Endogenous Financial Networks: Efficient Modularity and Why Shareholders Prevent It

Matthew Elliott and Jonathon Hazell*

February 2016

Abstract

We consider systemic risk in financial networks, by examining the conflict of interest between debt- and equity-holders. Through trading, banks can diversify their idiosyncratic risks and avoid failures following small shocks. However, the resulting interdependencies can cause multiple failures after large shocks. A social planner resolves this trade-off by creating a modular network structure with *fire breaks*, thereby preventing failures from small shocks while containing contagion. Socially efficient networks favor debt-holders over equity holders, meaning equity-holders can profitably trade away from these networks. Moreover, profitable trades for equity holders align their counter-parties' failures with their own, creating systemic risk.

---

*Elliott: Cambridge and California Institute of Technology; Hazell: MIT, email: jhazell@mit.edu. We thank Daron Acemoglu, Ben Golub, Sanjeev Goyal, Matthew Jackson, and Alp Simsek, as well as numerous seminar participants, for helpful comments and discussion. Funding from Caltech's Summer Research Fellowship programme is gratefully acknowledged.

1 Introduction

During the recent financial crisis, concerns over the systemic risk generated by financial interconnections underpinned policy interventions. Therefore, to design effective regulation we need a clear understanding of precisely how linkages between financial institutions contribute to systemic risk. Our comprehension of these issues has advanced rapidly. We now better understand which network structures are conducive to the spread of financial contagion, the mechanisms through which financial interconnections amplify real-economy shocks, and how to identify systemically important financial institutions. Nevertheless, our understanding of some fundamental questions is still somewhat limited. For example, what network of interdependencies would a benevolent social planner choose? What networks should we expect shareholder value maximizing banks to generate in a decentralized system? Would these differ from socially optimal networks—and if so, why? These questions are crucial for informed contemplation of regulation, and particularly for trying to anticipate the response of market participants.

Indeed, it is unclear that individually optimal behavior will necessarily generate excessive systemic risk, via the network structure. Cochrane (2014), for example, presents a critique of the so-called "domino effects" explanation of financial contagion:

> *"The dominoes or interconnectedness theory is a popular ... view of a crisis: A defaults on its debts to B, so B defaults on its debts to C, and so forth. ... [But] companies build buffers against dominoes."*

Thus in equilibrium, financial institutions might rearrange their interconnections to prevent large exposures to many counterparties, and so circumvent widespread financial contagion. In view of this critique, it is important to explain why privately optimising banks might deviate from socially efficient structures, and generate excessive systemic risk. We attempt to provide such an explanation, using a financial networks model. In particular, we study the role of limited liability together with the absence of debtor discipline, in creating a wedge between private and social interests.

We model n financial institutions, also termed banks, who each own proprietary assets that generate a random return. Banks can diversify their exposures to their proprietary assets by trading with each other. However, these trades create dependencies between banks' balance sheets. Overall market values are determined by the value of asset returns flowing to each bank. Banks have liabilities to external creditors, and the residual market value after these liabilities are settled becomes shareholder value. In states of the world where market values are less than the value of external debt, banks default. These defaults trigger discontinuous falls in the value of interbank contracts and external debt, due to bankruptcy costs. Thus bankruptcy costs may serve to propagate shocks originally from falls in asset returns. We make two crucial further assumptions. Firstly, shareholders are protected by limited liability, and so always have zero equity value in the event of bankruptcy. Secondly, there is no debtor discipline, so that only shareholders choose the structure of interbank contracts.

Our first set of results concerns socially efficient networks. We introduce a key trade-off in the social planner's problem, concerning small versus large shocks to asset returns. On the one hand, strong interconnections across many counterparties can prevent bankruptcies after relatively small shocks. However, they also render the system fragile to large shocks, by transmitting the shock across multiple counterparties and potentially leading to financial contagion.[1] Therefore we assess how the social planner optimally manages these two competing objectives. In our first key finding, the social planner's optimal response is to partition the financial system into "clusters" of banks, with strong dependencies within each cluster, and much lower dependencies across clusters. Small shocks are absorbed without triggering bankruptcies, and large shocks only cause single clusters to fail. Intuitively, the socially efficient structures create *firebreaks*, which stop the spread of financial contagion after large shocks and inhibits propagation via bankruptcy costs, but still prevent failures after small shocks. Firebreaks remain important when we allow for heterogeneous bank size. Networks somewhat resembling real-world networks, such as core-periphery structures, may be socially optimal. However these structures require firebreaks *between* core banks—which seems unlikely to hold in real-world financial systems.

Our first key result therefore matches a conjecture by Haldane (2010). Haldane notes that in settings as diverse as electrical engineering systems, computer manufacturing supply chains, forest fire management, the spread of infectious diseases, and terrorist organizations; networks are often optimally characterized by *modularity*.

> *Al'Qaeda ... operates not as a centralised, integrated organisation but rather as a highly decentralised and loose network of small terrorist cells. ... As events have shown, Al'Qaeda has exhibited considerable systemic resilience in the face of repeated and on-going attempts to bring about its collapse.*
>
> *These two characteristics are closely connected. A series of decentralised cells, loosely bonded, make infiltration of the entire Al'Qaeda network extremely unlikely. If any one cell is incapacitated, the likelihood of this undermining the operations of other cells is severely reduced. That, of course, is precisely why Al'Qaeda has chosen this organisational form. Al'Qaeda is a prime example of modularity and its effects in strengthening systemic resilience.*

Efficiently designed systems are partitioned, allowing only weak links across partitions in order to "strengthen system resilience." As Haldane observes, "banking ... has many of the same basic ingredients as other network industries, in particular the potential for viral spread and periodic systemic collapse." These observations suggest that a modular structure might also be optimal in financial systems, and we confirm this conjecture formally[2].

We next consider privately optimal behavior. We endogenize interbank relations, to consider whether individually optimising banks might arrange their interdependencies in a socially efficient pattern. In particular, we assume that banks trade on a bilateral basis, and aim to maximize shareholder value. Trades allow banks to hedge risks by diversifying away from their proprietary assets, thereby preventing failures in some states of the world. We find that socially efficient networks are typically not privately stable. This divergence between private and social incentives is driven by the interaction be-

[1] Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015a) discuss this "robust-yet-fragile" nature of financial system at length, which is also of concern to policymakers (Haldane, 2009). Cabrales, Gottardi, and Vega-Redondo (2014) consider similar issues.

[2] An alternative response by regulators to network risk, has been to introduce central clearing counterparties (CCPs). However, it seems that CCPs may not fully mitigate network-based risks—see Tarullo (2015) and BIS (2015a).

tween limited liability, and the absence of debtor discipline—both empirically relevant assumptions. These two assumptions create a conflict of interest between equity and debt holders, which generates a wedge between private and social optimality. The social planner maximizes the sum of shareholder and debt-holder value, which entails minimizing the expected number of defaults. To prevent defaults, the social planner redistributes surplus from the shareholders of healthy organizations to the debt holders of distressed organizations. Consequently the social planner's interests align with debt holders and against equity holders. To avoid this redistribution, shareholders seek to trade away from socially efficient networks. Though such trades may raise the probability of default, shareholders are protected by limited liability. While being exposed to other banks' defaults is costly for the shareholders of a solvent bank, a bank's default is not costly to its own shareholders at the margin.

Our second key finding is that limited liability causes banks to *prefer* trades which generate systemic risk, as opposed to idiosyncratic risk. We therefore potentially resolve the critique of Cochrane (2014). In the states in which a bank fails at the same time as its counterparties, shareholders are protected from counterparties' bankruptcy costs by limited liability. Instead, in these states bankruptcy costs are passed onto debt holders. Therefore, all else equal, shareholders aim to fail in the same states as their counterparties, even though this disadvantages creditors. Thus shareholders prefer systemic risk. They trade in order to correlate their failures with those of their largest counterparties. Moreover, this preference for systemic risk occurs even without the possibility of government intervention[3].

There are many reasons why financial connections between banks arise, other than the hedging motive that we consider. Intermediation (Acemoglu et al., 2015b; Farboodi, 2014) and the possibility of government bailouts (Eisert and Eufinger, 2013) are two important alternatives. We expect that possible bailouts would exacerbate the incentives that banks have to correlate their failures, which already exists in our model. By this token, part of our contribution is establishing that incentives to correlate failures exist, even without government bailouts. We also go on to discuss ways in which our model potentially captures some aspects of intermediation, albeit in a reduced form. Regardless, financial flows devoted to hedging are very large (see BIS, 2015b), undergirding the importance of the mechanisms we study.

1.1 Literature

The literature examining how the structure of exogenous financial networks affects systemic risk has expanded rapidly. Building on early important works, such as Allen and Gale (2000) and Freixas et al. (2000), this literature emphasizes that interconnections can facilitate the spread of contagion. Networks can generate systemic risk by facilitating the spread of relatively large shocks (Gai and Kapadia, 2010), or by interacting with various propagation mechanisms, such as bankruptcy costs (Elliott, Golub, and Jackson, 2014); uncertainty about banks' balance sheets (Caballero and Simsek, 2013; Alvarez and Barlevy, 2014); and fire sales more generally (Cifuentes, Ferrucci, and Shin, 2005). Since these papers consider exogenous networks, they do not examine socially efficient network structures, nor whether individual banks might choose to deviate from the social optimum. Our paper is also related to the empirical literature on interbank networks and systemic risk—see, for example, Denbee,

[3] Farhi and Tirole (2012), amongst others, discuss the incentives for banks to take on systemic risk in the context of government intervention. In our model, incentives to take on systemic risk are important regardless of policymakers' actions.

Julliard, Li, and Yuan (2014).

A smaller literature focuses on socially efficient network structures and the endogenous formation of financial networks, in the presence of systemic risk. Two particularly relevant papers[4] are Farboodi (2014) and Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015b). Each of these models endogenous network formation, and so examines why privately and socially optimal behavior might differ. In Farboodi (2014), banks form links due to intermediation, since some banks have access to risky investment opportunities or funding opportunities, and others do not. Banks also intermediate to capture rents. Farboodi (2014) is able to find equilibrium networks, and the equilibrium structure she identifies matches empirically observed financial networks—there is a core of investment banks connected to each other and each with links to a set a commercial banks. In her model private behavior is socially inefficient because core banks behave in an excessively risky manner, to capture intermediation rents. Acemoglu et al. (2015b) also focusses on financial intermediation as the reason for network formation. Privately and socially efficient behavior may diverge because of a financial network externality—banks contract on a bilateral basis, and so do not account for their role in creating a conduit that allows idiosyncratic shocks to develop into contagion. In our model, systemic risk also arises endogenously but from a different set of frictions, pertaining to shareholders' control rights and limited liability—moreover, the frictions in our model mean banks aim to correlate failures, and thus generate excess systemic risk.

Two other germane papers are Cabrales, Gottardi, and Vega-Redondo (2014) and Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015a). Both examine social efficiency in the context of financial contagion. Acemoglu et al. (2015a) identify a key tradeoff facing the social planner. Denser connections prevent bankruptcies from small shocks, but facilitate the spread of contagion from large shocks. Therefore highly connected networks are optimal if all shocks are relatively small, and facilitate cascades of failures if shocks are large. Conversely, if shocks are always large, then weak connections in the network can enhance stability by preventing the spread of contagion. Our paper builds on this analysis by considering socially efficient networks when shocks hitting the system can be *both* large or small with positive probability; and introduces a propagation mechanism via bankruptcy costs.

Like us, Cabrales et al. (2014) study socially efficient networks when shocks to asset values can be both relatively small or relatively large. In their model systemic risk arises from direct claims on other banks' projects. By considering some key benchmark networks, Cabrales et al. (2014) explore the social planner's preferences over network segmentation and link density. Dense interconnections prevent failures from small shocks, but facilitate the spread of contagion from large shocks. Building on their work we are able to solve the social planner's problem without restricting attention to benchmark networks. We also include bankruptcy costs, which provide the key propagation mechanism in our setting.[5]

The role of limited liability in generating financial instability has been widely discussed in the context of a single-bank framework (e.g. Merton, 1974; Jensen and Meckling, 1976; Rochet, 1992 and

[4] Other papers considering similar problems include Leitner (2005); Blume et al. (2011), Allen, Babus, and Carletti (2012); Babus (2013); Zawadowski (2013); Eisert and Eufinger (2013); Di Maggio and Tahbaz-Salehi (2014); Erol and Vohra (2014); Wang (2014); Cohen-Cole, Patacchini, and Zenou (2015); and Galeotti, Ghiglino, and Goyal (2015).

[5] Cabrales, Gale, and Gottardi (2015) explains this modelling difference in detail vis-a-vis Elliott et al. (2014), which provides the foundation of our model. Essentially, in our model banks hold claims on counterparties' balance sheets, as opposed to claims on underlying assets. This difference becomes important when bankruptcy costs are introduced.

Gollier et al., 1997). However, to the best of our knowledge we are the first to examine the effect of limited liability in a *network* setting.

We structure the paper as follows. Section 2 presents our model. Section 3 examines socially efficient networks. Section 4 discusses equilibrium networks. Section 6 concludes. Proofs are relegated to Appendix A.

2 The Model

2.1 Financial Institutions, Real Assets, Revenue, and Contracts

There is a set $N = \{1, \ldots, n\}$ of financial institutions, which we refer to as banks. Each bank holds a proprietary asset. This yields a return p_i for bank i, derived from i's investments in the real economy. Bank can diversify their exposures by holding claims on each other. We denote these claims on market values as *cross-holdings*. We let $C_{ij} \geq 0$ be the *cross-holding* of bank i in bank j.[6] We can represent $\mathbf{C}$ as a weighted, directed graph, where the banks are the nodes and the links represent their claims on each other. Equivalently, $\mathbf{C}$ is a *financial network*. After bank i has paid off all the claims on its own market value, there is a share $\widehat{C}_{ii} := 1 - \sum_{j \in N} C_{ji}$ left, assumed to be strictly positive[7]—it is the proportion of bank i's market value flowing to the outside shareholders of bank i. Setting $\widehat{C}_{ij} = 0$ for all $j \neq i$, $\widehat{\mathbf{C}}$ is a diagonal matrix.

The market value V_i of bank i is given by:

$$V_i = \sum_{j \in N} C_{ij} V_j + p_i.$$

In matrix notation, this is equivalent to

$$\mathbf{V} = \mathbf{C}\mathbf{V} + \mathbf{p} = (\mathbf{I} - \mathbf{C})^{-\mathbf{1}}\mathbf{p}.$$

The market value flowing to the shareholders of bank i, denoted v_i, is given by $\widehat{C}_{ii} V_i$ so that we have[8]:

$$\mathbf{v} = \widehat{\mathbf{C}}\mathbf{V} = \widehat{\mathbf{C}}(\mathbf{I} - \mathbf{C})^{-\mathbf{1}}\mathbf{p} = \mathbf{A}\mathbf{p} \tag{1}$$

As in Elliott et al. (2014), $\mathbf{A} = \widehat{\mathbf{C}}(\mathbf{I} - \mathbf{C})^{-1}$ is termed the *dependency* matrix—since equation (1) shows that the market value flowing to shareholders of a given bank depends ultimately on the returns from the real assets $\mathbf{p}$. In particular, the market value flowing to the shareholders of bank i is

$$v_i = \sum_{j \in N} A_{ij} p_j.$$

[6] We will often set $C_{ii} = 0$ for all i, but it will sometimes be helpful to allow banks to retain some claim on themselves in addition to the claims of their outside shareholders. These claims might, for example, represent internal claims within a bank that one division has on another.

[7] This assumption ensures that market value flowing to outside shareholders, discussed below, is well defined.

[8] Under the assumption that $\widehat{C}_{ii} > 0$ for all i, so that $\mathbf{C}$ is column substochastic, the inverse $(\mathbf{I} - \mathbf{C})^{-1}$ is well defined and non-negative.

Reassuringly, $\mathbf{A}$ is column stochastic[9], so that

$$\sum_{i \in N} A_{ij} = 1 \text{ for all } j \in N.$$

Thus the returns from every real asset are ultimately allocated to the outside shareholders of some bank. Banks can effectively construct portfolios of the underlying, primitive assets by holding claims on each others' market values—since banks all own and administer proprietary projects, claims on underlying assets cannot be held directly. Equally, the dependency matrix formulation reflects the fact that while banks negotiate contracts with one another, ultimately these contracts derive their value from returns to assets in the real economy.

To help fix ideas, it is worth further exploring the relationship between the dependency matrix $\mathbf{A}$, which represents the flow of returns from proprietary assets to banks; and the cross-holding matrix $\mathbf{C}$, which represents the flow of market value between banks. One representation of $\mathbf{A}$ is the infinite sum known as a Neumann series:

$$\mathbf{A} = \widehat{\mathbf{C}} \sum_{p=0}^{\infty} \mathbf{C}^p$$

This representation allows an intuitive interpretation of the relationship between the cross-holding and dependency matrices. In particular, the total claim of bank i on asset j comes from the sum of indirect claims on assets via claims on banking counterparties—who also have a series of indirect claims on assets via their own holdings on other counterparties, leading to an infinite sequence of indirect claims. Consequently, there can be a large difference between the direct exposure of bank j to bank i, given by the cross-holding C_{ij}; and bank j's ultimate exposure to bank i's asset, given by dependency matrix entry A_{ij}. In general there may be no cross-holding matrices $\mathbf{C}$ that generates a given dependency matrix $\mathbf{A}$, even when that dependency matrix is column stochastic and non-negative.[10]

2.2 External Liabilities, Bankruptcy, Seniority and the Value of Contracts

Firstly, it is worth briefly discussing the nature of the interbank contracts, which follow the model set out by Elliott, Golub, and Jackson (2014). We aim to capture a key trade-off. Financial contacts can be used to hedge a bank's exposure to its underlying assets but these contracts come at the cost of creating balance sheet interdependencies. As idiosyncratic risks are shared, banks' market values become correlated. Moreover, when one bank fails, it incurs costs that reduce the value of counterparties' contracts with the bank. To capture this tradeoff we introduce a relatively simple contracting space, in which the value of claims on a bank strictly decrease as the bank's market value falls, and these claims are *linear* before and after the failure of that bank. This contracting space is highly incomplete. The value of contracts cannot be conditioned directly on states of the world, the actions of the banks in the system, or the value of other contracts. Nor can they depend on the value of banks—except in a linear fashion. If the contracting space were sufficiently rich, then the trade-off of interest would not

[9] Column stochasticity of $\mathbf{A}$ follows from the column stochasticity of $\mathbf{C} + \widehat{\mathbf{C}}$.

[10] See the online appendix, Section A.5, of Elliott et al. (2014).

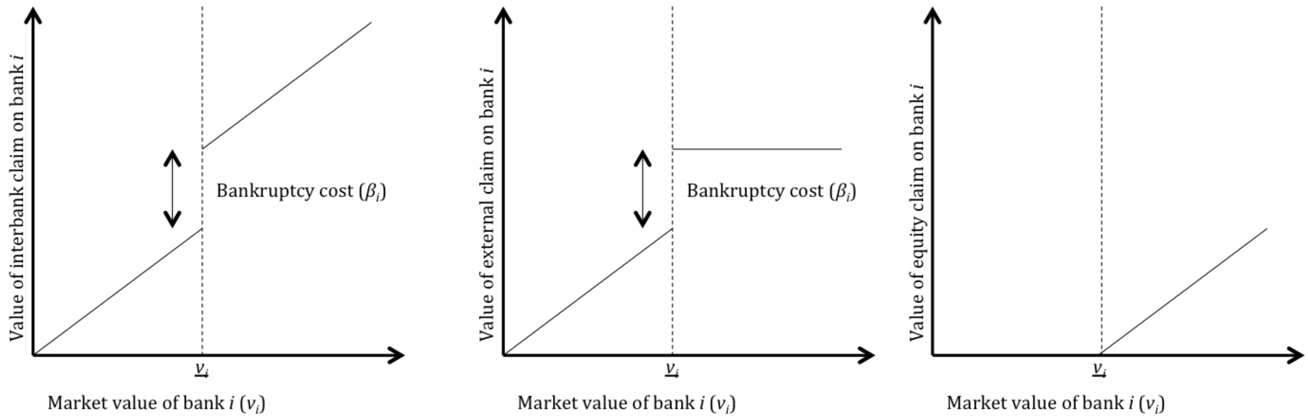

Figure 1: Value of interbank and external claims

exist in the model. Moreover, the interbank contracts we model generate networked interdependencies. A bank's value depends not just on its immediate counterparties, but also on its counterparties' counterparties, and so on.

Interbank contracts can take very many forms, and are often highly complex. However, we view linearity as a useful approximation to interbank contracts in general. Linearity captures a key fact, that the value of most interbank contracts is likely to decrease with an organization's value, especially in a distressed system.[11] Naturally, this is a feature of equity contracts. However, it is also likely that the value of debt contracts changes with the value of firms, as they approach their failure thresholds—and indeed, this is well documented empirically (e.g. Jones et al., 1984). We use linear contracts as a reduced form way of capturing a relatively wide range of contracts.

We now discuss banks' external liabilities, and also limited liability. It is helpful to make specific assumptions about the seniority of different claims. We choose assumptions that are hopefully simple and fairly natural, although we would expect to obtain similar results more generally. First, as bank i's market value falls, the value of other interbank claims on bank i also linearly fall—due to the linear nature of the interbank claims, as previously discussed. Secondly, while the value of interbank claims on bank i declines as i's market value falls, the value of bank i's external liabilities do not vary continuously with market value. However, for sufficiently large falls in market value, external liabilities can no longer be settled in full, and so bankruptcy must occur. Equity holders are residual claimants on market value—and so receive nothing after bankruptcy, given that they are protected by limited liability. If a bank formally enters bankruptcy proceedings, we assume all creditors are of equal seniority. Both interbank creditors and external creditors receive linearly rationed repayments. Moreover after bankruptcy proceedings, repayments are lowered by the failure costs associated with disorderly default. The features of these contracts are summarized in figure 1.

Formally, let bank i's external liabilities have a face value $\underline{v}_i$. Thus when we have $v_i \geq \underline{v}_i$, so that market value is greater than external liabilities, then external liabilities are settled in full. Remaining

[11] An alternative and popular modelling approach is to build on Eisenberg and Noe (2001), as in, for example Acemoglu et al. (2015a). They model debt contracts that retain their full value up until a firm fails, after which linear rationing occurs, as in our model. In practice, even the value of debt contracts decreases as firms approach their failure thresholds.

market value is rationed among interbank creditors and bank i's shareholders. When $v_i < \underline{v}_i$ firm i enters a disorderly bankruptcy phase. Outside shareholders now have no claim on remaining market values which are split between interbank claimants and external claimants in proportion to their claims (i.e. with equal seniority). We term $\underline{v}_i$ the *failure threshold* of bank i. After its default, the bank has to enter bankruptcy proceedings to settle with its creditors and liquidate its asset—this process is costly and irreversible. We capture this by subtracting bankruptcy costs of $\beta_i > 0$ from market values before they are rationed amongst creditors, in the case of bankruptcy.

Thus bank i's market value is:

$$V_i = \sum_{j \in N} C_{ij} V_j + p_i - \beta_i I_{v_i < \underline{v}_i},$$

where $I_{v_i < \underline{v}_i}$ is an indicator variable of value 1 if $v_i < \underline{v}_i$ and 0 otherwise. We can then augment the system given by equation (1) with the discontinuous, positive and finite failure costs, yielding:

$$\mathbf{v} = \widehat{\mathbf{C}}(\mathbf{I} - \mathbf{C})^{-1}(\mathbf{p} - \mathbf{b}(\mathbf{v})) = \mathbf{A}(\mathbf{p} - \mathbf{b}(\mathbf{v})),$$

and for a single bank:

$$v_i = \sum_{j \in N} A_{ij}(p_j - \beta_j I_{v_j < \underline{v}_j}).$$

It is worth briefly elaborating on the nature of bankruptcy costs. These are the costs of liquidating the proprietary asset, possibly at a discount during periods of financial turmoil[12]; of inefficient allocation of resources during the bankruptcy period; the cost of bankruptcy negotiations, administration and settlement; loss of human capital after workers are fired and so forth. Default costs are typically large and variable relative to firms' market values, as in James (1991) and Davydenko et al. (2012).

Letting π_i be the value accruing to banks i's outside shareholders, or equivalently bank i's equity value, and δ_i be the value of the claims held by outside debt holders, we obtain the following system of equations capturing the claims of all parties:

$$\begin{aligned} \mathbf{v} &= \mathbf{A}(\mathbf{p} - \mathbf{b}(\mathbf{v})) && (2)\\ \boldsymbol{\pi} &= \max\{\mathbf{v} - \underline{\mathbf{v}}, \mathbf{0}\} && (3)\\ \boldsymbol{\delta} &= \min\{\underline{\mathbf{v}}, \mathbf{v}\}, && (4) \end{aligned}$$

where $\mathbf{0}$ is the zero vector and the max and min operators are applied entry by entry.

Consequently, there is a settlement risk associated with the possibility of bank j defaulting, and thus transferring less funds to other banks, compounded by the discontinuous bankruptcy cost β_i. The dependency matrix determines how the market value generated by underlying assets flows to the owners of different banks. Bankruptcy costs are directly subtracted from these market values. Therefore bankruptcy costs end up being distributed according to the dependency matrix.[13]

[12]See Cifuentes et al. (2005), Gai and Kapadia (2010) or Caballero and Simsek (2013) for a more detailed treatment of fire sales.

[13]Due to bankruptcy costs, it would be more profitable for a bank i to directly hold a portfolio of assets on underlying assets, than to hold the equivalent portfolio indirectly through interbank claims. However, by assumption underlying assets are proprietary and not directly tradeable.

It is also worth emphasizing that the vector of shareholder values $\mathbf{v}$ reflects *limited liability*. Our model therefore captures the important difference between equity claims, and other claims on market value. When banks default shareholders receive nothing and external creditors are bailed in to become the new shareholders—the standard interpretation of bankruptcy under limited liability.[14]

2.3 Payment Equilibrium Existence and Multiplicity

A vector of shareholder market values $\mathbf{v}$ which solves equation (2) is a *payment equilibrium*—that is, a vector of shareholder market values which simultaneously satisfies all interbank payments. We require a payment equilibrium to always exist, so that there is always a fixed point of the system described by equation (2), for any return vector $\mathbf{p}$. As has been shown in a variety of related settings (e.g. Eisenberg and Noe, 2001; Acemoglu et al., 2015a) at least one payment equilibrium always exists. In the Supplementary Appendix, Section 3, we show that the set of payment equilibria in our model forms a complete lattice. Thus there is always a set of mutually consistent interbank payments solving equation (2), and there may be several.

The reason for the possible multiplicity of payment equilibria is the discontinuous default costs, which has two separate effects. Firstly, for a given bank i, there may be two separate values of shareholder market value v_i satisfying equation (2)—one where the bank has defaulted, so that $\pi_i = 0$ and $I_{v_i < \underline{v}_i} = 1$; and another where the bank has not failed, meaning $\pi_i \geq 0$ and $I_{v_i < \underline{v}_i} = 0$. This is essentially a form of the classic expectations-driven bank run, as in Diamond and Dybvig (1983). Secondly banks' market values are interdependent. Therefore, for any two banks i and j, there may be two shareholder market value vectors satisfying equation (2)—one in which both i and j fail, and the other in which neither fail. Intuitively, two banks may require high levels of debt repayment from one another to remain solvent, meaning that it is equally possible that they both default. Again, then, we can have multiple payment equilibria due to self-fulling expectations. Throughout, following Elliott et al. (2014), we focus on the best case equilibrium with the fewest failures. In any other equilibrium, all the organizations that fail in the best case equilibrium still fail, as well as some additional organizations. We could instead focus on the worst case equilibrium without appreciably altering our results.

2.4 Financial Contagion

We briefly elaborate on the nature of financial contagion in our model. We take contagion to be the failure of multiple banks in the network. Consider the expression for the market value of bank i:

$$v_i = \sum_{j \in N} A_{ij}(p_j - \beta_j I_{v_j < \underline{v}_j}).$$

Financial contagion can be driven by two complementary sources. The first is from correlated exposure to underlying asset holdings. Multiple banks may be simultaneously exposed to a large fall in underlying asset values—that is, from a large fall in value of some asset p_j. This interdependency

[14] For simplicity our formulation assumes that external creditors take on the claims of shareholders after a failure such that $\widehat{C}_{ii}$ remains constant before and after i has failed. Relaxing this assumption would result in a straightforward change in the dependency matrix before and after default, which would complicate notation without yielding additional insights.

arises when banks diversify their exposure away from their own underlying assets, meaning indirect claims on other banks' underlying assets, via cross-holdings on counterparties' balance sheets. Second, falls in asset values can then be amplified by bankruptcy costs, further depressing the value of interbank claims by some bankruptcy cost β_j. Thus bankruptcy costs serve as a propagation mechanism internal to the banking system. The existence of such propagation mechanisms appears to be extremely important in generating systemic risk in financial networks, both theoretically (Glasserman and Young, 2015) and empirically (Brunnermeier and Oehmke, 2012). It is straightforward to extend our model to include correlated holdings of underlying assets, as in Elliott et al. (2014). In one interpretation of our model, we implicitly consider a system that is already distressed—since the value of interbank claims on any bank falls as the bank's market value fall. This feature could, for example, reflect an unmodelled shock to some commonly held asset.

2.5 Hedging and Financial Intermediation

As we go on to explain further in subsequent sections, the reason for financial interconnections in our model is *hedging*, which is responsible for an important part of financial flows (BIS, 2015b). By exchanging financial connections with one another, banks can avoid failures in some states of the world – which may be desirable from the perspective of banks, and also the social planner. An important alternative reason for financial interconnections is intermediation (e.g. Farboodi, 2014; Acemoglu et al., 2015b). While our model is primarily motivated by hedging, it potentially captures some aspects of financial intermediation, albeit in reduced form. One could view the assets in our model as investment opportunities that arrive at a bank, through its connections to outside parties. The bank makes these investments, but seeks to diversify its exposure, hedging its position through trades with other banks. These trades transfer idiosyncratic risk to other banks, and mean the other banks effectively partly finance the investment. Thus the bank intermediates trade between the outside party with the investment opportunity, and other banks in the network. Moreover, in financial intermediation models such as Acemoglu et al. (2015a), the social planner aims to avoid early liquidation of investment projects, by rearranging the network architecture. This problem is very similar to the social planner's objective in our model, which, as we go on to discuss, is to prevent bankruptcies.

3 Socially Efficient Networks

We consider a constrained social planner who maximizes the sum of expected payments to bank shareholders and creditors, or equivalently, the sum of expected equity value and expected external debt repayments. All agents in our model are risk neutral. Therefore, assuming transfers can be made, an outcome fulfils the social planner's objective if and only if it is Pareto efficient. Such an outcome also corresponds to the objective function of a utilitarian social planner. As with the individual banks, the social planner knows the probability distribution of returns, but determines the socially efficient pattern of network links before returns are realized.

We constrain the social planner to choosing from the class of networks that are *fair*, in the following sense. Any network chosen by the social planner must preserve each bank's market value prior to the shock. Equivalently, any exchange of claims made to generate a socially efficient network must be individually rational for each bank, absent shocks. By assumption, all banks directly own a single proprietary asset that generates the same probability distribution over returns, though returns may differ across states. We require that any network chosen by the social planner leaves each bank no worse off, conditional on no banks failing. This restriction is equivalent to requiring that the $\mathbf{A}$ matrix is row stochastic, such that $\sum_{j \in N} A_{ij} = 1$ for all i.[15]

The social planner must therefore choose a feasible cross-holdings matrix to solve the following constrained optimization problem:

$$\max_{\mathbf{C}} \mathbb{E}\left[\sum_{i \in N} \pi_i(\mathbf{C}) + \delta_i(\mathbf{C})\right] \qquad \text{subject to} \quad \sum_{j \in N} A_{ij} = 1 \text{ for all } i. \tag{5}$$

3.1 Simplification and Shocks

We begin by working with a simplified model in which banks are symmetric[16], insofar as they have the same bankruptcy costs and the same levels of external liabilities, that is $\underline{v}_i = \underline{v}$ and $\beta_i = \beta$ for all $i \in N$.

We also assume that each bank owns a single underlying asset that generates a baseline return R. We let a random, firm-specific, finite shock of ε hit the underlying asset of one bank selected uniformly at random. Hence when the shock hits bank i, the vector of realized underlying asset returns is given by

$$\mathbf{p} = [R \ldots \underbrace{R - \varepsilon}_{\text{asset } i} \ldots R]^T.$$

We motivate our assumption that the shock hits only one bank by the premise that shocks are sufficiently rare for at most one shock to hit each period. Although stylized, we think of it as a useful benchmark—Caballero and Simsek (2013), Cabrales et al. (2014) and Acemoglu et al. (2015a) take a similar approach. The social planner's problem can be represented in a simple fashion.

REMARK **1.** The social planner's problem is equivalent to minimising the expected number of defaults subject to the same constraints.

Defining $\varepsilon^* := n(R - \underline{v})$, we let shocks $\varepsilon \in [R - \underline{v}, \varepsilon^*]$ be *small* and shocks $\varepsilon > \varepsilon^*$ be *large*. Thus, absent any cross-holdings, even small shocks are enough to cause an organization to fail. The threshold ε^* implies that at least one organization must fail following a large shock; although it is possible for no organizations to fail following a small shock. Intuitively, $n(R - \underline{v})$ is the total surplus equity value available in the system. When the shock is greater than total equity value, some defaults must occur for the system to clear and a payment equilibrium to form—conversely, for small shocks surplus equity value can always be distributed in such a way that defaults are prevented. A network structure that prevents defaults from small shocks, is when organizations arrange their interconnections so that their portfolios approach the market portfolio. By contrast, if all shocks are large the empty network

[15] We discuss this fairness assumption in the Supplementary Appendix (Section 2), and show we can relax it without changing our key findings.

[16] In section 5 we relax the assumption of size homogeneity. For a model that explores other important interbank heterogeneities, see Farboodi (2014).

is socially efficient, because only a single bank fails, namely the bank whose asset is hit by the shock. Consequently, a trade-off is apparent. More interconnections will tend to insure the system against failures from small shocks, but may also cause financial contagion by exposing more banks to large shocks, meaning potentially more failures.[17]

We study shock distributions where both small and large shocks are possible. Thus we examine how the socially efficient network structure balances between insuring against small shocks and limiting contagion following large shocks. In the interest of clarity, we initially adopt the simplest shock distribution capable of capturing such a tradeoff[18]. We assume ε is random and can be either large or small with some probability:

$$\varepsilon = \begin{cases} \varepsilon_L > \varepsilon^* & \text{with probability } q \\ \varepsilon_S \in [R - \underline{v}, \varepsilon^*] & \text{with probability } 1 - q. \end{cases} \tag{6}$$

We will focus on the case of relatively common small shocks and rare large shocks, that is, when q is relatively small. If large shocks are relatively common, the social planner simply aims to prevent the spread of contagion and constructs a network with little interconnection, such as the empty network. Only one bank fails, under any shock realisation, so that the social planner does not face a meaningful tradeoff. Conversely, when large shocks are rare, the social planner must balance preventing failures from small shocks with minimising the extent of contagion from large shocks. We define sufficiently rare shocks by:

$$q < \frac{1}{n^2}.$$

The reason for this choice will be explained in detail in the following section. Thus the social planner trades off preventing failures from small shocks, with preventing contagion after large shocks. It is crucial to emphasize that this tradeoff exists due to the highly incomplete contract space. Given that a small shock is realized, maximal risk sharing is optimal. Conversely, after a large shock the social planner wants to minimize risk sharing. However, the social planner is unable to select contracts that facilitate both these objectives simultaneously. The planner chooses contracts before asset returns are realized and cannot condition on states of the world. Nor can the planner condition on banks' values, except linearly, and thus cannot construct contracts that behave very differently in the presence of large versus small shocks. Hence the social planner must balance the two objectives against one another when selecting the network. This is an important tradeoff in the financial networks literature, since the seminal work Allen and Gale (2000), and also in more recent work (see Acemoglu et al., 2015a and Cabrales et al., 2014). The linear contracts in our model provide a way for the social planner to share risks between firms *prior* to the default of a given firm[19], and thereby attempt

[17]The discontinuity at ε^* in the socially optimal network, when shocks of a fixed size hit, is reminiscent of Propositions 2 and 3 in Acemoglu et al. (2015a). The idea contrasts with Allen and Gale (2000), who show that in a four-bank financial network with idiosyncratic shocks, it is generally better for each bank to diversify its holdings. However, in Allen and Gale (2000)'s setting, all shocks are small. Acemoglu et al. (2015a) and Cabrales et al. (2014) demonstrate in detail that a general trade-off exists between insuring against failures from small shocks, and preventing contagion from large shocks.

[18]In section 5 we generalize the shock distribution in several respects.

[19] With standard debt contracts, as in Eisenberg and Noe (2001), firms can only share risks *after* the default of a given firm. Therefore, these contracts may be less appropriate when considering the social planner's key tradeoff, since debt contracts

to minimize the number of defaults. With linear contracts the social planner must balance the twin objectives of preventing failures after a small shock and minimising the spread of contagion—they are thus a reasonable set of contracts to consider, given the social planner's tradeoff.

3.2 Socially Efficient Networks

In this section, we discuss our first key finding. We characterize the optimal network, conditional on no banks failing following a small shock. Observe that no banks fail following any small shock if and only if for all $i, j \in N$:

$$A_{ji}\varepsilon_S \leq \sum_{k \in N} A_{jk} R - \underline{v}.$$

Subject to this constraint, we seek to minimize the size of the set of banks that fail following a large shock to given asset i, denoting this set of banks by D_i.

LEMMA **1.** For all network structures such that no banks fail following any small shock, at least $\lceil d^* \rceil$ banks fail following any large shock to i, where d^* is the unique positive root of:

$$d^2(R-\underline{v})\beta + d[(R-\underline{v})(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta] + n\varepsilon_S(R-\underline{v}) - \varepsilon_S\varepsilon_L = 0.$$

For the remainder of this section we abstract from the integer problems. We assume that

(i) d^* is an integer; and

(ii) n/d^* is an integer.

Ignoring integer problems, we can partition the organizations into n/d^* groups of d^* agents. Define G_i as i's group. Let $\mathcal{A}$ be the set of feasible dependency matrices, which are non-negative and column stochastic. We then define the set of networks:

$$\mathcal{A}^*(d^*) := \left\{ \mathbf{A} \in \mathcal{A} : \begin{array}{rcll} |G_i| & = & d^* & \text{for all } i \in N \\ A^*_{ji} & = & \frac{R-\underline{v}}{\varepsilon_S} & \text{for all } i,j : G_i = G_j \\ A^*_{ji} & = & \frac{R-\underline{v}}{\varepsilon_L + \beta d^*} & \text{for all } i,j : G_i \neq G_j \end{array} \right\}.$$

Thus $\mathcal{A}^*$ characterizes symmetric networks where organizations are partitioned into groups of size d^*, which have strong connections within groups, and weak connections between groups. For reasons that will become clear, we refer to the across-group links as *firebreak* links, and refer to the value of these links as the *strength* of the firebreaks. Following a large shock to the asset of a given bank, all banks within the same group fail but no banks outside the group fail. In addition, all banks perfectly insure themselves against all small shocks. In general, we say that for networks in which banks can be partitioned into groups with weak links between groups, and strong links within groups, these groups are called *clusters*. Thus the networks in $\mathcal{A}$ exhibit clusters. The network representation of dependencies for an $\mathbf{A} \in \mathcal{A}^*(6)$ is shown in Figure 2.

cannot be used to prevent defaults.

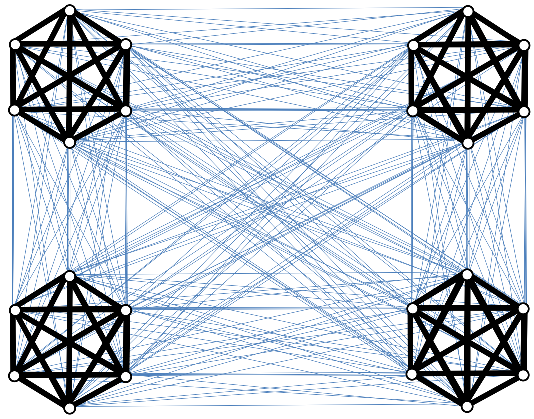

Figure 2: A network with 4 clusters and 6 banks in each cluster.

The set of networks $\mathcal{A}^*$ are in the space of banks' underlying claims on each others' assets, and it is these dependencies that ultimately matter in the analysis in this section. Indeed, working in this space of dependencies greatly simplifies the analysis. We first consider the best way to distribute losses from the shocks within the system, and then consider whether there exist underlying claims banks can have on one another to achieve this distribution of losses.

Dependencies $\mathbf{A} \in \mathcal{A}^*$ are generated by networks of cross-holdings with a similar structure, whereby banks have strong cross-holdings in other banks in the same group, and weak cross-holdings on banks in the rest of the network. Let $\mathcal{C}$ be the set of feasible cross-holding matrices, which are non-negative, column sub-stochastic and zero diagonal. Grouping banks in the same way as above, we define a set of networks in terms of cross-holdings:

$$\mathcal{C}^*(d^*) := \left\{ \mathbf{C} \in \mathcal{C} : \begin{array}{rcll} |G_i| & = & d^* & \text{for all } i \in N \\ C^*_{ii} & = & 0 & \text{for all } i \in N \\ C^*_{ji} & = & \kappa\phi & \text{for all } i,j : G_i \neq G_j \\ C^*_{ji} & = & \frac{1-\phi(1+(n-d^*)\kappa)}{d^*-1} & \text{for all } i,j : G_i = G_j \end{array} \right\}.$$

Since $\widehat{\mathbf{C}} + \mathbf{C}$ is column stochastic, it follows that $\widehat{\mathbf{C}}^*(\phi) = \phi I$.

REMARK **2.** Continuing to abstract from integer problems, for all d^* and all $\mathbf{A}^* \in \mathcal{A}^*(d^*)$, there exists $\kappa^* \geq 0$ such that:

$$\lim_{\phi \to 0} \widehat{\mathbf{C}}^*(\phi)(\mathbf{I} - \mathbf{C}^*(\kappa^*, \phi))^{-1} = \mathbf{A}^*.$$

PROPOSITION **1.** Abstracting from integer problems, under conditions (i) and (ii) above, if large shocks are sufficiently rare such that $q < \frac{1}{n^2}$, then a network $\mathbf{C} \in \mathcal{C}$ solves the social planner's problem if and only if it generates a dependency matrix $\mathbf{A}(\mathbf{C}) \in \mathcal{A}^*$.

Proposition 1 is our first key finding. It establishes that under rare enough large shocks, *all* socially efficient networks are characterized by a particular structure. Banks are grouped into clusters, as previously defined. All banks have symmetric cross-holdings and asset dependencies within-cluster and symmetric cross-holdings and asset dependencies between-cluster. Moreover, the result is derived under relatively general assumptions. We do not place restrictions on the possible size of any cross-holdings between banks; nor on the identities of the banks that have cross-holdings with one another. Although we permit the social planner to form asymmetric networks, the socially efficient networks are highly symmetric. Indeed the sole restriction that we place on the set of networks considered by the social planner beyond feasibility, is that the dependency matrix $\mathbf{A}$ is row stochastic.

Proposition 1 is intuitive. By construction, after a large shock at least one bank must default and under a small shock there are configurations such that no banks fail. Therefore a candidate for the socially efficient network is a configuration such that no banks default under small shocks, and as few as possible under large shocks. The networks satisfying this property exhibit clusters. Every bank in a cluster defaults when a large shock hits that cluster. However, links between clusters are sufficiently weak that the rest of the network will not fail after a large shock. Links within clusters and between clusters are together strong enough that when a small shock hits a given cluster, no banks default. Value is transmitted through the entire network, including outside the cluster, to the distressed bank—allowing the social planner to minimize the cluster size, and hence the number of defaults after a large shock. The condition $q < \frac{1}{n^2}$ ensures the large shocks are sufficiently rare for the empty network to be suboptimal.

Proposition 1 allows a better understanding of the role of bankruptcy costs in potentially propagating financial contagion—and how the socially efficient network structure prevents such propagation from actually occurring. A key role is played by the links between clusters. These links are relatively weak, but positively valued, since for any $\mathbf{A}^* \in \mathcal{A}^*$ we have:

$$0 < A^*_{ij} = \frac{R - \underline{v}}{\varepsilon_L + \beta d^*} < \frac{R - \underline{v}}{\varepsilon_S} = A^*_{jk} \quad \text{for } G_i \neq G_j = G_k.$$

Appealing to metaphor, and as already mentioned, we refer to the weak between-cluster links, given by A^*_{ij} above, as *firebreaks*. Firebreaks allow the social planner to prevent any "domino effect" from occurring between clusters. Banks who do not initially fail after a large shock to another bank might subsequently fail—exposure to bankruptcy costs could cause cascading secondary defaults. However, in the socially efficient network the cumulative impact of the large shock and bankruptcy costs does not cause failures beyond the cluster hit by the shock. Instead, the firebreaks ensure that links between clusters are too weak to transmit contagion. Equally, the weak links allowed by the firebreaks mean that value is still allocated from across the network into a cluster suffering a small shock. Consequently fewer banks within a cluster are required in order for them to absorb the impact of a small shock without default. Therefore firebreaks allow an optimal response by the social planner to the tradeoff identified in subsection 3.1. While preventing failures from small shocks, firebreaks also prevent the spread of contagion from large shocks.

It is suboptimal for clusters to be fully segmented, such that no bank holds claims to an asset from another cluster. While full segmentation would prevent any part of the large shock from being transmitted outside the cluster, it would also prevent any impact of the small shock from being diversified away from the cluster. To prevent failures following small shocks, clusters would then have to be larger, and more banks would fail following a large shock. Cabrales et al. (2014) compare the efficiency of various fully segmented network structures. Although we establish that fully segmented networks are not socially efficient, our clusters do share some important characteristics with these structures.[20]

Several observations sharpen the intuition behind Proposition 1.

REMARK **3.** In the social planner's solution:

(i) The optimal cluster size d^* is increasing in bankruptcy costs β, the size of the shocks ε_S and ε_L; and decreasing in firms' surplus value $R - \underline{v}$.

(ii) When $\varepsilon_S = \varepsilon^*$, all banks in the network are part of the same cluster and $d^* = n$.

(iii) The strength of firebreak links is decreasing in bankruptcy costs β, the size of shock ε_L; weakly decreasing in the size of the shock ε_S; and increasing in firms' surplus value $R - \underline{v}$.

The social planners' solution selects cluster networks and minimizes the size of clusters while preventing failures from small shocks. When the size of the small shock increases, or banks have less surplus value, larger clusters are required to absorb the shock. In the limiting case when the small shock is as large as $\varepsilon^* = n(R - \underline{v})$, the total surplus value in the network, all banks must be in the same cluster to prevent a failure. When large shocks are bigger, or bankruptcy costs are higher, preventing contagion from spreading beyond a cluster is also harder. Preventing contagion requires increasing the size of clusters, to keep more losses following a large shock within the cluster. The same forces require firebreak links to be weaker when large shocks are bigger or bankruptcy costs are higher, to continue preventing the transmission of contagion between clusters. Similarly, as the optimal cluster size increases, firebreak links must be weaker to prevent contagion, since following a large shock there are more total losses in the system, due to more bankruptcies.

Proposition 1 relies on several apparently restrictive assumptions. The support of ε has only two points, banks are of identical size, and the social planner is constrained to choosing from the class of fair networks. In section 5, we show that the key insight of Proposition 1—that firebreaks are socially efficient—is robust to relaxing these assumptions.

3.3 Shareholders vs. Creditors

The social planner aims to maximize the expected payments to shareholders *and* creditors. However, a crucial point underlying the rest of the paper is that, in a well-defined sense, the social planner's solution favors external creditors over shareholders. We illustrate this point with the following proposition.

PROPOSITION **2.** Consider the class of networks $\mathcal{A}'$ whereby no bank fails under a small shock. Abstracting from integer problems, the socially efficient networks $\mathcal{A}^* \subset \mathcal{A}'$ have the following properties:

(i) Any $A^* \in \mathcal{A}^*$ minimizes the sum of expected payments to shareholders in the class of networks $\mathcal{A}'$

[20] In different settings, Blume et al. (2011) and Erol and Vohra (2014) find that socially efficient networks are fully segmented.

(ii) Any $A^* \in \mathcal{A}^*$ maximizes the sum of expected payments to external creditors in the class of networks $\mathcal{A}'$

The social planner maximizes the expected sum of payments to all agents by minimising the expected number of defaults. External creditors receive back the face value of their debt if and only if their bank does not default. By minimising the likelihood of default, the social planner maximizes the probability that external creditors receive back the face value of their debt, so maximising their expected payments. In this sense, external creditors' incentives align with the social planner's. However, as discussed, the socially efficient network channels surplus value across the entire network to absorb shocks. Therefore shareholders have zero equity value in many states—in the socially efficient networks, shareholders' funds flow towards distressed banks instead of being retained as profit. Therefore the social planner aims to redistribute surplus from the *shareholders of healthy banks* to the *debtholders of distressed banks*. Hence the outside shareholders of bank i can raise their expected equity value by exchanging cross-holdings in ways that increase the probability of failure. For example, once bank i is at its failure threshold in some state, i's shareholders receive zero equity value in that state. All else equal, bank i suffers no further reduction in equity value from subsequently failing in this state, due to limited liability. Further falls in a bank's market value simply result in external debt write-downs. Concomitantly, bank i can raise its equity value in states when it does not fail, through making trades.

The fact that the socially optimal network favors creditors above shareholders is crucial to our analysis. It generates a split between society's and creditors' interests on the one hand, and shareholders' on the other. Moreover, we will assume in the next section that exclusively shareholders decide the structure of a bank's liabilities, so determining the network structure. Hence shareholders have incentives to form networks that differ from the socially optimal structures. Equivalently, a split between private and social interests in the network seems likely if external creditors do not have control rights over the decisions of organisations, and cannot impose "debtor discipline." Empirically, this failure of creditors to exact discipline on companies is well documented (Gorton and Santomero, 1990; Bliss and Flannery, 2002). Our aim is to introduce the shareholder/creditor conflict into a networked setting to explore its consequences for *systemic* risk, as opposed to merely its effects on a single firm. The systemic setting is important for the conflict of interests we identify. Recall that the social planner aims to redistribute surplus from the shareholders of healthy banks to the debtholders of distressed banks. As we go on to show in section 4, the tension between shareholders and creditors is indeed important, and means that the private trading networks can create more systemic risk than is socially optimal.

4 Equilibrium Networks

We assume that asset returns are unknown at the beginning of the period, although the distribution they are drawn from is common knowledge. Banks can exchange cross-holdings with one another. We assume that banks are risk-neutral equity value maximizers. In other words, bank i seeks to maximize

$$\mathbb{E}[\pi_i] = \sum_{\theta\in\Theta} P(\theta) \max\{\sum_{j\in N} A_{ij}(p_j(\theta) - \beta_j I_{v_j(\theta)<\underline{v}}) - \underline{v}, 0\}, \tag{7}$$

by trading direct cross-holdings with other banks[21], where $\theta \in \Theta = N \times \{S, L\}$ is a state of the world, and $P(\theta)$ is the probability of that state occurring. The different states of the world correspond to which bank is hit by a shock and the size of that shock. We denote the state in which a shock of size (amplitude) a hits bank k's asset by θ_k^a.

In our model, banks can form interconnections to *hedge* their risks. For example, consider two banks that have claims on each other. Suppose these claims are just insufficient for either to avoid failure following a small shock to their underlying assets and that failure costs are sufficiently large that both fail following a small shock to either. By increasing their claims on each other both can then avoid failing following a small shock and increase their expected equity values. More generally, by arranging their liabilities in certain ways, banks diversify away from exposure to their own project, and the idiosyncratic risk that it entails. Banks can therefore prevent failures in some states of the world, increasing the overall value in the system and potentially increasing their profits. Hedging seems to be a key reason for financial interconnections. To be sure, there exist numerous other causes, such as financial intermediation (Farboodi, 2014; Acemoglu et al., 2015b). Nevertheless financial flows spent on hedging instruments are extremely large (BIS, 2015b). Even if some of these flows are due to speculation, hedging is the main motivation for many key financial market participants.

Our model is well suited to assessing banks' medium term exposures, in the context of shocks to solvency. On the one hand, after short-run liquidity shocks in normal market conditions, banks have access to numerous ex post risk sharing opportunities, for example through wholesale lending markets. However over the medium term, hedging is important, because it allows banks to diversify away from their idiosyncratic investment risks.

4.1 External Creditors

Importantly, throughout the section on equilibrium networks, we assume that banks take the level of external liabilities as fixed. Equivalently, we assume that they are determined by some unmodelled process, before interbank claims are chosen. Therefore external liabilities cannot be adjusted after asset returns are realized. Nor can they be changed in response to the positions taken by the firm. Implicitly, external debt holders are unable to monitor the interbank contracts. Likewise, external debt contracts cannot be conditioned on the interbank network. This assumption is consistent with the *absence of debtor discipline*, given that external debtors do not determine banks' equilibrium behavior or the network structure.

[21] If contracts could be written directly on the underlying assets, banks would have no reason to ever have cross-holdings, as cross-holdings generate exposure to bankruptcy costs. Clearly, in our model banks can only write contracts on each others' *balance sheets*.

4.2 Trading Process and Stable Networks

First, we consider a very restrictive set of possible trades. We assume that two banks i and j can only exchange their cross-holdings in one another and that such trades are of symmetric size.

DEFINITION **1.** Crossholdings $\mathbf{C}' \in \mathcal{C}$ can be reached from $\mathbf{C}$ through a *feasible bilateral trade* between i and j if for all banks $k \in N$:

- $C_{kk} + C_{lk} = C'_{kk} + C'_{lk}$ for $k, l = i, j$ and $k \neq l$
- $C_{ii} - C'_{ii} = C_{jj} - C'_{jj}$
- $C_{kl} = C'_{kl}$ for either $k \neq i, j$ and/or $l \neq i, j$
- $\widehat{C}_{kk} = \widehat{C}'_{kk}$ for all k

We view the restriction to bilateral trading as appropriate for several reasons. Bilateral trading is intended to capture the decentralized nature of interbank over-the-counter markets. Detailed studies indicate that trading mainly takes place through bilateral lending relationships (e.g. Afonso et al., 2013), rather than multilateral arrangements. Moreover, in such markets trading partners typically trade contracts on one anothers' balance sheets, as opposed to instruments deriving value from third party banks.

We look for configurations of cross-holdings in which no pair of banks can raise their expected equity value by making a feasible bilateral trade, and refer to such configurations as stable. Conversely, if a pair of banks can raise expected equity value by making some feasible bilateral trade, then the current network is not stable.

DEFINITION **2.** A private trading network $\mathbf{C}$ is *stable* if and only if for every pair of banks $i, j \in N$, there are no feasible bilateral trades between i and j yielding crossholdings $\mathbf{C}'$ such that:

$$\mathbb{E}[\pi_i|\mathbf{C}'] > \mathbb{E}[\pi_i|\mathbf{C}] \text{ and } \mathbb{E}[\pi_j|\mathbf{C}'] > \mathbb{E}[\pi_j|\mathbf{C}].$$

4.3 Socially Efficient Networks and Private Instability

We next present a general result on the private stability of socially optimal networks.

PROPOSITION **3.** There exists a $\underline{\beta} > 0$ such that for all $\beta \in (0, \underline{\beta})$ no socially efficient network with at least two clusters and at most $N/2$ clusters is stable.

Proposition 3 shows that socially efficient networks are not stable for sufficiently small β. Even with a highly restrictive set of trades—that is, only symmetric trades in direct crossholdings between a pair of banks—we show that profitable trades away from socially efficient networks can be constructed. In Supplementary Appendix, Section 1, we show that Proposition 3 is driven by our dual assumptions that shareholders have limited liability and that there is an absence of debtor discipline. Relaxing either of these assumptions, we find that a socially efficient network is always stable.

To provide intuition for these results, we consider equation (7), replicated below for convenience.

$$\mathbb{E}[\pi_i] = \sum_{\theta \in \Theta} P(\theta) \max\{\sum_{j \in N} A_{ij}(p_j(\theta) - \beta_j I_{v_j(\theta) < \underline{v}}) - \underline{v}, 0\}.$$

The equation demonstrates the key importance of limited liability and the absence of debtor discipline, in creating a divergence between private and social incentives and generating greater systemic risk. Bank i's claims on some bank j's underlying assets, given by A_{ij}, have no value to i's shareholders in states where i fails, due to shareholder limited liability. Thus if i and j fail in the same states, i's shareholders are not exposed to the impact of j's bankruptcy costs. Bank i therefore prefers claims on banks who only fail in the same states as i, because of limited liability. More generally, bank i gains more equity value from claims on banks whose failures positively correlate with i's failure, all else equal. Importantly, when a set of banks fail in the same state, so that shareholders are protected by limited liability, it is the banks' external debt-holders who lose value from the resultant bankruptcy costs—due to debt write-downs. Therefore, due to limited liability, shareholders can redistribute value towards themselves and away from their external debt holders, by increasing claims on the counterparties with whom their failures are correlated. Since there is no debtor discipline in our model, in that only shareholders choose trades, external debt holders are unable to prevent shareholders from redistributing value in this way. We previously noted, in Proposition 2, that socially efficient networks act to redistribute value away from equity holders and towards external debt-holders. Limited liability allows equity holders to reverse this process. Overall, equity holders benefit from *systemic* risk, whereby banks fails at the same time as their largest counterparties; as opposed to idiosyncratic risk, whereby banks and their counterparties fail in different states.

In particular, in the proof of Proposition 3, banks i and j from different clusters deviate from the social optimum by executing a trade that lowers their cross-holdings of one another. The result is to increase both banks' claims on banks in their own clusters. In return, banks i and j reduce their claims on banks in each others' clusters.[22] Thus bank i becomes more heavily exposed to banks that default in the same states as i, namely those in its cluster. These bankruptcy costs do not affect i's equity value in such states, by limited liability. Conversely, bank i weakly reduces its exposure to banks in other clusters, and so is less exposed to the failure costs of these banks, in the states where i remains solvent but banks in other clusters fail. Hence i's overall equity value increases. Concomitantly, the trade lowers the expected value of bank i's external debt, as write downs become larger due to the greater exposure of i's external debt holders to counterparty shocks and bankruptcy costs in states where i fails. Furthermore the trade significantly increases the number of failures in certain states. However, i still finds the trade optimal since it already has zero equity value in all these states prior to the trade, so that the additional failures are not costly to i's shareholders at the margin due to limited liability. Thus i and j's privately optimal trade acts to generate greater systemic risk.

[22] We define bilateral trades to allow two banks to exchange cross-holdings in each other while preventing them from using these exchanges to directly increase their shareholder claims. Specifically, in the trade constructed to prove Proposition 3, we decrease i's holding in j, C_{ij}, and increase C_{ii} holding fixed $\widehat{C}_{ii}$. The alternative trade in which C_{ii} remains fixed and $\widehat{C}_{ii}$ increases is also profitable.

4.4 Trading and Systemic Risk

We have shown that trades can be profitable because banks prefer systemic risk to idiosyncratic risk—since banks aim to correlate their failures with their largest counterparties. This is the second key finding of our paper. To further illustrate this point, we relax our definition of feasible bilateral trades, to consider a much wider range of trades.

DEFINITION **3.** Crossholdings $\mathbf{C}' \in \mathcal{C}$ can be reached from $\mathbf{C}$ through a *relaxed bilateral trade* between i and j if:

- $C_{il} + C_{jl} = C'_{il} + C'_{jl}$ for all l
- $C_{kl} = C'_{kl}$ for $k \neq i, j$ and all l
- $\widehat{C}_{kk} = \widehat{C}'_{kk}$ for all k

In a relaxed bilateral trades i and j can exchange cross-holdings on any bank $k \neq i, j$ as well as in each other. These trades also no longer need to be symmetrically sized.

DEFINITION **4.** A private trading network $\mathbf{C}$ is *stable with respect to relaxed bilateral trades* if and only if for every pair of banks $i, j \in N$, there are no relaxed bilateral trades between i and j such that:

$$\mathbb{E}[\pi_i|\mathbf{C}'] > \mathbb{E}[\pi_i|\mathbf{C}] \qquad \text{and} \qquad \mathbb{E}[\pi_j|\mathbf{C}'] > \mathbb{E}[\pi_j|\mathbf{C}].$$

It is useful to introduce notation for the expected marginal value to i's shareholders of a claim on k's underlying assets:

$$\nu_{ik} := \sum_{\theta \in \Theta} P(\theta) \left(p_k(\theta) - \beta I_{v_k(\theta) < \underline{v}}\right) I_{v_i(\theta) \geq \underline{v}}. \tag{8}$$

Note that the expected marginal value of i's claim on j's underlying assets increases as i and j's failures become more closely correlated. Due to limited liability, when bank i fails counterparty bankruptcy costs are transferred to i's external creditors, absolving shareholders of the bankruptcy costs. Increased cross-holdings in a bank j increase indirect claims j's underlying asset, but also those of several other banks, as described by the $\mathbf{A}$ matrix. We therefore define the *personal price* shareholders of i place on a cross-holding in j by

$$p_{ij} := \sum_k A_{jk} \nu_{ik},$$

which captures the marginal value that i's shareholders place in a cross-holding in j—given the indirect claims on various assets k that the cross-holding in j entails, and noting that the value of these claims depend on the correlation of i and k's bankruptcies. The next result relates the stability of a financial network to the personal price ratios of banks.

PROPOSITION **4.** Suppose that in all states of the world no bank finds itself exactly on its failure threshold, that is, $v_k(\theta) \neq \underline{v}$ for all k and all θ. Also, suppose that $C_{il} > 0$ and $C_{jk} > 0$ for $i \neq j \neq k \neq l$. Then the network is stable with respect to relaxed bilateral trades only if $\frac{p_{il}}{p_{ik}} \geq \frac{p_{jl}}{p_{jk}}$.

Crucially, all else equal, p_{il} is higher when i's failure is highly correlated with the failures of the banks in which l has large indirect claims, since limited liability means that these claims are more valuable to i. Therefore, if possible, i and j will swap cross-holdings in third parties so that i and j's own failures become more closely correlated with the failures of banks on whom they have large claims. Proposition 4 emphasizes that equity holders value exposure to organizations whose failures correlate with their own. Equity holders therefore value the systemic risk generated by certain trades, since it allows them to default in the same states as the counterparties to whom they are most heavily exposed—thus emphasising our key result.

An immediate corollary of Proposition 4 is that there exist relaxed bilateral trades in which only claims on third parties are exchanged, which are profitable deviations from socially efficient networks. If only claims on third parties are exchanged, then we have $C_{kl} = C'_{kl}$ for any pair k, l such that $k \in \{i, j\}$ and $l \in \{i, j\}$.

COROLLARY **1.** There exists a $\underline{\beta} > 0$ such that for all $\beta \in (0, \underline{\beta})$, no socially efficient network with at least two clusters and at most $N/2$ clusters is stable with respect to relaxed bilateral trades between i and j in which only third party claims are exchanged.

In particular, if banks i and j are in different clusters they will want to trade in order to hold more cross-holdings on banks inside their own clusters. Thus bank i's value will be more reliant on the value of other banks in its cluster, who have more closely correlated failures.

Underlying Propositions 3 and 4 is that shareholders are indifferent between making no equity value while remaining in business, and defaulting—by definition, there is limited liability. Hence while bankruptcy is socially costly, it is not privately costly at the margin to shareholders. Thus situations exist in which a bank can raise its default probability *and* its expected equity value. While shareholders may sometimes prefer to raise a bank's probability of default, these trades are strictly worse for the bank's external creditors—who cannot act to prevent such trades, but will then face greater debt writedowns. Increasing the probability of default can therefore be privately optimal and socially suboptimal, and constitutes a conflict of interest between shareholders and bondholders. Moreover, this conflict creates incentives that favor systemic risk. We can therefore respond to the critique of Cochrane (2014), which hypothesizes that banks would endogenously act to prevent financial contagion arising from "domino effects". In our model, banks do not "build buffers against dominoes", in Cochrane's words. In fact, quite the opposite—given limited liability, banks *prefer* to correlate their failures with their largest counterparties, thus allowing domino effects to endogenously take place.

While our key result—that firms will seek to have correlated failures with their largest counterparties—is shown in a stylized contract space, using linear contracts, we expect similar findings to hold more generally. The key reason why a bank seeks to correlate failures with its largest counterparties is as follows. The value of a bank falls most when its largest counterparty fails, since the bank must absorb its counterparty's bankruptcy costs. Under limited liability, these bankruptcy costs are passed on to external creditors. It seems likely that both of these effects should operate with other contracts, such as the standard debt contracts of Eisenberg and Noe (2001).

The divergent interests of equity holders and debt holders reflects analysis from Merton (1974). Merton's classic insight was that under limited liability, equity is effectively a call option on a company's assets, with a strike price of the value of the company's debt. Consequently, it may well be optimal for shareholders to raise the probability of default under limited liability, since it increases risk and hence the option value of the equity claim. Notably, such a decision runs against the interests of creditors, whose claims lose value in the event of a default. Rochet (1992) and Gollier et al. (1997), amongst others, trace out the implications of these insights for the social efficiency of risk taking in individual banks. We study this divergence of interests in a systemic setting. Equity holders controlling the financial interconnections of their bank will deviate from socially optimal structure, by choosing interdependencies that generate systemic risk, rather than merely increasing idiosyncratic risk.

5 Firebreaks and Robustness

A key argument in our paper is that the social planner should introduce *firebreaks*, in order to enhance the resilience of financial systems. In Proposition 1, we demonstrate this result under several apparently restrictive assumptions, namely that the distribution of shock size ε has only two points of support, that all banks are of identical size, and that the social planner chooses from the set of fair networks (i.e. row stochastic $\mathbf{A}$ matrices). In the this section and in the Supplementary Appendix (Section 2), we argue that the importance of firebreaks is robust to relaxing these assumptions.

5.1 Shock Size Uncertainty

In this subsection we generalize our results by preventing the social planner from foreseeing the exact size of the small and large shock. Instead we suppose that the large shock is drawn from a distribution F_L with support $[\underline{\varepsilon}_L, \overline{\varepsilon}_L]$, the small shock is drawn from a distribution F_S with support $[\underline{\varepsilon}_S, \overline{\varepsilon}_S]$ and the probability of a large shock rather than a small shock hitting is drawn from a distribution F_q with support $[0, \overline{q}]$ for some $\overline{q} < \frac{1}{n^2}$. As before, each shock hits a single bank's underlying assets and that bank is selected uniformly at random. The social planner seeks to maximize the sum of equity and debt holder value. We consider the loss a social planner would incur by implementing the optimal network as if the large shock were $\overline{\varepsilon}_L$ for sure, the small shock were $\overline{\varepsilon}_S$ for sure and the probability of the large shock were $\overline{q}$. We abuse notation by writing $\mathcal{A}^*(\varepsilon_S, \varepsilon_L, q)$ to represent the set of networks $\mathcal{A}^*\left(d^*(\varepsilon_S, \varepsilon_L, q)\right)$.

PROPOSITION **5.** For all $\overline{q} < \frac{1}{n^2}$, any distribution F_L, any distribution F_S and any distribution F_q, the percentage of social surplus lost by a social planner selecting any network $\mathbf{A}^* \in \mathcal{A}^*(\overline{\varepsilon}_S, \overline{\varepsilon}_L, \overline{q})$ is bounded from above by:

$$\frac{\overline{q}\beta\left(d^*(\overline{\varepsilon}_S, \overline{\varepsilon}_L) - d^*(\underline{\varepsilon}_S, \underline{\varepsilon}_L)\right)}{nR - (1-\overline{q})\underline{\varepsilon}_S - \overline{q}\left(\underline{\varepsilon}_L + \beta d^*(\underline{\varepsilon}_S, \underline{\varepsilon}_L)\right)}.$$

Proposition 5 bounds the percentage losses incurred by a social planner implementing a network $\mathbf{A}^* \in \mathcal{A}^*(\overline{\varepsilon}_S, \overline{\varepsilon}_L, \overline{q})$ instead of an optimal network, for distributions F_L, F_S and F_q. Moreover, the upper bound on the losses from imposing network $\mathbf{A}^* \in \mathcal{A}^*(\overline{\varepsilon}_S, \overline{\varepsilon}_L, \overline{q})$ only occurs when the large shock has size $\underline{\varepsilon}_L$ for sure, and the small shock has size $\underline{\varepsilon}_S$ for sure. When there is uncertainty over shock sizes, that is $F_S(\underline{\varepsilon}_S), F_L(\underline{\varepsilon}_L) < 1$, the social planner loses strictly less than the bound on losses established in proposition 5, by choosing some network $\mathbf{A}^* \in \mathcal{A}^*(\overline{\varepsilon}_S, \overline{\varepsilon}_L, \overline{q})$. Therefore the clustered networks we previously characterized are likely to be close to socially efficient, as long as there is not too much uncertainty about the size of the large or small shocks. We conclude that the general trade-off facing the social planner faces remains relevant beyond the very simple two point support shocks we considered earlier, as does the importance of introducing firebreaks.

Using Proposition 5, we can also solve the optimal network design problem for an ambiguity averse social planner. Suppose the social planner knows the support of F_L, F_S and F_q but not the distributions. Furthermore, suppose the social planner is ambiguity averse insofar as her objective is to maximize the minimum possible social surplus, so that she solves

$$\max_{\mathbf{C}} \min_{\varepsilon_S \in [\underline{\varepsilon}_S, \bar{\varepsilon}_L], \varepsilon_L \in [\underline{\varepsilon}_L, \bar{\varepsilon}_L], q \in [0, \bar{q}]} \mathbb{E}\left[\sum_{i \in N} \pi_i(\mathbf{C}) + \delta_i(\mathbf{C})\right] \qquad \text{subject to} \quad \sum_{j \in N} A_{ij} = 1 \text{ for all } i.$$

COROLLARY **2.** For all $\bar{q} < \frac{1}{n^2}$, any distributions F_L, F_S and F_q the network $\mathbf{A}^*(\bar{\varepsilon}_S, \bar{\varepsilon}_L, \bar{q})$ is optimal for an ambiguity averse social planner.

Corollary 2 reinforces the key insights from Proposition 1. Indeed ambiguity aversion, as modelled, might provide a reasonable objective function for policy-makers such as central banks.

5.2 Adding Heterogeneity

So far we have only considered homogeneous banks, which are ex-ante identical. All banks are assumed to have claims on returns R, to have the same bankruptcy costs and the same failure threshold. In this section we will demonstrate how networks in $\mathcal{A}^*$, as defined in Proposition 1, can be extended to incorporate heterogeneous bank sizes, and show that they remain socially optimal. To show this result, we return to the setup of Proposition 1, so that the shock size again takes only two values.

DEFINITION **5.** Consider a financial system $(\mathbf{A}, N)$ and set of banks $M \subseteq N$, to be merged into a new bank i. The financial system $(\mathbf{A}', N')$ results from a (M, i)-merger of $(\mathbf{A}, N)$ if:

- $A'_{ij} = \sum_{k \in M} A_{kj}$ and $A'_{ji} = \frac{1}{|M|} \sum_{k \in M} A_{jk}$ for all $j \in N$.
- $A'_{jk} = A_{jk}$ for $j, k \notin M$.
- $\beta_i = \sum_{m \in M} \beta_m, \quad \underline{v}_i = \sum_{m \in M} \underline{v}_m, \quad p_i = \sum_{m \in M} p_m.$
- $N' = (N \cup \{i\}) \setminus M.$

Therefore a bank i created from the merger of banks M inherits the cumulative claims of the banks in M. The merger preserves the overall claims of other banks and of outside shareholders. Overall bankruptcy costs, asset values and failure thresholds are also preserved. Finally, we preserve the same overall shock probability. As before, a random, firm specific shock ε hits the underlying asset of one bank. However, the probability of a given asset being hit by a shock is now proportional to the *size* of bank. We index the size of a bank by the value of its asset in the absence of a shock, given by the value of the assets belonging to the previously merged banks. Bank i is of size σ_i if its asset pays a return of $\sigma_i R$ in the absence of any shocks. Hence the probability of a small shock hitting bank i is

$(1-q)\left(\frac{\sigma_i}{\sum_{j\in N}\sigma_j}\right)$ and the probability of a large shock hitting bank i is $q\left(\frac{\sigma_i}{\sum_{j\in N}\sigma_j}\right)$. By construction $\mathbf{A}'$ is column stochastic, so is still a valid dependency matrix. For ease of exposition we state mergers in terms of the dependency matrix $\mathbf{A}$. These dependencies emerge naturally from merging cross-holding claims (see the Supplementary Appendix, Section 4.2).

We will consider sequences of mergers between organizations that are in the same cluster. We assume the social planner continues to take the size of organizations as given and chooses a financial network to maximize the expected sum of payments to creditors and organisations, $\mathbb{E}\left[\sum_{i\in N}\pi_i+\delta_i\right]$, while choosing from within the class of fair networks (i.e. networks where the $\mathbf{A}$ matrix is row stochastic).

REMARK **4.** Following mergers, the social planner aims to minimize the expected sum of default costs rather than simply the number of organizations that fail.

We can then extend our analysis of the social planner's problem to include size heterogeneity. Suppose homogeneous banks are connected by a socially optimal financial network, $\mathbf{A}^*\in\mathcal{A}^*$ as defined in Proposition 1. We say that a merger is within-cluster when it comprises of only firms from the same cluster. Following such a merger, we label the newly created bank as belonging to the same cluster as its constituent parts.

PROPOSITION **6.** Consider a socially optimal financial network $\mathbf{A}^*$, as defined in Proposition 1, connecting homogeneous banks. For any sequence of within-cluster mergers the resulting network $\mathbf{A}'$ is still socially optimal.

An immediate implication of the proof of Proposition 6 is that the opposite exercise to merging organizations can also be considered. We can contemplate splitting an organization i into two identical copies i_1, i_2, such that organization i would result from i_1 merging with i_2. Note that if we took a network $\mathbf{A}^*$ and did this with all organizations, the optimal cluster size would double to $2d^*$. Repeatedly this exercise k times, we would have $2^k d^*$ organizations in each cluster. We could then apply Proposition 6 to create different sized organizations with much more variation than initially possible. Indeed, for sufficiently large k—after splitting each organization sufficiently many times—we can reconstruct a socially optimal network arbitrarily close to any size distribution within a cluster. Thus although the size heterogeneity allowed by Proposition 6 allows is very structured and initially appears quite restrictive, it accommodates size heterogeneity in a relatively flexible manner.

Figure 3 illustrates a socially optimal network in which there are three clusters, each with five banks. Within each cluster there is one bank that is as large as all other banks within the cluster combined. We term the relatively large organizations as *core* banks and other organizations as *periphery* banks. Each core banks could have been created by the merger of 4 periphery banks. The width of the lines represents the strength of dependencies. With the heterogeneous sizes the strongest links end up being across cluster, between the core banks—these dependencies are equal to 16 weak relationships between periphery banks from different clusters. If we focus attention on only sufficiently strong links, then the socially optimal networks associated with significant size heterogeneity closely resemble the core-periphery networks often observed in practice. However, an important property of the socially efficient networks is that there is no contagion between clusters—ensured by the relatively

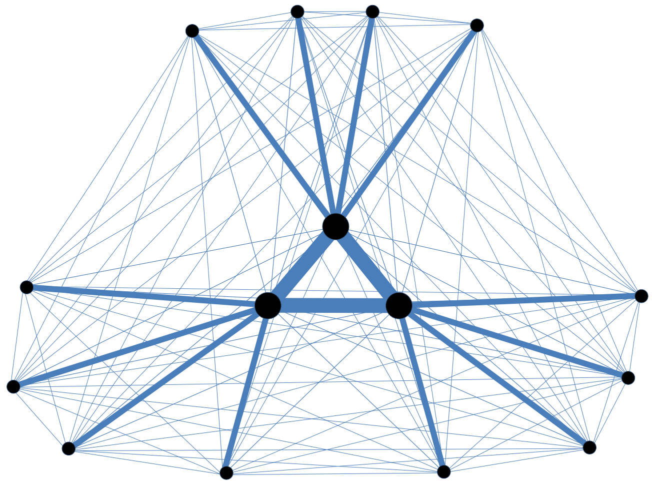

Figure 3: A socially optimal network with size heterogeneity.

weak strength of the firebreak links between clusters. Therefore, even though the strongest links are between core banks, contagion will not spread between core banks and no two core banks will fail in the same state of the world.

5.3 More Shock Sizes

To further emphasize the importance of firebreaks, we now consider a more general discrete shock distribution. Let ε be distributed such that $\varepsilon = \varepsilon_w$ with probability q_w for $w = 1, ..., \bar{w}$. We continue to assume that the shock hits one bank's proprietary asset and that the bank hit is selected uniformly at random. Without loss of generality we let $\varepsilon_w < \varepsilon_{w+1}$ for all $w < \bar{w}$.

Let D_i^w be the set of banks that fail following a shock of size ε_w to i and set[23] $d_i^w = |D_i^w|$. For notational convenience later, define $D_i^{\bar{w}+1} \equiv N$ for all i and $D_i^0 = \emptyset$ for all i. The social planner's problem is

$$\min_{\mathbf{A}\in\mathcal{A}} \sum_{i\in N} \frac{1}{n} \sum_{w=1}^{\bar{w}} q_w d_i^w \quad \text{subject to} \sum_{k\in N} A_{jk} = 1 \text{ for all } j \in N, \tag{9}$$

so that the social planner chooses from the class of fair networks as before. Let $\mathbf{A}^*$ be a network that solves the social planner's problem and let $\mathbf{d}_i^*(\mathbf{A}^*) \in \mathcal{N}^{\bar{w}}$, with wth element $d_i^{w*}(\mathbf{A}^*)$, denote the number of failures that result from shocks of different sizes hitting bank i when $\mathbf{A}^*$ is implemented.

To describe the social planner's solution, we need to define a family of networks.

DEFINITION **6.** We define $\bar{w}$-*layered-clustered* networks as follows:

- For all $w = 1, ..., \bar{w}$, the banks are partitioned into sets of equal cardinality. We let G_i^w denote the partition element, or set of banks, that i belongs to.

[23]Note that for all $w' < w$, $D_i^{w'} \subseteq D_i^w$, given that for a given network, each bank's value is weakly decreasing in the size of the realized shock.

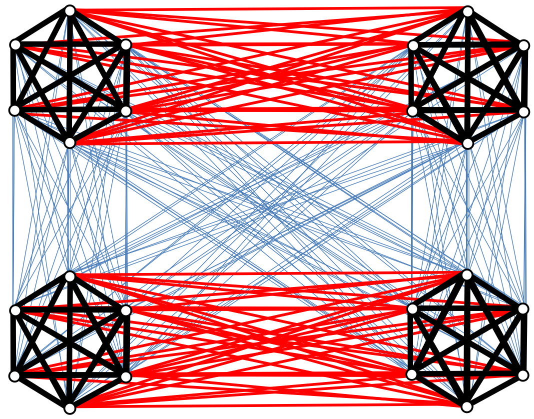

Figure 4: A layered-clustered network. Thicker lines represent stronger dependencies.

- For all $i \in N$ and for all $w \geq w'$, $G_i^w \supseteq G_i^{w'}$
- For all $w = 1, ..., \bar{w}$, and for all i and all j such that $j \in G_i^w \backslash G_i^{w-1}$, $A_{ij} = \bar{A}_w \leq \bar{A}_{w-1}$.

See figure 4 for an example. Intuitively, banks are arranged into *layers*. In layer 1 banks are partitioned into small groups that have high dependencies $\bar{A}_1$ on each other. In layer 2, banks are more coarsely partitioned. These partitions are created by combining layer 1 groups. The dependencies between banks in the same layer 2 groups, but different layer 1 groups, are $\bar{A}_2 \leq \bar{A}_1$. Layer 3 groups are then constructed by combining layer 2 groups, and so on. It will be convenient to define $G_i^{\bar{w}+1} \equiv N$ for all i and $G_i^0 = \emptyset$ for all i. We will use this class of networks to describe the solution to the social planner's problem.

To derive the socially efficient networks, we first examine a relaxed social planner's problem, and then find conditions under which the solution also solves the full social planner's problem. If a shock hits bank i, by definition banks $j \in D_i^w \backslash D_i^{w-1}$ for $w = 1, \ldots, \bar{w}+1$ do not fail after a shock ε_{w-1} and so

$$A_{ji}\varepsilon_{w-1} + \sum_{k \in D_i^{w-1}} A_{jk}\beta \leq R - \underline{v}. \tag{10}$$

These inequalities provide some constraints on $\mathbf{A}$. In addition, any feasible network $\mathbf{A}$ also satisfies column stochasticity. We define the relaxed social planner's problem as follows. The social planner knows which bank will be hit by a shock, though not the size of the shock; and chooses the number of banks that fail subject to choosing a network that satisfies equations (10), row stochasticity and column stochasticity. Therefore, fixing some i, the social planner's relaxed problem is:

$$\min_{\mathbf{d}_i \in \mathcal{N}^{\bar{w}}} \sum_{w=1}^{\bar{w}} q_w d_i^w \tag{11}$$

subject to

(i) $d_i^w = |D_i^w(\mathbf{A})|$ for some network $\mathbf{A}$ satisfying the conditions $(ii) - (iv)$.

(ii) $A_{ji}\varepsilon_{w-1} + \sum_{k \in D_i^{w-1}} A_{jk}\beta \leq R - \underline{v}$. for all $j \in D_i^w$ and all $w = 1, \ldots, \bar{w}$.

(iii) $\sum_k A_{kj} = 1$ for all $j \in N$.

(iv) $\sum_k A_{jk} = 1$ for all $j \in N$.

We denote the solution to this problem with the vector $\mathbf{d}_i' \in \mathcal{N}^{\bar{w}}$. This formulation relaxes two constraints in the unrelaxed problem. First, the social planner can condition on the shock hitting a given bank i. Second, the social planner chooses the dependency network $\mathbf{A}$ without requiring it to correspond to a cross-holding network $\mathbf{C}$.

We say that $\mathbf{d}_i'$ satisfies a *generalized integer assumption* if two conditions hold. Firstly, $d_i^{w+1'} = k_w d_i^{w'}$ where $k_w \in \mathcal{N}$ for all w. Secondly, the social planner's solution cannot be improved upon by implementing any sequence of demergers, in which which each bank is split arbitrarily many times into two identical banks.

The first of these conditions requires that the minimal number of banks failing after any shock to i exactly divide one another. To understand the second of these conditions, recall that in subsection 5.2, we demonstrated that a bank could be split into two identical banks in a way that preserved the claims of other banks on underlying assets and such that the two banks collectively had the same claims as the initial bank. When the social planner's solution does not change after arbitrarily many such demergers, there is no "spare" equity value that can be reallocated to reduce the number of failures. In effect, the social planner can choose $\mathbf{d}_i' \in \mathbb{R}_+^{\bar{w}}$ instead of $\mathbf{d}_i' \in \mathcal{N}^{\bar{w}}$.

PROPOSITION **7.** Under the generalized integer assumption, the $\bar{w}$-layered-clustered network defined by

- $|G_i^w| = d_i^{w'}$ for all w and for all i
- $A_{ji} = \frac{R - \underline{v}}{\varepsilon_{w-1} + \beta d_i^{w-1'}}$ for all $j \in G_i^w \backslash G_i^{w-1}$, for all $i \in N$, and for all w

solves the social planner's problem.

Proposition 7 shows that under relatively strong integer assumptions[24] a layered-clustered network solves the social planner's problem with multiple shocks.[25] Given this structure, the social planner now imposes *multiple* firebreaks, where firebreaks now refer to the links across each layer as in figure 4. Thus we show that even with multiple shocks, firebreaks are still optimal.

The reason for firebreaks is the same. Suppose a given shock causes some failures in the network. Firebreaks optimally insulate the rest of the network from contagion, from both the shock itself, and the associated bankruptcy costs. The proposition also introduces a new insight. Efficient networks can have a fractal-like quality. When there are multiple distinct layers, the layered-clustered network

[24]The conditions in the proof require that the size of the default set in the relaxed problem, after a shock of a given size, exactly divides the size of the default sets for larger shocks. Thus the generalized integer assumption does more work as the number of layers increases.

[25] Unlike in Section 3 we are not able to characterize the optimal number of failures for shocks of different sizes. Rather, we show that minimising the expected number of failures can be achieved by a $\bar{w}$-layered-clustered network.

looks similar at different scales, with strong links to the layers corresponding to smaller shocks, and weak links to the layers corresponding to larger shocks. Thus when a shock of size ε_w hits any bank i in the network only banks in G_i^w fail, and the rest are unaffected—allowing the social planner to lower the expected number of failures. For example, suppose that the network in figure 4 is optimal for shock sizes $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$. After the largest shock hits a bank i, all the banks except those with the weakest links to i will fail. After an intermediate-sized shock, only the banks with the strongest connections to i will fail. After a small shock, no banks will fail.

In general, for some shock sizes w, we can have $|G_i^w| = |G_i^{w-1}|$, that is, different shock sizes can correspond to layers containing the same banks. The number of distinct layers depends on the model parameters. Proposition 7 also nests our previous results on firebreaks in Proposition 1. For example, if we have $\sum_{w:\varepsilon_w<\varepsilon^*} q_w > \frac{1}{n^2}$ and the set $\varepsilon_w : \varepsilon_w > \varepsilon^*$ is a singleton, then a socially efficient network is given by the simple clustered structure discussed in Subsection 3.2. Also, if all shocks are small, that is, $\varepsilon_{\bar{w}} \leq \varepsilon^*$, then the socially efficient network is complete—conversely, if all shocks are large, so that $\varepsilon_1 > \varepsilon^*$, then the socially efficient network is empty. Of course, the proposition is also more general. For example, if we have $\sum_{w:\varepsilon_w<\varepsilon^*} q_w > \frac{1}{n^2}$, and only two large shocks, then a structure similar to figure 4 is possible, for certain values and probabilities of the two large shocks. Nevertheless, the key point is that firebreaks remain optimal when we consider more general shock distributions.

6 Conclusion

The recent financial crisis has emphasized the importance of understanding how the structure of financial networks might generate systemic risk. A key question is how networks should be structured to minimize systemic risk—and why privately optimal behavior, in a decentralized financial system, might depart from this and generate excess systemic risk. Clearly, we need to understand the reasons behind any putative divergence between private and social incentives when considering regulations.

This paper seeks to explain why banks might generate systemic risk in a networks setting, by emphasising the interaction between limited liability and the absence of debtor discipline. In our first set of results, we derive the social planner's solution, in an environment of both rare, large shocks; and common, small shocks. We show that, in general, socially efficient networks are characterized by *firebreaks*. That is, in the optimal network structure banks are partitioned into sets that we term clusters, with strong interdependencies within clusters and weak dependencies across clusters. This structure prevents failures after small shocks, and also minimizes the spread of contagion after large shocks. Thus our result formalizes the conjecture by Haldane (2010), that the architecture of socially efficient financial networks should exhibit *modularity*.

In our second set of results, we demonstrate that limited liability and the absence of debtor discipline will mean socially efficient networks are typically not stable. Instead, they cause banks to generate excess systemic risk, by trading in order to correlate failures with their largest counterparties. Limited liability prevents failure from being costly at the margin to equity holders. Banks can then make profitable trades that increase their exposures to their largest counterparties. While a counterparty failing is costly for the equity holders of a bank that is solvent, it is not costly for the equity holders of a bank that fails in the same states. Thus shareholders prefer systemic risk to idiosyncratic risk, in that they prefer to fail in the same state as their largest counterparties. In this sense, we show that banks have incentives to endogenously allow inefficient "domino effects" to occur.

It is well understood that banks may seek to correlate their risks, due to the possibility of government intervention (e.g. Farhi and Tirole, 2012). Our findings show that banks seek to correlate their failures even in the absence of government intervention—since they prefer to fail in the same states as their largest counterparties. Therefore our analysis suggests different policy implications, around breaking the confluence of shareholder control and limited liability, which is the underlying cause of excessive systemic risk in our model. According to our results, one set of useful policies involves reducing the hazardous effects of limited liability, by making default costly for banks at the margin[26]. Our findings also support policies that aim to prevent the breakdown of debtor discipline, by ensuring that some degree of control rights belong to creditors, especially in a distressed financial system. Certain forms of convertible debt (e.g. Bulow and Klemperer, 2015) might abet this outcome. Moreover, our work also helps justify policies that aim to rectify systemic risk directly, through our analysis of firebreaks. These arguments provide a justification for regulation that explicitly limits the size of bilateral exposures between different counterparties, in the interests of reducing systemic risk[27].

Finally, our paper suggests empirical research. We have found that modularity is an important feature of socially efficient networks. An key question is how the degree of modularity in financial systems has varied in recent history, or across different financial systems. Equally, it might be fruitful to investigate the impact of changes in some measure of modularity on levels of systemic risk.

[26]Our model therefore provides support for various pieces of financial regulation. For example, Title II of the Dodd-Frank Act contains a "clawback" provision in the event of failure, allowing the FDIC to recover previously paid executive compensation. Similarly, the Squam Lake Report (French et al., 2010) recommends that executives in systemically important institutions forfeit compensation in the event of bankruptcy.

[27]An example of such legislation is the single counterparty exposure restriction in the Dodd Frank Act.

References

ACEMOGLU, D., A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2015a): "Systemic Risk and Stability in Financial Networks," *American Economic Review*, 105, 564–608.

ACEMOGLU, D., A. E. OZDAGLAR, AND A. TAHBAZ-SALEHI (2015b): "Systemic risk in endogenous financial networks," *Available at SSRN 2553900*.

AFONSO, G., A. KOVNER, AND A. SCHOAR (2013): "Trading partners in the interbank lending market," .

ALLEN, F., A. BABUS, AND E. CARLETTI (2012): "Asset commonality, debt maturity and systemic risk," *Journal of Financial Economics*, 104, 519–534.

ALLEN, F. AND D. GALE (2000): "Financial contagion," *Journal of political economy*, 108, 1–33.

ALVAREZ, F. AND G. BARLEVY (2014): "Mandatory disclosure and financial contagion," Tech. rep., Working Paper, Federal Reserve Bank of Chicago.

BABUS, A. (2013): "The formation of financial networks," .

BIS (2015a): "Central clearing : trends and current issues 1," 59–76.

——— (2015b): "Semiannual OTC derivatives statistics," .

BLISS, R. R. AND M. J. FLANNERY (2002): "Market discipline in the governance of US bank holding companies: Monitoring vs. influencing," *European Finance Review*, 6, 361–396.

BLUME, L., D. EASLEY, J. KLEINBERG, R. KLEINBERG, AND É. TARDOS (2011): "Network formation in the presence of contagious risk," in *Proceedings of the 12th ACM conference on Electronic commerce*, ACM, 1–10.

BRUNNERMEIER, M. K. AND M. OEHMKE (2012): "Bubbles , Financial Crises , and Systemic Risk ," 1–89.

BULOW, J. AND P. KLEMPERER (2015): "Equity Recourse Notes: Creating Counter-cyclical Bank Capital," *forthcoming, Economic Journal*.

CABALLERO, R. J. AND A. SIMSEK (2013): "Fire sales in a model of complexity," *The Journal of Finance*, 68, 2549–2587.

CABRALES, A., D. GALE, AND P. GOTTARDI (2015): "Financial contagion in networks," *forthcoming, Oxford Handbook on the Economics of Networks*.

CABRALES, A., P. GOTTARDI, AND F. VEGA-REDONDO (2014): "Risk-sharing and contagion in networks," *CESifo Working Paper Series*.

CIFUENTES, R., G. FERRUCCI, AND H. S. SHIN (2005): "Liquidity risk and contagion," *Journal of the European Economic Association*, 3, 556–566.

COCHRANE, J. H. (2014): "Toward a run-free financial system," *Available at SSRN 2425883*.

COHEN-COLE, E., E. PATACCHINI, AND Y. ZENOU (2015): "Static and dynamic networks in interbank markets," *Network Science*, 98–123.

DAVYDENKO, S. A., I. A. STREBULAEV, AND X. ZHAO (2012): "A market-based study of the cost of default," *Review of Financial Studies*, hhs091.

DENBEE, E., C. JULLIARD, Y. LI, AND K. YUAN (2014): "Network Risk and Key Players : A Structural Analysis of Interbank Liquidity ," .

DI MAGGIO, M. AND A. TAHBAZ-SALEHI (2014): "Financial intermediation networks," *Columbia Business School Research Paper*.

DIAMOND, D. W. AND P. H. DYBVIG (1983): "Bank runs, deposit insurance, and liquidity," *The journal of political economy*, 401–419.

EISENBERG, L. AND T. H. NOE (2001): "Systemic risk in financial systems," *Management Science*, 47, 236–249.

EISERT, T. AND C. EUFINGER (2013): "Interbank network and bank bailouts: Insurance mechanism for non-insured creditors?" .

ELLIOTT, M., B. GOLUB, AND M. O. JACKSON (2014): "Financial Networks and Contagion," *American Economic Review*, 104, 3115–3153.

EROL, S. AND R. VOHRA (2014): "Network Formation and Systemic Risk, Second Version," *PIER Working Paper*.

FARBOODI, M. (2014): "Intermediation and voluntary exposure to counterparty risk," Tech. rep., Working Paper, University of Chicago.

FARHI, E. AND J. TIROLE (2012): "Collective Moral Hazard, Maturity Mismatch, and Systemic Bailouts," *The American Economic Review*, 102, 60–93.

FREIXAS, X., B. M. PARIGI, AND J.-C. ROCHET (2000): "Systemic risk, interbank relations, and liquidity provision by the central bank," *Journal of money, credit and banking*, 611–638.

FRENCH, K., M. BAILY, J. CAMPBELL, J. COCHRANE, D. DIAMOND, D. DUFFIE, A. KASHYAP, F. MISHKIN, R. RAJAN, D. SCHARFSTEIN, ET AL. (2010): "The squam lake report: Fixing the financial system*," *Journal of Applied Corporate Finance*, 22, 8–21.

GAI, P. AND S. KAPADIA (2010): "Contagion in financial networks," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, rspa20090410.

GALEOTTI, A., C. GHIGLINO, AND S. GOYAL (2015): "Financial Linkages, Portfolio Choice and Systemic Risk," .

GLASSERMAN, P. AND H. P. YOUNG (2015): "How likely is contagion in financial networks?" *Journal of Banking & Finance*, 50, 383–399.

GOLLIER, C., P.-F. KOEHL, AND J.-C. ROCHET (1997): "Risk-taking behavior with limited liability and risk aversion," *Journal of Risk and Insurance*, 347–370.

GORTON, G. AND A. M. SANTOMERO (1990): "Market discipline and bank subordinated debt: Note," *Journal of Money, Credit and Banking*, 119–128.

HALDANE, A. G. (2009): "Rethinking the financial network," *Speech delivered at the Financial Student Association, Amsterdam, April.*

——— (2010): "The hundred billion dollar question," *Speech at the Institute of Regulation and Risk, North Asia (IRRNA), in Hong Kong.*

JAMES, C. (1991): "The losses realized in bank failures," *The Journal of Finance*, 46, 1223–1242.

JENSEN, M. C. AND W. H. MECKLING (1976): "Theory of the firm: Managerial behavior, agency costs and ownership structure," *Journal of financial economics*, 3, 305–360.

JONES, E. P., S. P. MASON, AND E. ROSENFELD (1984): "Contingent claims analysis of corporate capital structures: An empirical investigation," *Journal of Finance*, 611–625.

LEITNER, Y. (2005): "Financial networks: Contagion, commitment, and private sector bailouts," *The Journal of Finance*, 60, 2925–2953.

MERTON, R. C. (1974): "On the pricing of corporate debt: The risk structure of interest rates*," *The Journal of Finance*, 29, 449–470.

PETERSEN, K. B., M. S. PEDERSEN, ET AL. (2008): "The matrix cookbook," *Technical University of Denmark*, 7, 15.

ROCHET, J.-C. (1992): "Capital requirements and the behaviour of commercial banks," *European Economic Review*, 36, 1137–1170.

TARULLO, D. K. (2015): "Annual Conference on Evaluating Macroprudential Tools: Complementarities and Conflicts Arlington, Virginia January 30, 2015," .

WANG, J. J. (2014): "Distress Dispersion and Systemic Risk in Networks," *Available at SSRN.*

ZAWADOWSKI, A. (2013): "Entangled financial systems," *Review of Financial Studies.*

A Omitted Proofs

A.1 Proof of Remark 1

The social planner's problem is to maximize:

$$\mathbb{E}\left[\sum_{i\in N}\pi_i+\delta_i\right] \text{ subject to } \sum_{i\in N}A_{ij}=1 \text{ for all } j\in N$$

Using equation (7) we have that:

$$\begin{aligned}
\mathbb{E}\left[\sum_{i\in N}\pi_i+\delta_i\right] &= \mathbb{E}\left[\sum_{i\in N}\max\{v_i-\underline{v},0\}+\min\{\underline{v},v_i\}\right]\\
&= \mathbb{E}\left[\sum_{i\in N}v_i\right]\\
&= \mathbb{E}\left[\sum_{i\in N}\sum_{j\in N}A_{ij}(p_j-\beta I_{v_j<\underline{v}})\right]\\
&= \sum_{i\in N}\sum_{j\in N}A_{ij}\mathbb{E}[p_j]-\beta\sum_{i\in N}\sum_{j\in N}A_{ij}\mathbb{E}[I_{v_j<\underline{v}}]
\end{aligned}$$

Noting that $\mathbb{E}[p_j]=R-\varepsilon/n$; and that $\sum_{i\in N}A_{ij}=1$ because the dependency matrix is column stochastic; it follows that:

$$\begin{aligned}
\mathbb{E}\left[\sum_{i\in N}\pi_i+\delta_i\right] &= nR-\varepsilon-\beta\mathbb{E}\left[\sum_{j\in N}I_{v_j<\underline{v}_j}\right]\\
&= nR-\varepsilon-\beta\mathbb{E}[\textit{no. of defaults}]
\end{aligned}$$

Since all but the final term of the above equation is exogenously given, the social planner maximizes $\mathbb{E}\left[\sum_{i\in N}\pi_i+\delta_i\right]$ by minimising the expected number of defaults. □

A.2 Proof of Lemma 1

For any network $\mathbf{A}$ there is a minimum set of organizations $D_i(\mathbf{A})$ that will fail following a large shock to i. Letting $d_i(\mathbf{A})=|D_i(\mathbf{A})|$, we seek to minimize the number of organizations that fail when a large shock hits i, subject to the constraint that no organizations ever fail following a small shock *to any organization*. Where there should be no confusion we abuse notation and drop the arguments of functions.

$$\text{P1:}\quad \min_{\mathbf{A}\in\mathcal{A}} d_i(\mathbf{A})\quad \text{subject to}\quad A_{jk}\varepsilon_S\le R-\underline{v}\quad \text{for all } j,k\in N$$

As organizations $j\notin D_i$ do not fail following a large shock to i:

$$A_{ji}\varepsilon_L+\sum_{k\in D_i}A_{jk}\beta\le R-\underline{v}. \tag{12}$$

Thus an upper bound on the losses absorbed by banks not in D_i after a large shock to i is

$$(n - d_i)(R - \underline{v}).$$

The remaining losses must be absorbed by the remaining banks. Thus collectively organizations in D_i incur losses of at least

$$\varepsilon_L + d_i\beta - [n(R - \underline{v}) - d_i(R - \underline{v})],$$

and so

$$\sum_{j \in D_i} \left[\sum_{k \in D_i} A_{jk}\beta + A_{ji}\varepsilon_L \right] \geq \varepsilon_L + d_i\beta - [n(R - \underline{v}) - d_i(R - \underline{v})]. \tag{13}$$

No organization fails following a small shock to i, so by the constraints in P1:

$$\sum_{j \in D_i} A_{ji} \leq \frac{d_i(R - \underline{v})}{\varepsilon_S}. \tag{14}$$

Moreover, by the constraints in P1 none of the organizations $j \in D_i$ can fail when a small shock hits any $k \in D_i$. The above condition must therefore hold when a small shock hits any $k \in D_i$. Thus

$$\sum_{j \in D_i} \sum_{k \in D_i} A_{jk} \leq \frac{d_i^2(R - \underline{v})}{\varepsilon_S}. \tag{15}$$

Combining inequalities (13), (14) and (15) we have:

$$\varepsilon_L + d_i\beta - [n(R - \underline{v}) - d_i(R - \underline{v})] \leq \frac{d_i^2(R - \underline{v})}{\varepsilon_S}\beta + \sum_{j \in D_i} A_{ji}\varepsilon_L \leq \frac{d_i^2(R - \underline{v})}{\varepsilon_S}\beta + \frac{d_i(R - \underline{v})}{\varepsilon_S}\varepsilon_L.$$

Rearranging:

$$f(d_i) := d_i^2(R - \underline{v})\beta + d_i\left((R - \underline{v})(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta\right) + \varepsilon_S\left(n(R - \underline{v}) - \varepsilon_L\right) \geq 0. \tag{16}$$

The constant term of the quadratic equation is always negative, according to the definition of ε_L; and the quadratic coefficient is always positive according to the constraint in P1. It follows from the quadratic formula that $f(d_i)$ always has exactly one positive real root. Moreover, for values of d_i above this real root $f(d_i)$ is increasing in d_i. Thus there is a unique minimized value of d_i that satisfies the constraint $f(d_i) \geq 0$ denoted d_i^*, and it is implicitly defined by the positive solution to $f(d_i) = 0$. Thus, for any network structure that prevents any bank failing following a small shock to i, at least $\lceil d_i^* \rceil$ banks fail when bank i is hit by a large shock. □

A.3 Proof of Remark 2

See the Supplementary Appendix, Section 4, in which we prove a more general result.

A.4 Proof of Proposition 1

We first show that all networks $\mathbf{A} \in \mathcal{A}^*$ are socially optimal.

For rare enough large shocks it is always optimal for a social planner to perfectly insure against small shocks. An upper bound on the expected number of failures conditional on no failures following a small shock is qn. A lower bound on the number of failures whenever at least one bank fails following some small shock is $1/n$, since any bank that fails after a small shock to some asset will also fail after a large shock to the same asset. Thus if $qn < 1/n$, or equivalently $q < \frac{1}{n^2}$, a network that permits a bank to fail following some small shock can never be optimal.

Lemma 1 identifies a lower bound on the number of failures when a large shock hits an organization i. This lower bound, denoted d^*, is given by the positive root of $f(d_i)$ as defined in equation (16) and assumed the planner knew the shock would hit i. As it is suboptimal for the social planner to allow failures following a small shock, a network that simultaneously achieved the lower bound of d^* failures following a large shock to i for all $i \in N$, and no failures otherwise, would be socially optimal. We show now that all networks $\mathbf{A} \in \mathcal{A}^*$ achieve this bound.

The class of networks $\mathcal{A}^*$ is defined by partitioning the banks into groups such that:

(i) $A_{ji} = \frac{R-\underline{v}}{\varepsilon_S}$ if $G_i = G_j$.

(ii) $A_{ji} = \frac{R-\underline{v}}{\varepsilon_L+\beta d^*}$ if $G_i \neq G_j$.

(iii) $|G_i| = d^*$ for all $i \in N$.

If a small shock hits a bank i then j's equity value will be $R - \underline{v} - A_{ji}\varepsilon_S$. Substituting in the possible values of A_{ji} above, if $j \in G_i$ then this equity value is weakly positive and j does not fail while if $j \notin G_i$ then this equity value is strictly positive and j does not fail. Thus there are no failures following a small shock. If a large shock hits any bank i then by construction j fails if $j \in G_i$, but if $j \notin G_i$ then j's equity value is $R - \underline{v} - A_{ji}\varepsilon_L - A_{ji}d^*\beta$, which is weakly positive and so j does not fail. Thus, in expectation there are qd^* failures and all networks $\mathbf{A} \in \mathcal{A}^*$ are socially optimal.

We now show that only networks $\mathbf{A} \in \mathcal{A}^*$ are optimal by showing that only these networks can achieve d^* failures following a large shock to any $i \in N$, and no failures otherwise.

As no organizations can fail following a small shock in an optimal network we must have

$$A_{jk} \leq \frac{R-\underline{v}}{\varepsilon_S} \quad \text{for all } j,k \in D_i \text{ and for all } i \in N. \tag{17}$$

Inequalities (13), (14) and (15) are all used to construct $f(d_i)$, and all three must bind for the lower bound d^* to be achieved. Thus by equation (15)

$$\sum_{j\in D_i}\sum_{k\in D_i} A_{jk} = \frac{d_i^2(R-\underline{v})}{\varepsilon_S}. \tag{18}$$

The only way for equations (17) and (18) to hold is for

$$A_{jk} = \frac{R-\underline{v}}{\varepsilon_S} \quad \text{for all } j,k \in D_i \text{ and for all } i \in N. \tag{19}$$

Thus, for any $j, k \in D_i$, we have $A_{jk} = \frac{R-\underline{v}}{\varepsilon_S}$ and so $j \in D_k$. Hence $D_k \supseteq D_i$. As $i \in D_k$, an equivalent analysis for shocks hitting k leads to the conclusion that $D_i \subseteq D_k$. Combining set inclusions we conclude that $D_i = D_k$ for all $k \in D_i$. Hence to achieve the lower bound for all $i \in N$, the set of banks N must be partitioned into disjoint subsets such that when a large shock hits the asset of any bank in the set, all banks in the set default.

The upper bound on losses absorbed by all banks $j \notin D_i$ after a large shock to asset i must also bind—and again, this must hold for all $i \in N$. Therefore by equation (12) we have;

$$A_{jh} = \frac{R - \sum_{k \in D_i} A_{jk}\beta - \underline{v}}{\varepsilon_L} \text{ for all } j \notin D_i \text{ and for all } h \in D_i.$$

Thus $A_{jk} = A_{jh}$ for all $j \notin D_i$ and for all $k, h \in D_i$, and so

$$A_{jh} = \frac{R - \underline{v}}{\varepsilon_L + \beta d^*} \text{ for all } j \notin D_i, \text{ for all } h \in D_i. \tag{20}$$

As there are by assumption no integer problems, d^* is an integer and n/d^* is an integer. Combining the conditions we have found, any network $\mathbf{A}$ that achieves the lower bound must satisfy:

(i) $A_{ij} = \frac{R-\underline{v}}{\varepsilon_S}$ if $D_i = D_j$.

(ii) $A_{ij} = \frac{R-\underline{v}}{\varepsilon_L+\beta d^*}$ if $D_i \neq D_j$.

(iii) $|D_i| = d^*$ for all $i \in N$.

As we have found that the agents must partitioned into groups such that $i \in D_j$ if and only if $j \in D_i$, the above conditions define the set $\mathcal{A}^*$ with $D_i = G_i$ for all i. Thus all socially optimal networks are members of $\mathcal{A}^*$. □

A.5 Proof of Remark 3

For this proof we relax our integer assumption and let $d^* \in \mathbb{R}$. The assumption that d^* is an integer constrains the values that parameters can take, but the claims continue to hold.

(i) Recall that the socially efficient cluster size d^* is given by the positive root of:

$$d^2(R-\underline{v})\beta + d[(R-\underline{v})(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta] + n\varepsilon_S(R-\underline{v}) - \varepsilon_S\varepsilon_L = 0 \tag{21}$$

Implicitly differentiating equation (21) with respect to $\beta, \varepsilon_L, \varepsilon_S$ and $R-\underline{v}$, respectively, yields:

$$\frac{\partial d^*}{\partial \beta} = \frac{d^*(\varepsilon_S - d^*(R-\underline{v}))}{2d^*\beta(R-\underline{v}) + (R-\underline{v})(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta}, \tag{22}$$

$$\frac{\partial d^*}{\partial \varepsilon_L} = \frac{\varepsilon_S - d^*(R-\underline{v})}{2d^*\beta(R-\underline{v}) + (R-\underline{v})(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta}, \tag{23}$$

$$\frac{\partial d^*}{\partial \varepsilon_S} = \frac{d[(R-\underline{v}) + \beta] + \varepsilon_L - n(R-\underline{v})}{2d^*\beta(R-\underline{v}) + (R-\underline{v})(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta}, \tag{24}$$

$$\frac{\partial d^*}{\partial (R-\underline{v})} = \frac{-[d^{*2}\beta + d^*(\varepsilon_L - \varepsilon_S) + \varepsilon_S n]}{2d^*\beta(R-\underline{v}) + (R-\underline{v})(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta}. \tag{25}$$

We will sign these expressions. Rearranging equation (21) yields

$$d^*(R-\underline{v}) = \varepsilon_S - (n-d)\frac{R-\underline{v}}{\varepsilon_L + \beta d^*},$$

and so

$$d^*(R-\underline{v}) < \varepsilon_S. \tag{26}$$

Since d^* is the unique positive root of equation (21), applying the quadratic formula:

$$d^* = \frac{-[(R-\underline{v})(\varepsilon_L-\varepsilon_S)-\varepsilon_S\beta] + \sqrt{[(R-\underline{v})(\varepsilon_L-\varepsilon_S)-\varepsilon_S\beta]^2 - 4(R-\underline{v})\beta\left[n\varepsilon_S(R-\underline{v})-\varepsilon_S\varepsilon_L\right]}}{2(R-\underline{v})\beta}.$$

Rearranging yields

$$\begin{aligned}&2d^*\beta(R-\underline{v}) + (\varepsilon_L-\varepsilon_S)(R-\underline{v}) - \varepsilon_S\beta \\ &\qquad = \sqrt{[(R-\underline{v})(\varepsilon_L-\varepsilon_S)-\varepsilon_S\beta]^2 - 4(R-\underline{v})\beta\left[n\varepsilon_S(R-\underline{v})-\varepsilon_S\varepsilon_L\right]} > 0.\end{aligned} \tag{27}$$

We can now sign equations (22), (23), (24) and (25). In all cases the denominator is positive by equation (27). The numerator is positive in equations (22) and (23) according to equation (26); and positive in equation (24) and negative in equation (25) by the definition of a large shock.

(ii) Substituting $\varepsilon_S = n(R-\underline{v})$ into equation (21) and dividing through by ε_S yields:

$$\frac{d^2\beta}{n} + d[\frac{\varepsilon_L - n(R-\underline{v})}{n} - \beta] + n(R-\underline{v}) - \varepsilon_L = 0$$

which has $d^* = n$ as its (weakly) positive root.

(iii) Let i and j be in different clusters such that $A^*_{ij} = \frac{R-\underline{v}}{\varepsilon_L+\beta d^*}$ is a firebreak link. Using the results from part (i):

$$\begin{aligned}\frac{dA^*_{ij}}{d\beta} &= \frac{\partial A^*_{ij}}{\partial \beta} + \frac{\partial A^*_{ij}}{\partial d^*}\frac{\partial d^*}{\partial \beta} < 0\\ \frac{dA^*_{ij}}{d\varepsilon_L} &= \frac{\partial A^*_{ij}}{\partial \varepsilon_L} + \frac{\partial A^*_{ij}}{\partial d^*}\frac{\partial d^*}{\partial \varepsilon_L} < 0\\ \frac{dA^*_{ij}}{d\varepsilon_S} &= \frac{\partial A^*_{ij}}{\partial \varepsilon_S} + \frac{\partial A^*_{ij}}{\partial d^*}\frac{\partial d^*}{\partial \varepsilon_S} \leq 0\\ \frac{dA^*_{ij}}{d(R-\underline{v})} &= \frac{\partial A^*_{ij}}{\partial (R-\underline{v})} + \frac{\partial A^*_{ij}}{\partial d^*}\frac{\partial d^*}{\partial (R-\underline{v})} > 0.\end{aligned}$$

□

A.6 Proof of Proposition 2

1. For any $\mathbf{A} \in \mathcal{A}'$ there are, by definition, no failures following a small shock. Thus the sum of all banks' equity values is $n(R-\underline{v}) - \varepsilon_S$. After a large shock, by limited liability, the sum of all banks'

equity values is weakly positive. We show now that for all $\mathbf{A}^* \in \mathcal{A}^* \subset \mathcal{A}'$ the sum of all banks' equity values is zero, and hence, that all networks in $\mathcal{A}^*$ minimize the sum of the banks' expected equity values within the class of networks $\mathcal{A}'$.

Consider any network $\mathbf{A}^* \in \mathcal{A}^*$ and suppose a large shock hits a bank i. Then, by construction, all banks in the group G_i fail and have equity values equal to zero. A bank $j \notin G_i$ has an equity value given by $\max\{0, R - \underline{v} - A_{ji}\varepsilon_L - \sum_{k \in G_i} A_{jk}\beta\}$. By the definition of the class of networks $\mathcal{A}^*$ this simplifies to

$$\max\{0, R - \underline{v} - \left(\frac{R - \underline{v}}{\varepsilon_L + d^*\beta}\right)\varepsilon_L - \left(\frac{R - \underline{v}}{\varepsilon_L + d^*\beta}\right) d^*\beta\} = 0.$$

Thus the equity value of all banks is 0 following a large shock to any bank.

2. The social planner maximizes the expected sum of equity and debt holder value: $\mathbb{E}\left[\sum_i (\pi_i + \delta_i)\right]$. Thus all networks $\mathbf{A}^* \in \mathcal{A}^*$ maximize $\mathbb{E}\left[\sum_i (\pi_i + \delta_i)\right]$. As we have already shown that all networks in $\mathcal{A}^*$ minimize $\mathbb{E}\left[\sum_i \pi_i\right]$ within the class of networks $\mathcal{A}'$, thus all networks in $\mathcal{A}^*$ must maximize $\mathbb{E}\left[\sum_i \delta_i\right]$ within the class of networks $\mathcal{A}'$. □

A.7 Proof of Proposition 3

We begin by stating a Lemma. It is proved in the next subsection.

LEMMA **2.**

$$\frac{\partial A_{gh}}{\partial C_{ij}} = \frac{A_{gi}A_{jh}}{\widehat{C}_{jj}} \text{ for all } j, g, h \in N$$

We now turn to proving the proposition. By definition

$$\begin{aligned} v_i(\theta) &= \sum_{l \in N} A_{il}(p(\theta) - \beta I_{v_l(\theta) < \underline{v}}) \\ \pi_i(\theta) &= [v_i(\theta) - \underline{v}] I_{v_i(\theta) < \underline{v}}. \\ \mathbb{E}[\pi_i] &= \sum_{\theta \in \Theta} P(\theta)\pi_i(\theta). \end{aligned}$$

We will first construct a marginal trade that is strictly profitable, at the margin, under the assumption that $\beta = 0$. We then show that this trade can be scaled up to generate a strictly profitable non-marginal trade for $\beta = 0$. Finally, we argue that for sufficiently small $\beta > 0$ this trade remains profitable.

The trade we will consider is between banks i and j in different clusters. We let i exchange a marginal unit of its claims on j, C_{ij}, for a marginal unit of j's claim on i. There are thus four changes of equal magnitude to cross-holdings: C_{ij} decreases, C_{ii} increases, C_{ji} decreases and C_{jj} increases. Applying Lemma 2, the marginal change in an organization g's ultimate claims on h's underlying assets is then:

$$
\begin{aligned}
\frac{\partial A_{gh}}{\partial C_{ii}} - \frac{\partial A_{gh}}{\partial C_{ij}} + \frac{\partial A_{gh}}{\partial C_{jj}} - \frac{\partial A_{gh}}{\partial C_{ji}} &= \frac{A_{gi}A_{ih}}{\widehat{C}_{ii}} - \frac{A_{gi}A_{jh}}{\widehat{C}_{jj}} + \frac{A_{gj}A_{jh}}{\widehat{C}_{jj}} - \frac{A_{gj}A_{ih}}{\widehat{C}_{ii}} \\
&= \left(\frac{A_{ih}}{\widehat{C}_{ii}} - \frac{A_{jh}}{\widehat{C}_{jj}}\right)(A_{gi} - A_{gj}) \qquad (28)
\end{aligned}
$$

Therefore the change in i's claims on h's underlying assets is:

$$
\left(\frac{A_{ih}}{\widehat{C}_{ii}} - \frac{A_{jh}}{\widehat{C}_{jj}}\right)(A_{ii} - A_{ij})
$$

Initializing the network to be socially efficient, since i is in a different cluster from j we have $A_{ii} > A_{ij}$. Moreover, in the socially efficient network $\widehat{C}_{ii} = \widehat{C}_{jj}$ so i's claims on h strictly increase if and only if $A_{ih} > A_{jh}$. Thus the trade increases i's claims on all banks within i's cluster and reduces i's claims on all banks in j's cluster. Moreover, as for all $\mathbf{C} \in \mathcal{C}^*$, $\widehat{C}_{ii} = \widehat{C}_{jj}$ and $A_{ih} = A_{jh}$ for all $h \notin G_i \cup G_j$, by equation 28 the trade with j leaves i's claims on banks from other clusters unchanged.

By assumption $\beta = 0$, and by the symmetry of the trade, A remains row stochastic. Therefore i's equity value in state θ_k^a is:

$$
\pi_i(\theta_k^a) = \max\{(R - \underline{v}) - A_{ik}\varepsilon_a, 0\}.
$$

Consider bank i. Initially we have $\pi(\theta_k^a) = 0$ for all $k \in G_i$ and $a \in \{L, S\}$. As the trade increases A_{ik} for $k \in G_i$, i is exposed to more of a shock to a bank in its cluster, and so $v_i(\theta_k^a)$ falls for all $k \in G_i$ and $a \in \{L, S\}$. Thus i fails in all states θ_k^a for $k \in G_i$ after the trade. However, in these states, we still have $\pi(\theta_k^a) = 0$ by limited liability. In states θ_k^a for $k \notin G_i$, i does not fail prior to the trade. After the trade, for $k \notin G_i, G_j$, A_{ik} remains constant; and for $k \in G_j$, A_{ik} strictly decreases. Thus in states θ_k^a such that $k \notin G_i$, i is exposed to weakly less of the shock and $v_i(\theta_k^a)$ weakly increases. Hence $\pi(\theta_k^a)$ weakly increases for $k \notin G_i, G_j$ and strictly increases for $k \in G_j$. As i's equity value weakly increases at the margin in all states of the world and strictly increases in some states, the trade strictly increases i's expected equity value at the margin. By symmetry the trade also strictly increases j's expected equity value at the margin.

Suppose that i exchanges $\eta > 0$ units of C_{ij} with j for η units of C_{ji}. This scales up the marginal trade already considered. Let $\mathbb{E}[\pi_i(\eta)]$ be the expected shareholder value of i after a trade of size η. We can then consider the impact of a change in η on $\mathbb{E}[\pi_i(\eta)]$. First note that $\mathbb{E}[\pi_i(\eta)]$ is continuously differentiable in $v_i(\theta)$ for all θ at all $v_i(\theta) < \underline{v}$ and at all $v_i(\theta) > \underline{v}$. After the marginal trade $v_i(\theta_k^a) < \underline{v}$ for all $k \in G_i$, $v_i(\theta_k^a) > \underline{v}$ for all $k \in G_j$ and, as A_{ik} remains fixed for all $k \notin G_i \cup G_j$, $v_i(\theta_k^a)$ remains fixed for all $k \notin G_i \cup G_j$. Note too that $v_i(\theta)$ is continuously differentiable in A_{il} for all l and A_{il} is a continuously differentiable in C_{ij}, C_{ji}, C_{ii} and C_{jj}. Thus the value of the marginal change in $\mathbb{E}[\pi_i(\eta)]$ changes continuously as η is increased from 0. Therefore since the constructed trade is strictly profitable at the margin, there exists an $\eta > 0$ such that the proposed trade is strictly profitable for the equity holders of i. By symmetry such a trade is also strictly profitable for the equity holders of j.

Finally, observe that $v_i(\theta)$ is continuous in β and so $\mathbb{E}[\pi_i(\eta)]$, is continuous in β. Thus there exists a

$\overline{\beta} > 0$ such that the η trade constructed above is strictly profitable for all $\beta < \overline{\beta}$. □

A.8 Proof of Lemma 2

We will use the following result (Petersen et al., 2008; Section 2). For a matrix M:

$$\frac{\partial(\mathbf{M}^{-1})_{kl}}{\partial M_{ij}} = -(\mathbf{M}^{-1})_{ki}(\mathbf{M}^{-1})_{jl}.$$

Thus, for all $j, g, h \in \{1, \ldots, n\}$

$$\begin{aligned}
\frac{\partial A_{gh}}{\partial C_{ij}} &= \frac{\partial(\widehat{\mathbf{C}}(\mathbf{I}-\mathbf{C})^{-\mathbf{1}})_{gh}}{\partial C_{ij}} \\
&= \widehat{C}_{gg}\frac{\partial((\mathbf{I}-\mathbf{C})^{-\mathbf{1}})_{gh}}{\partial-(\mathbf{I}-\mathbf{C})_{ij}} \\
&= \widehat{C}_{gg}((\mathbf{I}-\mathbf{C})^{-\mathbf{1}})_{gi}((\mathbf{I}-\mathbf{C})^{-\mathbf{1}})_{jh} \\
&= \frac{A_{gi}A_{jh}}{\widehat{C}_{jj}}.
\end{aligned}$$

□

A.9 Proof of Proposition 4

The proof is by contradiction. Consider a stable network with $C_{il} > 0$ and $C_{jk} > 0$ and consider a marginal trade between i and j in which i exchanges a marginal unit of C_{il} for λ units of C_{jk}. As a result of such a trade four entries in C change: C_{ik}, C_{il}, C_{jk} and C_{jl}. As the changes are marginal and all other entries in C remained fixed, the effect on A can be decomposed into the following effects, where the equalities follow from Lemma 2:

$$\frac{\partial A_{ig}}{\partial C_{ik}} - \frac{\partial A_{ig}}{\partial C_{jk}} = \frac{A_{ii}A_{kg}}{\widehat{C}_{kk}} - \frac{A_{ij}A_{kg}}{\widehat{C}_{kk}} = (A_{ii}-A_{ij})\frac{A_{kg}}{\widehat{C}_{kk}} \tag{29}$$

$$\frac{\partial A_{ig}}{\partial C_{jl}} - \frac{\partial A_{ig}}{\partial C_{il}} = \frac{A_{ij}A_{lg}}{\widehat{C}_{ll}} - \frac{A_{ii}A_{lg}}{\widehat{C}_{ll}} = -(A_{ii}-A_{ij})\frac{A_{lg}}{\widehat{C}_{ll}}, \tag{30}$$

as by assumption no banks end up exactly on their failure threshold for any possible shock absent trade, and as we are constructing marginal trades the failure set does not change following any shock. Recall that:

$$\mathbb{E}[\pi_i] = \sum_{\theta\in\Theta} P(\theta)\max\{\sum_{g\in N} A_{ig}(p_g(\theta) - \beta_g I_{v_g(\theta)<\underline{v}}) - \underline{v}, 0\}.$$

By assumption the failure sets are unchanged by a marginal trade. Utilizing equations (29) and (30) the marginal value to i of the trade is then:

$$\sum_{\theta\in\Theta} P(\theta)\left(\sum_{g\in N}\left[(A_{ii}-A_{ij})\left(\lambda\frac{A_{kg}}{\widehat{C}_{kk}}-\frac{A_{lg}}{\widehat{C}_{ll}}\right)\right](p_g(\theta)-\beta_g I_{v_g(\theta)<\underline{v}})\right) I_{v_i(\theta)\geq\underline{v}}. \tag{31}$$

Substituting in the definition of ν_{ig} from equation (8), the trade is profitable to i if:

$$(A_{ii}-A_{ij})\sum_g\left(\lambda\frac{A_{kg}}{\widehat{C}_{kk}}-\frac{A_{lg}}{\widehat{C}_{ll}}\right)\nu_{ig}>0 \qquad\iff\qquad \lambda>\frac{\sum_g \frac{A_{lg}}{\widehat{C}_{ll}}\nu_{ig}}{\sum_g \frac{A_{kg}}{\widehat{C}_{kk}}\nu_{ig}}. \tag{32}$$

By symmetry the condition for the trade to be profitable for j is that:

$$(A_{jj}-A_{ji})\sum_g\left(\frac{A_{lg}}{\widehat{C}_{ll}}-\lambda\frac{A_{kg}}{\widehat{C}_{kk}}\right)\nu_{jg}>0 \qquad\iff\qquad \lambda<\frac{\sum_g \frac{A_{lg}}{\widehat{C}_{ll}}\nu_{jg}}{\sum_g \frac{A_{kg}}{\widehat{C}_{kk}}\nu_{jg}}. \tag{33}$$

Thus there exists a λ that makes the trade described profitable at the margin if and only if

$$\frac{p_{jl}}{p_{jk}}>\frac{p_{il}}{p_{ik}}.$$

□

A.10 Proof of Corollary 1

As the proof closely mirrors that of Proposition 3 we only provide an outline here. First, it easily verified that at for a socially efficient network with at least two clusters and at most $N/2$ clusters:

$$\frac{p_{il}}{p_{ik}}<\frac{p_{jl}}{p_{jk}},$$

when $i\neq k$ are in the same cluster and $j\neq l$ are in a different cluster. By the proof of Proposition 4 there then exists a strictly profitable marginal trade in which i exchanges marginal claims on l for j's marginal claims on k. Following the proof strategy in Proposition 3 this strictly profitable marginal trade can then be scaled up to generate a strictly profitable non-marginal trade. Finally, bankruptcy costs can be increased a little without making this trade unprofitable as equity values are continuous in bankruptcy costs.

A.11 Proof of Proposition 5

Fix any network $\mathbf{A}$ and any $q\in(0,\overline{q}]$. By equation (2), as failures are complementary and as $A_{ij}\geq 0$ for all i,j, each bank's market value is weakly decreasing in the realized value of ε. Thus, the expected number of failures is smallest when $F_L(\underline{\varepsilon}_L)=1$ and $F_S(\underline{\varepsilon}_S)=1$.

The optimal network in this best case scenario is $A^*(\underline{\varepsilon}_S,\underline{\varepsilon}_L)$ by Proposition 1. Recall that for all $q\in(0,\overline{q}]$ the optimal network $\mathbf{A}^*$ does not depend on q. Thus for any shock distribution, a lower bound on the expected number of failures is given by $qd^*(\underline{\varepsilon}_S,\underline{\varepsilon}_L)$ and an upper bound on the expected social surplus is

$$\overline{\mathcal{S}} := nR - (1-q)\int_{\underline{\varepsilon}_S}^{\overline{\varepsilon}_S} \varepsilon_S dF_{\varepsilon_S} - q\left(\int_{\underline{\varepsilon}_L}^{\overline{\varepsilon}_L} \varepsilon_L dF_{\varepsilon_L} + \beta d^*(\underline{\varepsilon}_S, \underline{\varepsilon}_L)\right).$$

Furthermore, as for any network $\mathbf{A}$ and any $q \in (0, \overline{q}]$ all banks market values are decreasing in the realized value of ε, the expected number of failures is maximized when $F_L(\varepsilon_L) = 0$ for all $\varepsilon_L < \overline{\varepsilon}_L$ and $F_S(\varepsilon_S) = 0$ for all $\varepsilon_S < \overline{\varepsilon}_S$. Thus a lower bound on the surplus a social planner can obtain in general, is that achieved by the network $\mathbf{A}^*(\overline{\varepsilon}_S, \overline{\varepsilon}_L)$. For any shock distribution, the expected social surplus from implementing this network is at least

$$\underline{\mathcal{S}} := nR - (1-q)\int_{\underline{\varepsilon}_S}^{\overline{\varepsilon}_S} \varepsilon_S dF_{\varepsilon_S} - q\left(\int_{\underline{\varepsilon}_L}^{\overline{\varepsilon}_L} \varepsilon_L dF_{\varepsilon_L} + \beta d^*(\overline{\varepsilon}_S, \overline{\varepsilon}_L)\right).$$

Thus for any distributions F_S and F_L the percentage of social surplus lost by choosing network $A^*(\overline{\varepsilon}_S, \overline{\varepsilon}_L)$ instead of whatever network is optimal for the exact distributions F_S and F_L is bounded from above by

$$\frac{\overline{\mathcal{S}} - \underline{\mathcal{S}}}{\overline{\mathcal{S}}} = \frac{q\beta\left(d^*(\overline{\varepsilon}_S, \overline{\varepsilon}_L) - d^*(\underline{\varepsilon}_S, \underline{\varepsilon}_L)\right)}{nR - (1-q)\int_{\underline{\varepsilon}_S}^{\overline{\varepsilon}_S} \varepsilon_S dF_{\varepsilon_S} - q\left(\int_{\underline{\varepsilon}_L}^{\overline{\varepsilon}_L} \varepsilon_L dF_{\varepsilon_L} + \beta d^*(\underline{\varepsilon}_S, \underline{\varepsilon}_L)\right)}.$$

Finally, note that

$$\frac{q\beta\left(d^*(\overline{\varepsilon}_S, \overline{\varepsilon}_L) - d^*(\underline{\varepsilon}_S, \underline{\varepsilon}_L)\right)}{nR - (1-q)\int_{\underline{\varepsilon}_S}^{\overline{\varepsilon}_S} \varepsilon_S dF_{\varepsilon_S} - q\left(\int_{\underline{\varepsilon}_L}^{\overline{\varepsilon}_L} \varepsilon_L dF_{\varepsilon_L} + \beta d^*(\underline{\varepsilon}_S, \underline{\varepsilon}_L)\right)} \leq \frac{\overline{q}\beta\left(d^*(\overline{\varepsilon}_S, \overline{\varepsilon}_L) - d^*(\underline{\varepsilon}_S, \underline{\varepsilon}_L)\right)}{nR - (1-\overline{q})\underline{\varepsilon}_S - \overline{q}\left(\underline{\varepsilon}_L + \beta d^*(\underline{\varepsilon}_S, \underline{\varepsilon}_L)\right)},$$

which completes the proof. □

A.12 Proof of Corollary 2

As shown in the proof of Proposition 5, fixing for any network $\mathbf{A}$, and fixing any $q \in (0, \overline{q}]$, for any F_S and any F_L, the social surplus is minimized by a shock distribution in which a shock $\overline{\varepsilon}_S$ occurs with probability $(1-q)$ and a shock $\overline{\varepsilon}_L$ occurs with probability q. Thus applying Proposition 1 the optimal network for the social planner is $A^*(\overline{\varepsilon}_S, \overline{\varepsilon}_L, q) = A^*(\overline{\varepsilon}_S, \overline{\varepsilon}_L, \overline{q})$. □

A.13 Proof of Remark 4

Denote the size of firm i as σ_i, and consider the set of firms N after any mergers. The social planner maximizes:

$$\mathbb{E}\left[\sum_{i\in N} \pi_i + \delta_i\right] \text{ subject to } \sum_{i\in N} A_{ij} = 1 \text{ for all } j \in N.$$

We have:

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i\in N}\pi_i+\delta_i\right] &= \mathbb{E}\left[\sum_{i\in N}\max\{v_i-\underline{v},0\}+\min[\underline{v},v_i]\right] \\
&= \mathbb{E}\left[\sum_{i\in N}v_i\right] \\
&= \mathbb{E}\left[\sum_{i\in N}\sum_{j\in N}A_{ij}(p_j-\beta\sigma_j I_{v_j<\underline{v}})\right] \\
&= \sum_{i\in N}\sum_{j\in N}A_{ij}\mathbb{E}[p_j]-\beta\sum_{i\in N}\sum_{j\in N}A_{ij}\sigma_j\mathbb{E}[I_{v_j<\underline{v}}] \\
&= \sum_{j\in N}\mathbb{E}[p_j]-\beta\sum_{j\in N}\sigma_j\mathbb{E}[I_{v_j<\underline{v}}].
\end{aligned}
$$

Where the last line uses the fact that $\mathbf{A}$ is column stochastic. Note that after mergers, $\mathbb{E}[p_j]=\sigma_j R-\left(\frac{\sigma_j}{\sum_{i\in N}\sigma_i}\right)(q\varepsilon_L+(1-q)\varepsilon_S)$ so that:

$$
\mathbb{E}\left[\sum_{i\in N}\pi_i+\delta_i\right]=\sum_{j\in N}\sigma_j R-(q\varepsilon_L+(1-q)\varepsilon_S)-\beta\sum_{j\in N}\sigma_j\mathbb{E}[I_{v_j<\underline{v}}].
$$

Since all but the last term is exogenously given, the social planner minimizes the expected size of default costs. $\square$

A.14 Proof of Proposition 6

Starting from an initial pre-merger network among homogeneously sized banks, $\mathbf{A}$, consider a (M,i)-merger and denote the resulting network $\mathbf{A}''$. Let $\mathcal{A}$ be the set of feasible initial networks and let $\mathcal{A}'\subset\mathcal{A}$ be the subset of these feasible networks, still among n banks, such that $A'_{kj}=A'_{lj}$ for all $k,l\in M$.

First we will show that any post-merger network $\mathbf{A}''$ among $n-m+1$ banks can be mapped into a network $\mathbf{A}'\in\mathcal{A}'$ among n banks, such that the expected default costs obtained in the post-merger network equal the expected default costs in the network $\mathbf{A}'$. For any network $\mathbf{A}$ and a (M,i)-merger, define $\mathbf{A}'$ by setting $A'_{kj}=\frac{\sum_{k\in M}A_{kj}}{|M|}$ for all $k\in M$ and $A'_{kj}=A_{kj}$ for all other entries. The resulting network $\mathbf{A}'$ is feasible as it can be generated by a feasible cross-holding matrix (see the Supplementary Appendix, Section 4.2). Observe that:

(i) $v'_j=v'_k$ for all $j,k\in M$ in all states of the world

(ii) After the merger, $v''_i=\sum_{j\in N}A''_{ij}(p_j-\beta''_j I_{v_j<\bar{v}})$

(iii) By the definition of a merger, $A''_{ij}=\sum_{k\in M}A_{kj}$

Thus, in the network $\mathbf{A}'$, banks $k\in M$ fail, such that $v'_k<\underline{v}$, if and only if i fails in $\mathbf{A}''$ such that $v''_i<m\underline{v}$. Hence the same failure costs are incurred in every state of the world in the pre-merger network $\mathbf{A}'$ as in the post merger network $\mathbf{A}''$.

As every post merger network can be mapped to some pre-merger network in $\mathcal{A}'$, a lower bound on the expected default costs a social planner can achieve post merger, can be found by minimizing expected default costs over networks in the set $\mathcal{A}'$. Moreover, as $\mathcal{A}' \subseteq \mathcal{A}$ the social planner must be able to do weakly better when choosing from $\mathcal{A}$. Socially optimal post-merger expected default costs must be weakly higher than socially optimal pre-merger expected default costs.

We have shown that the expected default costs obtained pre-merger are a lower bound on the expected default costs a social planner can achieve post-merger. We now show that this lower bound can be achieved by the networks that result from taking a socially optimal network $\mathbf{A}^*$ and merging banks within cluster. Hence we conclude that these networks are socially optimal for the post-merger size distribution.

By Proposition 1, the socially optimal networks chosen from $\mathcal{A}$ are such that $A^*_{kl} = A^*_{jl}$ for all k, j in the same cluster. Take a socially optimal network $\mathbf{A}^*$ and merge banks within a cluster to generate network $\mathbf{A}''$. This network generates equivalent expected default costs to $\mathbf{A}'$, where $A'_{kj} = \frac{\sum_{k \in M} A^*_{kj}}{m}$ for all $k \in M$ and $A'_{kj} = A^*_{kj}$ for all other entries. However, as all banks M are within cluster and the initial network is socially optimal, $A^*_{kj} = A^*_{lj}$ for all $k, l \in M$ and all j. Thus $A'_{kj} = A^*_{kj}$ for all $k \in M$ and $\mathbf{A}' = \mathbf{A}^*$, so the lower bound on expected default costs is achieved. □

A.15 Proof of Proposition 7

In any solution to the relaxed social planner's problem, the constraints on the solution must bind. These constraints are

$$A_{ji}\varepsilon_{w-1} + \sum_{k \in D_i^{w-1}} A_{jk}\beta \leq R - \underline{v}. \quad \text{for all } j \in D_i^w \text{ and all } w = 1, \ldots, \bar{w}.$$

Suppose, towards a contradiction, that we are at an optimal solution for the planner and the inequality corresponding to bank j's holdings did not bind. Then bank j has strictly positive equity value in all states of the world in which it does not fail. It is thus possible to increase j's dependency on i and reduce the dependency of other banks on i without violating column stochasticity and with bank j failing in the same states. Thus, as we are ignoring integer problems under the generalized integer assumption, dependencies can be adjusted so that weakly few banks fail after any shock to i, and strictly fewer fail after some shock to i.

Take any vector $\mathbf{d}'_i(\mathbf{A})$ which solves the relaxed social planner's problem. As the inequality constraints bind,

$$A_{ji}\varepsilon_{w-1} + \sum_{k \in D_i^{w-1}} A_{jk}\beta = R - \underline{v}. \quad \text{for all } j \in D_i^w \text{ and all } w = 1, \ldots, \bar{w}.$$

We will show that we can construct a $\bar{w}$-layered-clustered network satisfying these conditions and with the same number of failures following each shock. Thus we show now that it is feasible to create a $\bar{w}$-layered-clustered network such that $|G_i^w| = d_i^{w\prime}(\mathbf{A})$ for all w. Consider a $\bar{w}$-layered-clustered network such that

$$A_{ji} = \frac{R - \underline{v}}{\varepsilon_{w-1} + \beta d_i^{w-1'}} \quad \text{for all } j \in G_i^w \backslash G_i^{w-1}\text{, for all } i \in N\text{, and for all } w.$$

This pins down A_{ji} for all i, j, and so completely defines the network. It is easily verified that this network satisfies the (binding) inequality constraints on the relaxed social planner's problem and, by construction, achieves the same number of failures following each possible shock to i. It is also both row stochastic and column stochastic. It thus satisfies all the constraints in the relaxed social planner's problem.

We now show that this network solves the initial (unrelaxed) social planner's problem. Suppose that instead of conditioning on a shocks hitting bank i we conditioned on it hitting bank $j \neq i$. As banks are ex-ante identical, there would exist a solution to the relaxed social planner's problem in which the same distribution of banks failures occurred following shocks of different sizes as found before. Thus, an upper bound on the unrelaxed social planner's solution is one that achieves this distribution of failures conditional on a shock hitting *any* bank. By symmetry, the constructed $\bar{w}$-layered-clustered network obtains the (same) optimal distribution of failures conditional on a shock hitting *any* $i \in N$. Thus the constructed $\bar{w}$-layered-clustered network obtains this upper bound. Second, we verify in the Supplementary Appendix, Section 4, that any $\bar{w}$-layered-clustered network $\mathbf{A}$ can generated by some feasible cross-holdings $\mathbf{C}$. $\square$