billett, M. T., M. J. Flannery, and J. A. Garfinkel. 1995. The effect of lender identity on a borrowing firm's equity return. *Journal of Finance* 50:699–718.

Cabrales, A., P. Gottardi, and F. Vega-Redondo. 2014. Risk-sharing and contagion in networks. Working Paper, Universitat Pompeu Fabra.

Chang, B., and S. Zhang. 2015. Endogenous market making and network formation. Working Paper, London School of Economics.

Cheng, I.-H., S. Raina, and W. Xiong. 2014. Wall Street and the housing bubble. *American Economic Review* 104:2797–829.

Demyanyk, Y., and O. Van Hemert. 2011. Understanding the subprime mortgage crisis. *Review of Financial Studies* 24:1848–80.

Di Maggio, M., and A. Kermani. 2014. Credit-induced boom and bust. Working Paper, U. C. Berkeley.

Di Maggio, M., and A. Tahbaz-Salehi. 2014. Financial intermediation networks. Working Paper, Columbia University.

Eisenberg, L., and T. H. Noe. 2001. Systemic risk in financial systems. *Management Science* 47:236–49.

Elliott, M., B. Golub, and M. O. Jackson. 2014. Financial networks and contagion. *American Economic Review* 104:3115–53.

Friedkin, N. E., and E. C. Johnsen. 1999. Social influence networks and opinion change. *Advances in Group Processes* 16:1–29.

Froot, K. A., D. Scharfstein, and J. C. Stein. 1993. Risk management: Coordinating corporate investment and financing decisions. *Journal of Finance* 48:1629–58.

Glaeser, E. L., and C. G. Nathanson. 2017. An extrapolative model of house price dynamics. *Journal of Financial Economics*. Advance Access published June 24, 2017, 10.1016/j.jfineco.2017.06.012.

Glasserman, P., and H. P. Young. 2015. How likely is contagion in financial networks? *Journal of Banking and Finance* 50:383–99.

Gofman, M. 2011. A network-based analysis of over-the-counter markets. Working Paper, University of Wisconsin, Madison.

Granovetter, M. 1985. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91:481–510.

Jackson, M. O., and D. López-Pintado. 2013. Diffusion in networks with heterogeneous agents and homophily. *Network Science* 1:49–67.

Jackson, M. O., and A. Wolinsky. 1996. A strategic model of social and economic networks. *Journal of Economic Theory* 71:44–74.

Kelly, B., H. Lustig, and S. Van Nieuwerburgh. 2013. Firm volatility in granular networks. Working Paper, NBER.

Kermani, A. 2014. Cheap credit, collateral and the boom-bust cycle. Working Paper, U. C. Berkeley.

López-Pintado, D. 2012. Influence networks. *Games and Economic Networks* 75:776–87.

Loutskina, E., and P. E. Strahan. 2009. Securitization and the declining impact of bank finance on loan supply: Evidence from mortgage origination. *Journal of Finance* 64:861–89.

MacKenzie, D. 2011. The credit crisis as a problem in the sociology of knowledge. *American Journal of Sociology* 116:1778–841.

Mian, A., and A. Sufi. 2009. The consequences of mortgage credit expansion: Evidence from the U.S. mortgage default crisis. *Quarterly Journal of Economics* 124:1449–96.

———. 2011. House prices, home equity-based borrowing, and the U.S. household leverage crisis. *American Economic Review* 101:2132–56.

———. 2014. *House of debt*. Chicago: University of Chicago Press.

Nadauld, T. D., and S. M. Sherlund. 2013. The impact of securitization on the expansion of subprime Ccredit. *Journal of Financial Economics* 107:454–76.

Nier, E. W., J. Yang, T. Yorulmazer, and A. Alentorn. 2008. Network models and financial stability. Working Paper, Bank of England.

Palmer, C. J. 2015. Why did so many subprime borrowers default during the crisis: Loose credit or plummeting prices? Working Paper, U. C. Berkeley.

Poon, M. A. 2008. From New Deal institutions to capital markets: Commercial consumer risk scores and the making of subprime mortgage finance. *Accounting, Organizations and Society* 35:654–74.

Rajan, R. 2010. *Fault lines: How hidden fractures still threaten the world economy*. Princeton: Princeton University Press.

Shiller, R. J. 2014. Speculative asset prices. *American Economic Review* 104:1486–517.

Stanton, R., J. Walden, and N. Wallace. 2014. The industrial organization of the U.S. residential mortgage market. *Annual Review of Financial Economics* 6:259–88.

Uzzi, B. 1997. Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly* 42:35–67.

Zhong, Z. 2014. The risk sharing benefit versus the collateral cost: The formation of the inter-dealer network in over-the-counter trading. Working Paper, University of Melbourne.