Firebreaks and Risk-Shifting in Financial Networks

Matthew Elliott and Jonathon Hazell*

May 2017, first draft September 2015

Abstract

Financial connections hedge risks, but also propagate large shocks. We characterise socially optimal financial networks in the presence of this tradeoff. These networks divide banks into groups, with strong connections within groups but weak connections between groups. The weak connections create firebreaks, which limit the spread of contagion. When financial distress costs are below a certain threshold, socially efficient networks cannot form in equilibrium. Shareholders engage in risk-shifting by altering financial connections—thereby increasing equity value, but also raising the likelihood of contagion. Equally, socially efficient networks are an equilibrium when financial distress costs are above this threshold. Contagion becomes so costly that banks are not willing to take the other side of a risk-shifting trade.

1 Introduction

A central concern in financial regulation is the systemic risk generated by counterparty relations. Can the failure of some financial institution cause its counterparties to default, and precipitate a cascade of failures? We need to understand what regulation might mitigate this possibility, and what features make the financial system resilient against contagion. However, designing regulation raises questions about institutions incentives. Without regulation, will the financial system arrange itself in a resilient pattern? If not, can regulation encourage these features? Could misguided regulation inadvertently undermine these goals? In the context of a parsimonious model, we aim to clarify some of these issues.

We identify a key feature that makes financial systems resilient, which we term firebreaks. Financial systems face a tradeoff, between preventing the spread of contagion when large shocks hit the system, and preventing failures during more tranquil times. The system that optimally balances this tradeoff partitions institutions into distinct groups, with strong links within groups and weak links between groups. The weak links between groups help to insure against failures during relatively normal times. However, they prevent the spread of contagion—appealing to metaphor, we dub these weak links firebreaks.

*Elliott: Cambridge and California Institute of Technology; Hazell: MIT, email: jhazell@mit.edu. This paper was previously circulated under the title “Endogenous Financial Networks: Efficient Modularity and Why Shareholders Prevent It.” We thank Daron Acemoglu, Agostino Capponi, Ben Golub, Sanjeev Goyal, Matthew Jackson, and Alp Simsek, as well as numerous seminar participants, for helpful comments and discussion. Elliott acknowledges funding from the Cambridge Endowment for Research in Finance. Hazell acknowledges funding from the Caltech Summer Research Fellowship programme.
Does the financial system endogenously generate firebreaks and become resilient to contagion? We find that two opposing forces govern whether these structures emerge in equilibrium. First, shareholders have incentives to transfer value away from debt holders and to themselves by engaging in risk-shifting (Jensen and Meckling, 1976). The opportunity for risk-shifting comes through choosing which other banks to hedge risks with. We find that banks like to positively correlate their failures with their largest counterparties, and so risk-shifting increases the likelihood of contagion. However, there is a countervailing force. While shareholders aim to engage in risk-shifting, they also try to avoid contagion generated by the failures of other banks. Shareholders may prevent their counterparties from engaging in risk-shifting by refusing to enter into otherwise profitable trades with them.

Costs of financial distress modulate the strength of these two forces. When these costs are large, and contagion becomes sufficiently harmful, firms prevent their counterparties from risk-shifting. Therefore financial distress costs determine whether resilient financial networks form in equilibrium—which has implications for policy. Regulation can introduce a novel moral hazard problem in a network context, separate from the traditional distortions. In particular, intervention may undermine banks’ incentives to prevent other banks from risk-shifting.

We model a finite set of financial institutions, also termed banks, who each own proprietary investments that generate a random cashflow. Banks can diversify their exposures to their proprietary investments by trading with each other. However, these trades create dependencies between banks. Overall market values are determined by the value of investment cashflows flowing to each bank. Banks have liabilities to external debt holders, and the residual market value after these liabilities are settled becomes equity value. In states of the world where market values are less than the value of external debt, banks fail. These failures trigger discontinuous falls in the value of inter-bank contracts and external debt, due to costs of financial distress. Thus financial distress costs amplify shocks to real investments. We make two crucial further assumptions. Shareholders are protected by limited liability, and external debt holders have no control over cashflows.

Our first set of results concerns socially efficient networks. The social planner seeks to maximize the combined value of all equity and debt in the presence of shocks to real investments. Strong interconnections across many counterparties can prevent bankruptcies after relatively small shocks, avoiding additional losses from the costs of financial distress. However, interconnections also render the system fragile to large shocks, by transmitting the shock across multiple counterparties, and potentially leading to financial contagion and many failures. 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). Facing the possibility of small and large shocks to real investments, our first key finding is that the social planner’s solution partitions 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 a single cluster of banks to fail. The weak links between groups are firebreaks, which limit the spread of contagion after large shocks, while still helping to prevent failures after small shocks.

We therefore identify firebreaks as an important feature of resilient financial systems. Indeed, our findings match a conjecture by Haldane (2010) about how to regulate financial systems. 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. 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.¹

Firebreaks make financial networks resilient. However, to inform regulations, we need to analyse whether firebreaks can emerge in a decentralised equilibrium—or whether banks have incentives to avoid such structures. To this end, we endogenize banks financial positions with each other. 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 investments, thereby preventing failures in some states of the world. Two competing forces govern whether firebreaks emerge in equilibrium.

First, we take a classic mechanism, risk-shifting. Risk-shifting is caused by limited liability, along with the absence of debtor control over cashflows. In certain states, banks can raise both their probability of failure and their equity value. Since shareholders are protected by limited liability, external debt holders suffer any costs of failure. It is well understood that risk-shifting creates excess risk at the individual firm level. We show that risk-shifting generates incentives to create excess systemic risk at the level of the entire network. At the system level, risk-shifting induces banks to correlate their failures with their largest counterparties, which increases the risk of contagion. Therefore all else equal, risk-shifting undermines banks’ incentives to generate firebreaks.

However, there is a countervailing force. Shareholders want their own bank to engage in risk-shifting, but they prefer other banks to minimise risk. Consider two banks in a financial network. Suppose that the second engages in risk-shifting and fails more frequently. If the second bank is a counterparty of the first, these failures directly reduce the expected value of contracts between the two banks. Moreover, the failure of the second bank might trigger a subsequent cascade of failures in some states. Then the value of the first bank’s contracts with many counterparties might fall. Thus the first bank may prefer to avoid trades that allow the second bank to take on extra risk.

In equilibrium, the relative strength of these two forces depends on the cost of counterparty financial distress. Trades away from socially efficient networks increase the chance of a cascade of failures. When the costs of counterparty financial distress are above a certain threshold, the shareholders of a bank will not enter into trades that permit their counterparties to engage in risk-shifting. The socially efficient networks are then stable. Conversely, for lower costs of counterparty financial distress, there are always profitable trades away from socially efficient networks.

Overall, regulation has subtle effects. For a given financial network, policies that reduce the costs of financial distress can limit the amplification of shocks to real investments, and prevent a cascade of failures. However, in response to a reduction in financial distress costs, banks may cease to prevent each other from engaging in risk-shifting. As a result, systemic risk may actually increase.

There are many reasons why financial connections between banks arise, such as intermediation (Acemoglu et al., 2015b; Farboodi, 2014). We go on to discuss ways in which our model potentially

¹An 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)).
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 (Cifuentes, Ferrucci, and Shin, 2005). Allen and Babus (2009) and Capponi (2016) survey this literature. 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 inter-bank networks and systemic risk—see, for example, Denbee, 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 papers studying socially efficient network structures, are Cabrales, Gottardi, and Vega-Redondo (2017) 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. 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 an amplification and propagation mechanism via costs of financial distress.

Cabrales et al. (2017) study the same key tradeoff governing social efficiency as in our paper—how to limit contagion, while also allowing risk-sharing. However, these authors study continuous shock distributions. Their results complement our analysis. When shocks are sufficiently fat tailed, the tradeoff is resolved by minimizing the chance of contagion and forming a maximally segmented network. When shocks are sufficiently thin tailed, the tradeoff is resolved by maximizing risk-sharing in a fully integrated network structure. For intermediate shock distributions, they find conditions under which the strength of connections in the network takes at most two values. As they do not model any costs of financial distress, only the number of links of each type held by each bank matters, and not the pattern in which these links are organized. We find that when financial distress costs are included, the pattern in which links are organized is crucial. We characterize the socially optimal patterns, which exhibit firebreaks. Links between different groups of highly connected banks must be weak, to contain financial distress costs after one group fails.

Farboodi (2014) and Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015b) consider endogenous net-
work formation. Each 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. Impressively, Farboodi (2014) succeeds in finding equilibrium networks, and the equilibrium structure she identifies matches empirically observed financial networks—there is a core of investment banks connected to each other, 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 focuses 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 risk-shifting. The frictions in our model generate incentives to generate excess systemic risk, that depend on the size of financial distress costs. By contrast, Cochrane (2014) argues that banks have strong endogenous incentives to prevent cascading defaults from occurring in equilibrium. We clarify his logic—banks are unwilling to enter trades that might cause cascading failures of their counterparties, when financial distress costs are high. Conversely, when these costs are small, banks’ incentives to engage in risk-shifting dominate.

Outside the networks literature, the role of bailouts in generating systemic risk has been carefully explored (see e.g. Farhi and Tirole (2012)). In a network context, Erol (2015) also considers the effect of bailouts on equilibrium networks. Erol shows that the possibility of bailouts will increase connectivity in equilibrium, and lead to excess systemic risk. Core-periphery networks can also emerge in equilibrium due to bailouts. By contrast, our model highlights the role of costs of financial distress, which mean an intermediate or even negative bailout can be optimal.

The role of risk-shifting in generating financial instability has been widely discussed in the context of a single-firm framework (e.g. Jensen and Meckling, 1976). Relatively little work considers risk-shifting in a systemic setting. A notable exception is Acharya (2009).

We structure the paper as follows. Section 2 presents our model. Section 3 examines socially efficient networks. Section 4 discusses equilibrium networks. Section 5 discusses some extensions. Section 6 concludes. Proofs are relegated to Appendix A.

## 2 The Model

### 2.1 Banks, Real Investments and Contracts

There is a set $N = \{1, \ldots, n\}$ of financial institutions, which we refer to as banks. Each bank has a proprietary real investment that yields a stochastic cashflow. We let $p_i$ denote the random cashflow generated by bank $i$’s real investment, which we interpret broadly to encompass bank $i$’s idiosyncratic financial exposures.

---

2 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); Galeotti, Ghiglio, and Goyal (2015), Cabrales et al. (2017) and Erol (2015).
Banks can hedge their idiosyncratic exposures by swapping claims on each others’ investments. We denote these claims by a matrix $A$ where $A_{ij} \geq 0$ is the inter-bank claim of bank $i$ on bank $j$’s investment. We can represent $A$ as a weighted, directed graph, where the banks are the nodes and the links are claims between counterparty banks. We refer to $A$ as a financial network.

Banks are funded by a combination of debt and equity. In period 0, each bank issues debt with face value $v$, to entities external to the financial network. In period 1 banks form the financial network. In period 2 cashflows from the real investments realize. The inter-bank contracts $A$ and external debt $v$ determine the distribution of cashflows. Thus financial networks are formed before cashflows realise, and are fixed when banks receive these cashflows.

We do not model banks’ choice of capital structure in period 0. The face value of external debt is fixed, and not contingent on the bank’s choice of claims in the network. There are no mechanisms, in our model, such as debt covenants, that condition the face value of external debt on inter-bank claims. One could imagine, equivalently, that external creditors cannot observe inter-bank claims when they write contracts. In equilibrium, debt holders should correctly anticipate the structure of inter-bank claims. They will demand returns that compensate them for any risk associated with inter-bank claims, and $v$ will adjust accordingly. We do not model this step explicitly.

The market value of bank $i$ is denoted by $v_i$. Market value depends on the claims bank $i$ has on different investments, so that

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

In matrix notation we have:

$$v = Ap$$

Since the total percentage of claims on a banks $j$’s investment must sum to 1, we have

$$\sum_{i \in N} A_{ij} = 1$$

for all $j \in N$.

### 2.2 Costs of Financial Distress

We now introduce costs of financial distress. When $v_i < v$, bank $i$ cannot repay its debt holders in full, which causes $i$ to fail. The associated costs of financial distress lower the cashflows generated by $i$’s investment. We let $\beta$ denote the size of these costs and define

$$b_i(v) = \begin{cases} 
\beta & \text{if } v_i < v, \\
0 & \text{otherwise}.
\end{cases}$$

---

3. Caballero and Simsek (2013) point out that inter-bank financial networks are highly opaque, especially to outsiders.
4. Elliott, Golub, and Jackson (2014) obtain a similar formulation by letting banks have claims on each others’ balance sheets rather than directly on projects. Beyond differences of interpretation, this alternative formulation imposes restrictions on the structure of feasible networks. See also Cabrales, Gale, and Gottardi (2015).
Letting $\pi_i$ be the equity value of $i$ and $\delta_i$ be the external debt value, we have

$$v_i = \sum_{j \in N} A_{ij} (p_j - b_j(v_j))$$  \hspace{1cm} (2)$$

$$\pi_i = \max \{ v_i - \underline{v}, 0 \}$$  \hspace{1cm} (3)$$

$$\delta_i = \min \{ \underline{v}, v \}.$$  \hspace{1cm} (4)$$

A simple thought experiment illustrates the mechanisms at work. Suppose that the cashflow of $i$’s investment falls by some shock of size $\varepsilon$, while all other investment cashflows remain constant. Imagine that after this shock no banks fail. The market value of bank $j$, $v_j$, falls as $j$’s claims on $i$ become worth less—specifically, $j$’s claims on $i$’s project fall in value by $A_{ji}\varepsilon$. The equity value of bank $i$, $\pi_i$, also falls by $A_{ii}\varepsilon$. Since, by assumption $i$ does not fail, the value of bank $i$’s external liabilities remain constant.

Now suppose that $\varepsilon$ is sufficiently large, so that after the shock $v_i < \underline{v}$ and $i$ fails, but small enough that no other bank fails. As bank $i$ fails it incurs financial distress costs that further decrease the cashflows of its investment. $i$’s equity holders are residual claimants, and so receive nothing after $i$ fails, since the external debt cannot be fully settled. Bank $j$’s claim on $i$’s project now falls by $A_{ji}(\varepsilon + \beta)$, due to the added impact of counterparty financial distress costs. Let $v'_i$ be the value of $i$ prior to the shock. The value of external debt in $i$ after the shock is $v_i = v'_i - A_{ii}(\varepsilon + \beta) < \underline{v}$. The debt value falls by $A_{ii}(\varepsilon + \beta)$, less $i$’s equity value prior to the shock.

Finally, suppose instead that after the shock both $i$ and $j$ both fail, but no other banks fail. In this case, $j$’s claim on $i$ continues to fall by $A_{ji}(\varepsilon + \beta)$. However, $i$’s debt holders are now exposed to $j$’s financial distress costs. The debt value of $i$ now falls by $A_{ii}(\varepsilon + \beta) + A_{ij}\beta$, less $i$’s equity value prior to the shock. The key features of these contracts are summarized in figure 5. We discuss these contracts further in Section 2.4.

The shareholder claims $\pi$ reflect limited liability. Our model therefore captures the important difference between equity claims, and other claims on bank $i$. When $i$ defaults its shareholders receive nothing and its external creditors effectively become the new shareholders—the standard interpretation of bankruptcy under limited liability.\(^5\)

In vector notation banks’ market values $\mathbf{v}$, equity values $\mathbf{\pi}$ and debt values $\mathbf{\delta}$ after cashflows realise are

$$\mathbf{v} = A(p - b(\mathbf{v}))$$  \hspace{1cm} (5)$$

$$\mathbf{\pi} = \max \{ \mathbf{v} - \mathbf{\underline{v}}, 0 \}$$  \hspace{1cm} (6)$$

$$\mathbf{\delta} = \min \{ \mathbf{\underline{v}}, \mathbf{v} \}.$$  \hspace{1cm} (7)$$

A vector of market values $\mathbf{v}$ which solves equation (5) is a payment equilibrium. As shown in a variety of related settings (e.g. Eisenberg and Noe, 2001; Acemoglu et al., 2015a) at least one payment equilib-\(^6\)

\(^5\)For simplicity our formulation assumes that external creditors have the same proportional claims on $i$’s project when $i$ fails, as $i$’s shareholders had before $i$ failed. Relaxing this assumption would complicate notation without yielding additional insights.
Figure 1: Let $v'_i$ be the value of $i$'s equity prior to a shock $\varepsilon$ to $i$'s investment. For shocks $\varepsilon < (v'_i - v)/A_{ii}$ bank $i$ does not fail and the losses due to the shock are spread among banks in proportion to their claims on the investment. The shock reduces the equity value of $i$ (Panel (a)) and the value of a bank $j$ with claims $A_{ji} < A_{ii}$ (Panel (c)). For larger shocks $i$ fails, resulting financial distress costs of $\beta$ that further reduce the cashflows of $i$'s investment. These losses reduce both the value of external debt (Panel (b)) and also the value of bank $j$'s claim on $i$'s project (Panel (c)).

2.3 Shocks

We consider only substantial shocks to investment cashflows. All shocks are larger than a given bank's equity value, and occur with probability $r$. They are also rare, so that $r \to 0$. In normal times, which occur with probability $1 - r$, each bank's project generates a baseline return of $R$. When a shock arrives it is specific to a given investment. At most one investment is hit by a shock, and each bank's investment is equally at risk. When the shock $\varepsilon$ hits bank $i$, the vector of realized investment cashflows is

$$p = [R \ldots R - \varepsilon \ldots R]^T.$$  

If shocks were uncorrelated across investments and sufficiently rare, the probability of multiple shocks arriving simultaneously would be negligible. This observation motivates our assumption that a shock affects only one investment. Although stylized, this assumption is a useful benchmark—Caballero

---

6Rogers and Veraart (2013) and Elliott, Golub, and Jackson (2014) derive similar results.

7A trivial generalization introduces some correlation between investment cashflows, and changes none of the results. Consider a second cashflow distribution in which whenever a shock arrives, as modelled previously, all other investment cashflows fall by an identical amount $\varepsilon C$. The first and second cashflow distributions are equivalent if cashflows $R'$ under the second distribution satisfy $R' = R - \varepsilon C$.  

---

\[\text{Size of shock (}\varepsilon)\]
\[\text{Size of shock (}\varepsilon)\]
\[\text{Size of shock (}\varepsilon)\]
\[\{v'_i - v\}/A_{ii}\]
\[\{v'_i - v\}/A_{ii}\]
\[\{v'_i - v\}/A_{ii}\]
\(0\)
\(\varepsilon - A_i \beta\)
\(A_j R\)
\(A_j \beta\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
\(\text{Size of shock (}\varepsilon)\)
and Simsek (2013), Cabrales et al. (2017) and Acemoglu et al. (2015a) take a similar approach. We extend our model to consider more general shock distributions in Section 5.1.

We are interested in how private agents balance the benefits of financial connections, coming from insurance against relatively small shocks; and the costs of financial connections, coming from the potential for contagion. We initially adopt the simplest shock distribution capable of capturing such a tradeoff. We assume $\varepsilon$ is random and can be either large or small with some probability:

$$
\varepsilon = \begin{cases} 
\varepsilon_L > n(R - \psi) & \text{with probability } qr \\
\varepsilon_S \in [R - \psi, n(R - \psi)] & \text{with probability } (1 - q)r.
\end{cases}
$$ (8)

Absent exchanging claims, and in a state in which no shock hits, the equity value of bank $i$ is $\pi_i = (R - \psi)$. Thus the large shock, $\varepsilon_L$, is greater than the combined equity value of all banks in the system when no shocks occur. Shocks of this magnitude are possible in a highly leveraged financial system. If there were never any large shocks it would be possible to construct a financial network in which there were never any failures. In comparison, shocks $\varepsilon_S$ are relatively small. A shock $\varepsilon_S$ is smaller than the cumulative equity value in the system, but still larger than the equity value of a single bank, absent trade and when no shocks hit. Thus financial connections can redistribute exposures and prevent failures when a small shock hits, but cannot avoid failures altogether when a large shock hits.

Although any shock is rare, small shocks are relatively common compared to large shocks. Thus $q$ is small. If $q$ were large, a sparse architecture such as the empty network would minimise the expected number of failures. We define sufficiently rare shocks by:

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

2.4 Discussion of the Model

Our model attempts to capture a key tradeoff. Financial contracts prevent failures in some states, by hedging against banks’ idiosyncratic risk from their real investments. However financial contracts also generate systemic risk, whereby multiple banks might fail due to greater connectivity. As idiosyncratic risks are shared, market values become correlated. Moreover, when one bank fails, its financial distress costs reduce counterparties’ market values. These forces allow multiple failures, and hence financial contagion, to potentially occur.

To capture this tradeoff, we need to consider contracts that satisfy two key properties. First, the contracts must help banks to avoid failures from small shocks by allowing them to hedge their exposures to their proprietary investments. A bank’s contracts must spread potential losses from its real investment amongst its counterparties before it fails. Standard debt agreements do not permit this feature. Second, there must be counterparty risk. If a bank $j$ has a contract with a bank $i$, then the failure of bank $i$ must reduce the value of bank $j$’s contract all else equal. In practice there is huge

---

8 In section 5 we generalize the shock distribution in several respects.
9 Allen and Gale (2000) take a different approach. In their model failure occurs due to a lack of liquidity, which banks can overcome by providing short term funding to one another. In our setting failure only occurs when banks become insolvent, due to a fall in the value of their investments.
10 See Capponi (2013) for a discussion of contracts exhibiting counterparty risk.
heterogeneity in the financial contracts banks have with one another and many contracts satisfy these two properties.\footnote{For example, unsecured, bilaterally traded, over-the-counter swaps of risky cashflows between banks have these two properties. When \(i\) and \(j\) exchange claims, they are pledging risky cashflows from their real investments to one another. Post-financial crisis, inter-bank swaps are relatively heavily collateralised, which is not modelled in our contracting space. Nevertheless, we would expect the value of an inter-bank swap that bank \(j\) has with bank \(i\) to fall were bank \(i\) to fail.} Instead of picking one or more such contract and modeling them in detail, we take a parsimonious approach and consider the simplest contracts with the key features we need.

With a richer contracting space the tradeoff between hedging and minimising exposure to contagion would disappear. For example, suppose banks \(i\) and \(j\) could write a contract that induced a cash flow from \(i\) to \(j\) if and only if a small shock (but not a large shock) hit \(i\), and a cash flow from \(j\) to \(i\) if and only if a small shock (but not a large shock) hit \(j\). Banks \(i\) and \(j\) could have strong financial interdependencies following small shocks but no interdependencies following large shocks. These contracts always prevent failures after a small shock, and allow only a single failure after large shocks.

The model permits multiple interpretations of the costs of financial distress. There may be costs associated with liquidating the real investment, possibly at a discount during periods of financial turmoil\footnote{See Cifuentes et al. (2005), Gai and Kapadia (2010) or Caballero and Simsek (2013) for a more detailed treatment of fire sales in financial networks.}; of inefficient allocation of resources during the bankruptcy period; a loss of human capital associated with firing workers; a redirection of cash flows generate by the project towards the cost of bankruptcy negotiations, administration and settlement before they are shared; and so forth. In practice, costs of financial distress are likely to be large.\footnote{For example, Hennessy and Whited (2007) estimate costs of financial distress of 8-15 percent of capital, for a panel of non-financial firms. Given the likelihood of fire sales during periods of financial contagion, the costs for financial institutions may be even higher.}

The reason for financial interconnections in our model is hedging, which is responsible for a large part of financial flows (BIS, 2015b). By exchanging financial connections with one another, banks can sometimes avoid failures when a shock hits their real investment—which may be desirable from the perspective of banks, and also a hypothetical social planner. Another important reason for financial interconnections is intermediation (e.g. Farboodi, 2014; Acemoglu et al., 2015b). To some extent, hedging and intermediation may resemble one another. In 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. Moreover, our model captures some aspects of intermediation directly. Bank \(i\) makes a real investment on behalf of its equity holders, whose funds come from outside the network. Bank \(i\) hedges through trades with counterparties. These trades transfer cashflows from the counterparties’ investments to \(i\)’s equity holders. In effect, the bank intermediates between \(i\)’s equity holders, and other banks’ investments, and so allows the equity holders to hedge their exposure to \(i\)’s investment.

2.5 Example

To illustrate the model, we consider an example. We consider networks with various strengths of inter-bank claims. Inter-bank claims can serve to prevent failures after small shocks, but also may cause multiple failures after large shocks.

Suppose there are \(n = 6\) banks, which we label \(A, B, C, D, E, F\). We let each bank have an equity
value, conditional on no shocking hitting, of $\pi_i = R - v = 1$; let the large shock be $\varepsilon_L = 10 > n\pi_i$; the small shock be $\varepsilon_S = 3$, with $\pi_i < \varepsilon_S < n\pi_i$, and also set financial distress costs to be $\beta = 1$. Finally, the probability of a small shock is 0.0098 and the probability of a large shock be 0.0002. Thus there is a one in a hundred chance of a shock occurring and a one in five thousand chance of a large shock occurring. Some natural networks to consider are the empty network, a network with two groups of 3 banks, a network with 3 groups 2 banks and the complete network, with perfect risk-sharing within-group and no links between groups. In the network with two groups of three banks, we let group 1 be $G_1 = \{A, B, C\}$ and group 2 be $G_2 = \{D, E, F\}$. We then set $A_{ij} = 1/3$ if $i$ and $j$ are in the same group and $A_{ij} = 0$ otherwise. In the network with three groups of two banks we let $G_1 = \{A, B\}$, $G_2 = \{C, D\}$, $G_3 = \{E, F\}$ and set $A_{ij} = 1/2$ if $i$ and $j$ are in the same group and $A_{ij} = 0$ otherwise. In the complete network we set $A_{ij} = 1/6$ for all $i$ and all $j$ and in the empty network we set $A_{ij} = 0$ for all $i$ and all $j \neq i$. These networks are illustrated in Figure 2.

![Network Diagrams](image)

(a) Empty network  (b) Three groups of two  (c) Two groups of three  (d) Complete network

Figure 2: Four possible financial networks among six banks. The nodes represent the banks. In these networks all holdings are symmetric such that $i$ has the same size claim in $j$’s project as $j$ has in $i$’s project. The width of the link connecting $i$ to $j$ is proportional to the size of these claims.

Table 1 shows how many failures occur for each of the candidate networks illustrated in Figure 2 following a small shock and following a large shock. The last column reports the overall expected number of failures. The network comprising of three groups of two banks obtains the lowest expected number of failures. It is sufficiently connected to avoid any failures following the relatively common small shock, while the number of failures following a large shock are also limited by separating the banks into two groups.
<table>
<thead>
<tr>
<th></th>
<th>Number of failures after a small shock</th>
<th>Number of failures after a large shock</th>
<th>Expected number of failures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empty</td>
<td>1</td>
<td>1</td>
<td>0.0100</td>
</tr>
<tr>
<td>Three groups of two</td>
<td>2</td>
<td>2</td>
<td>0.0200</td>
</tr>
<tr>
<td>Two groups of three</td>
<td>0</td>
<td>3</td>
<td>0.0006</td>
</tr>
<tr>
<td>Complete</td>
<td>0</td>
<td>6</td>
<td>0.0012</td>
</tr>
</tbody>
</table>

Table 1: The performance of different networks in avoiding failures.

3 Socially Efficient Networks

Our model captures a key tradeoff of financial connections. They prevent failures after relatively small shocks, but spread contagion after large shocks. This section considers how to optimally structure networks given this tradeoff—how can the social planner prevent failures from small shocks, while at the same time minimizing contagion?

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. The social planner therefore has a utilitarian objective function. However, as all agents in our model are risk neutral and utility is transferable, an outcome maximizes this objective if and only if it is Pareto efficient.

We constrain the social planner in two ways. First, as with the individual banks, the planner knows the probability distribution of cashflows and not the realization. Second, we impose a participation constraint on banks. The planner must choose a network that does not decrease the expected market value of any bank below its autarky market value,\(^{14}\) so that \(v_i(A) \geq R - (r/n) (q\varepsilon_L + (1-q)\varepsilon_S + \beta)\) for all \(i\).

The social planner must therefore choose a feasible financial network to solve the following optimization problem:

\[
\max_{A \in \mathcal{A}} \mathbb{E} \left[ \sum_{i \in N} \pi_i(A) + \delta_i(A) \right],
\]

where \(\mathcal{A}\) is the set of column stochastic and non-negative matrices \(A\) satisfying the participation constraints. A final remark is worth making.

Remark 1. The social planner's problem is equivalent to minimizing the expected number of defaults subject to the same constraints.

3.1 Example Continued

Consider again the example from Section 2.5.\(^{15}\) Recall that we considered organizing the banks into separate groups, with perfect risk sharing within group, and no links between groups. Within this class of networks, the best option for the social planner was to organize the banks into two groups of three.

\(^{14}\) I.e. the expected market value if the bank has no inter-bank assets or liabilities.

\(^{15}\) There are \(n = 6\) banks \(A, B, C, D, E, F\) with equity values \(\pi = 1\), shock sizes \(\varepsilon_L = 10\) and \(\varepsilon_S = 3\) and bankruptcy costs \(\beta = 1\). The probability of a small shock is 0.0098 and the probability of a large shock is 0.0002.
In this case, three banks would fail following a large shock but none would fail following a small shock, and expected failures were 0.0006. Can the social planner improve on this outcome?

Suppose the social planner added weak links across groups—as opposed to leaving them fully segmented. The network still contains three groups. As long as these weak links are sufficiently small, banks outside the group will not fail when a small shock hits. However, the weak links allow banks outside the group to absorb some of the shock, and potentially prevent failures in certain states. Suppose we organize the banks into three groups of two and set $A_{ji}^* = 1/3$ when $i$ and $j$ are in the same group, while if $i$ and $j$ are in different groups $A_{ji}^* = 1/12$.\(^{16}\) This network is shown in Figure 3. There are no failures when a small shock hits, since losses to each bank within the group are $\varepsilon S/3 \leq 0.1$. There are only two failures when a large shock hits. Banks inside the group fail, but each bank outside the group absorbs losses of only $(\varepsilon L + 2\beta)/12 \leq 0.1$, and so remains solvent. Expected failures from this network are then 0.0004, lower than any network with no links between groups. Expected failures fall by $1/3$ relative to organizing the banks into two groups of three without links between groups.

![Network with Between-Group Claims](image)

**Figure 3:** A financial networks among six banks with three groups of two banks such that there are strong claims within group and weak claims across group. The nodes represent the banks. In the network, inter-bank claims are symmetric such that $i$ has the same size stake in $j$’s project as $j$ has in $i$’s project. The width of the link connecting $i$ to $j$ represents the size of these claims.

**Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015a)** also emphasise that weak links between groups can reduce defaults. However, it is not clear whether this structure is generally optimal. We turn now to this question.

### 3.2 Social Planner’s Solution

Before characterising the social planner’s solution we define a particular network structure. First we define $d^*$ to be the unique positive root of

$$d_i^2 (R - \nu)\beta + d_i ((R - \nu)(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta) + \varepsilon_S(n(R - \nu) - \varepsilon_L) = 0.\quad (10)$$

\(^{16}\) This network is feasible as $\sum_j A_{ji} = 2(1/3) + 4(1/12) = 1$ for all $i$. 

---

12
For the rest of this section we abstract from the integer problems by assuming that

(i) \( d^* \) is an integer; and

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

Under these integer assumptions we can partition the banks into \( n/d^* \) groups of \( d^* \) banks. Now we define the class of networks \( A^*(d^*) \). Letting \( G_i \) be the group that \( i \) belongs to:

\[
A^*(d^*) := \left\{ A \in A : \begin{array}{l}
|G_i| = d^* \quad \text{for all } i \in N \\
A_{ji}^* = \frac{R-v}{\varepsilon_S} \quad \text{for all } i, j : G_i = G_j \\
A_{ji}^* = \frac{R-v}{\varepsilon_L + \beta d^*} \quad \text{for all } i, j : G_i \neq G_j
\end{array} \right\}.
\]

For the example in Figure 3, \( d^* = 2 \) and the network is in the class of networks \( A^*(2) \). More generally, the class of networks \( A^*(d^*) \) are symmetric networks where organizations are partitioned into groups of size \( d^* \), with strong connections within groups, and weaker connections between groups. By construction, when a small shock hits any bank \( i \), \( A_{ji}^* \varepsilon_S \leq \sum_{k \in N} A_{jk} R - v \) so that no banks fail. When a large shock hits a bank \( i \), banks in \( G_i \) fail but banks outside \( G_i \) remain solvent. In particular, for a bank \( j \notin G_i \) we have

\[
v_j = \sum_{k \notin G_i} A_{jk}^* R + \sum_{k \in G_i \backslash \{i\}} A_{jk}^* (R - \beta) + A_{jk}^* (R - \beta - \varepsilon_L) = v.
\]

Thus \( d^* \) banks fail following a large shock to \( i \) and no banks fail following a small shock to any bank.

Appealing to metaphor, we refer to the across-group links as firebreak links—since they prevent failures from spreading beyond the group hit by a large shock. We say that a network \( A \in A^*(d^*) \) is \( d^* \)-clustered, and call a given group in this network a cluster. The network representation of claims for an \( A \in A^*(6) \) is shown in Figure 4.

We can now state our first key result.

**Proposition 1.** Under the maintained assumptions, namely integer conditions (i) and (ii) above and \( q < \frac{1}{n^2} \), as \( r \to 0 \) a network \( A \) solves the social planner’s problem if and only if \( A \in A^* \).

### 3.3 Proof Outline

There are several steps to proving Proposition 1. First, large shocks are rare, since \( q < \frac{1}{n^2} \). Full risk sharing, such that \( A_{ij} = 1/n \) for all \( i, j \), has fewer expected defaults than any network in which a small shock results in a failure—since with full risk sharing, banks only fail after large shocks. Second, as \( r \to 0 \), the participation constraints imply that the social planner must choose a row stochastic matrix \( A \) for the network. The social planner’s problem therefore simplifies to choosing a doubly stochastic network \( A \) that minimizes the expected number of failures from large shocks, conditional on no failures from small shocks.

While these steps make the problem more tractable, the space of possible networks remains large and is hard to optimize over. We make progress by first considering a simpler problem. The solution

\[17\] To verify the last equality, equation (10), which holds at \( d = d^* \) by the integer assumption, has to be substituted in along with the values of \( A^* \).
LEMMA 1. For all doubly stochastic network structures $A$ such that no banks fail following any small shock, at least $\lceil d^* \rceil$ banks fail following a large shock to $i$.

Lemma 1 shows that at least $\lceil d^* \rceil$ banks will fail following a large shock to a bank $i$, in any network that is doubly stochastic and avoids failures after a small shock, where $\lceil d^* \rceil$ is the smallest integer greater than $d$. By the integer assumptions $d^* = \lceil d^* \rceil$ and networks $A(d^*)$ are feasible. While information about where the large shock hits might be expected to be useful for the social planner, so that the bound identified in Lemma 1 is not tight, any network $A \in A^*(d^*)$ achieves $d^*$ failures when a large shock hits $i$, and no failures otherwise.

Lemma 1 works by combining inequalities. Let $D_i$ be the set of banks that fail when a large shock hits $i$. Collectively banks outside $D_i$ cannot have overly large claims on banks in $D_i$, otherwise they would fail following a large shock. Banks within $D_i$ can have stronger claims, but not so strong that they fail when a small shock hits a bank in $D_i$. Minimising the size of $|D_i|$ subject to these constraints implies minimising $d_i$, subject to the inequality

$$d_i^2(R - \bar{v}) \beta + d_i ((R - \bar{v})(\varepsilon_L - \varepsilon_S) - \varepsilon_S \beta) + \varepsilon_S (n(R - \bar{v}) - \varepsilon_L) \geq 0.$$ 

As networks $A \in A^*(d^*)$ achieve the upper bound on the planner’s objective identified in Lemma 1, we have established the if part of Proposition 1. Showing that only networks $A \in A^*(d^*)$ are socially optimal is more tricky. Obtaining the upper bound in Lemma 1 requires several inequalities to bind, which pins down the strength of each bank’s claims. Specifically, it can be shown that it is necessary
but not sufficient for bank $i$ to have $d^*$ strong links with weight $(R - v)/\varepsilon_S$ and $n - d^*$ weak links with weight $(R - v)/(\varepsilon_L + d^*\beta)$. It turns out that the only way to wire the network, so that each bank has links with these values and the bound in Lemma 1 is achieved, is to arrange banks in a $d^*$-clustered network.

### 3.4 Discussion

Proposition 1 is our first key finding. The social planner optimally chooses a network that is $d^*$-clustered. We characterise socially efficient networks despite considering a large space of possible network structures. We do not place restrictions on the possible size of any inter-bank claims; nor on the identities of the banks that have cross-holdings with one another. Even under the participation constraints, the social planner has a great deal of flexibility. For example, some banks could have very high claims on a few projects, while other banks might have low claims on many projects.

Proposition 1 allows a better understanding of the role of financial distress 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 firebreak links between clusters. These links are relatively weak, but positively valued, since for any $A^* \in A^*$ we have:

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

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 financial distress costs could cause cascading secondary defaults. However, in the socially efficient network, the cumulative impact of the large shock and financial distress 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. Still, the weak links allowed by the firebreaks enable risk-sharing between clusters after a small shock. Consequently fewer banks within a cluster are required in order 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 2.3. While preventing failures from small shocks, firebreaks also limit the spread of contagion from large shocks.

It is suboptimal for clusters to be fully segmented, such that no bank holds claims to investments in another cluster. While full segmentation would prevent any part of the large shock from being transmitted outside the cluster, it would also prevent sharing of the small shock between clusters. To prevent failures following small shocks, clusters would then have to be larger, and more banks would fail following a large shock\(^{18}\). Consequently in the example of subsection 3.1, fully segmenting the network and allowing no links between groups, substantially increases the expected number of failures. Weak firebreak links across clusters are helpful for preventing failures.

Figure 5 illustrates how the socially optimal network varies for all possible shock sizes $\varepsilon_S$ and $\varepsilon_L \geq \varepsilon_S$, including those not considered in Proposition 1. As can be seen, our formal analysis focuses on shock sizes for which neither the empty or complete network is optimal. The argument in

\(^{18}\)In different settings, Blume et al. (2011) and Erol and Vohra (2014) find that socially efficient networks are fully segmented.
Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015a) that networks may be ‘robust yet fragile’, can also be seen in Figure 5. Set the size of the small shock to zero, so that in effect there is only one shock size as in Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015a). Moving up the $y$-axis, the optimal network changes discontinuously from the complete network to the empty network. Studying the 45 degree line also reduces the model to one shock, with the same outcome.

Considering continuous shock distributions, two important results from Cabrales et al. (2017) are that when the shock distribution is sufficiently fat tailed, maximal segmentation of banks is optimal, while the maximal integration of banks is optimal when the distribution is sufficiently thin tailed. Intuitively, in the fat tailed case, large shocks are sufficiently common that the tradeoff between avoiding failures from small shocks and preventing contagion from large shocks is resolved by the corner solution minimizing contagion. The opposite is true in the thin tailed case. Our focus is on the intermediate case—where it is not optimal to ignore the risk of large cascades of failures; nor to neglect the hedging role of the financial system in helping to avoid small shocks. Cabrales et al. (2017) also consider this intermediate case. However, the absence of financial distress costs from their model prevents the pattern of connections in socially optimal networks from being pinned down. In this intermediate case, modeling financial distress costs, we find that socially optimal networks are $d^*$-clustered. This yields the insight that firebreaks are optimal. To contain contagion propagated by the costs of financial distress, the social planner would like to group highly interdependent banks together, with only weak links going outside of these groups. Finally, in our model, when the costs of firebreaks are 0, the social planner is indifferent about all network structures. It is costs of financial distress that make the structure of financial networks matter.

Some comparative statics help sharpen the intuition behind Proposition 1.

Remark 2. The following comparative statics results hold regarding the social planner’s solution:
(i) The optimal cluster size \(d^*\) is increasing in financial distress costs \(\beta\), the size of the shocks \(\varepsilon_S\) and \(\varepsilon_L\); and decreasing in firms’ equity value \(R - v\).

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

(iii) The strength of firebreak links is decreasing in financial distress costs \(\beta\), the size of shock \(\varepsilon_L\); weakly decreasing in the size of the shock \(\varepsilon_S\); and increasing in firms’ equity value \(R - v\).

The social planner’s solution selects \(d^*\)-clustered networks, minimizing the extent of cascades conditional on preventing failures from small shocks. When the size of the small shock increases, or banks have less equity value, larger clusters are required to absorb small shocks and avoid failures within cluster. In the limiting case when the small shock is as large as \(n(R - v)\), the total equity value in the network, all banks must be in the same cluster to prevent a failure. When large shocks are bigger, or financial distress costs are higher, preventing contagion from spreading beyond a cluster is also more difficult. Preventing contagion requires increasing the size of clusters, to absorb more losses following a large shock within the cluster. The same forces require firebreak links to be weaker when large shocks are bigger or financial distress 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—because following a large shock there are more total losses in the system, due to more bankruptcies.

Proposition 1 relies on some strong assumptions. In Section 5, we show some ways in which the key insight of Proposition 1—that firebreaks are optimal—generalizes.

### 3.5 Shareholders vs. Debt holders

The social planner aims to maximize the expected payments to shareholders and debt holders. Crucially, and in a well-defined sense, this turns out to imply that the social planner’s solution favors external debt holders 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 \(\mathcal{A}'\).

(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 debt holders 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 debt holders receive back the face value of their debt, so maximising their expected payments. In this sense, external debt holders’ incentives align with the social planner’s. The social planner would like to channels value across the network from the shareholders of healthy banks to the debt holders of distressed banks to avoid failures. As a result, the socially efficient networks
have the property that following a shock, many outside shareholders’ find themselves with little equity value.

That the socially optimal network favors debt above equity is crucial to our analysis. It generates a split between society’s and debt holders’ interests on the one hand, and shareholders’ on the other. Moreover, we will subsequently assume that exclusively shareholders determine the network structure.

There is a conflict of interest between debt and equity, stemming from a combination of limited liability and the inability of debt holders to control cashflows. This conflict is risk-shifting, and dates back to Jensen and Meckling (1976). It is well understood at the individual firm level. Our aim is to examine the shareholder/debt holder conflict in a networked setting. We aim to explore the consequences of risk-shifting for systemic risk, as opposed to the effects for a single firm. Recall that the social planner redistributes surplus from the shareholders of healthy banks to the debt holders of distressed banks. As we go on to show in section 4, if shareholders undo this redistribution, they generate more systemic risk than is optimal.

4 Equilibrium Networks

In this section, we characterise the private stability of socially efficient networks. Shareholders tradeoff two competing forces. Due to risk-shifting, shareholders will have a conflict with debt holders and want to form financial networks that take on more risk than the debt holders would like. Proposition 2 shows that this extends to a conflict of interest with the social planner. On the other hand, if another bank takes on excess risk, shareholders can be exposed to a possible cascade of failures that devalues expected shareholder value. As a result shareholders might be reluctant to enter into trades that result in their counter-party taking on excessive risk. In this section we study, when, if ever, such concerns can enable the financial system to self-regulate to the extent that efficient networks are stable.

We assume that investment cashflows are unknown at the beginning of the period, although the distribution they are drawn from is common knowledge. Banks can exchange claims on each others’ investments. We assume that banks are risk-neutral equity value maximizers. Bank \( i \) maximizes

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

by trading direct claims with other banks, where \( \Theta \) is the set of possible states of the world, and \( P(\theta) \) is the probability of that state occurring. The different states of the world correspond to whether a shock hits, and if so, to which bank, and of what size.

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 investments. Suppose too that financial distress costs are sufficiently large that after one fails following a small shock to its proprietary investment, the other bank also fails. By increasing their claims on each other both banks may avoid failing following a small shock, thereby increasing their expected equity values. More generally, by arranging their liabilities in certain ways, banks diversify away from exposure to their own investment, 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 equity value. Hedging seems to be a key reason for financial interconnections. To be sure, there exist numerous other causes, such as financial intermediation (see e.g. Farboodi, 2014). 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.

4.1 External Creditors

Importantly, we assume that banks take the level of external debt as fixed when forming financial connections. Equivalently, we assume that these levels are determined by some unmodelled process before inter-bank claims are chosen, such that external liabilities cannot be adjusted after inter-bank contracts have been formed. Implicitly, external debt holders are unable to monitor the inter-bank contracts; and cannot issue debt conditional on the inter-bank network formed. We expect debt holders to adjust the cost of their debt to account for risks associated with the formation of the financial network—while leaving this step unmodelled.

4.2 Trading Process and Stable Networks

To give the efficient networks the best chance of being stable, we limit the opportunities of banks to trade away from an efficient configuration of claims by only permitting banks to engage in fairly limited set of bilateral trades. Specifically, we permit banks to exchange claims in each other’s proprietary investments.

**Definition 1.** Claims $A' \in A$ can be reached from $A$ through a *feasible bilateral trade* between $i$ and $j$ if for all banks $k \in N$:

- $A'_{ij} - A_{ij} = A_{jj} - A'_{jj}$ and $A'_{ji} - A_{ji} = A_{ii} - A'_{ii}$
- $A_{kl} = A'_{kl}$ for either $k \neq i, j$ or $l \neq i, j$

Bilateral trading also fits well with our focus on hedging. Bilateral trades are sufficient for this purpose, especially given that we do not restrict any two banks from trading with each other.\(^\text{19}\)

We look for configurations of claims 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 unstable.

**Definition 2.** A private trading network $A$ 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 claims $A'$ such that:

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

\(^\text{19}\) For studies of network formation and intermediation in financial markets see Farboodi (2014); Erol and Vohra (2014); Erol (2015); or Acemoglu et al. (2015b). The overwhelming share of inter-bank derivatives trading is over-the-counter, as opposed to exchange traded. While OTC contracts are increasingly cleared through a central counterparty, a substantial share remains bilaterally traded (BIS, 2015b).
4.3 Example Continued

Consider again the example from Section 2.5. There are \( n = 6 \) banks \( A, B, C, D, E, F \) with equity values, conditional on no shock occurring, of \( \pi = R - \nu = 1 \), shock sizes \( \varepsilon_L = 10 \) and \( \varepsilon_S = 3 \) and financial distress costs \( \beta = 1 \). The probability of a small shock is 0.0098 and the probability of a large shock is 0.0002. By Proposition 1, the network illustrated in Figure 3 is socially optimal and minimizes the expected number of failures. In this network claims are \( A_{ij} = 1/3 \) if \( i \) and \( j \) are in the same group, and \( A_{ij} = 1/12 \) if \( i \) and \( j \) are in different groups. However, this network is not be stable.

Suppose that bank \( A \) and bank \( C \) traded to reduce their claims on each others’ projects. Specifically, consider the feasible bilateral trade which generates new inter-bank claims \( A'_{AC} = A'_{CA} = 0 \) and \( A'_{AA} = A'_{CC} = 5/12 \). Table 2 shows the equity values of banks \( A \) and \( C \) before and after this trade in the different possible states of the world.

<table>
<thead>
<tr>
<th></th>
<th>No Shock</th>
<th>Small Shock</th>
<th>Large Shock</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td><strong>Before Trade</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \pi_A )</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \pi_C )</td>
<td>1</td>
<td>3/4</td>
<td>3/4</td>
</tr>
<tr>
<td><strong>After Trade</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \pi'_A )</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \pi'_C )</td>
<td>1</td>
<td>11/12</td>
<td>3/4</td>
</tr>
</tbody>
</table>

Table 2: Equity values of banks \( A \) and \( C \), before and after the trade between banks \( A \) and \( C \), in different states of the world. Failures are indicated in bold and red.

From Table 2 it can be seen that the trade induces additional failures. With the trade Bank \( A \) fails following a small shock to it. Moreover, this failure of \( A \) reduces the value of \( B \)'s claims in \( A \) and, although not shown in Table 2, causes bank \( B \) to also fail. Nevertheless, \( A \)'s equity value remains unchanged at 0 because of limited liability. Following a small shock to bank \( A \), bank \( C \)'s equity value changes in two ways. Without the trade, bank \( C \) has a 1/12th claim on \( A \)'s project, and the small shock causes \( C \)'s equity value to fall to 3/4. With the trade bank \( C \) no longer has a claim on \( A \)'s project, and so bank \( C \) does not suffer directly from a shock to \( A \). However, as a small shock to \( A \) now causes bank \( B \) to fail, the value of \( C \)'s claim in \( B \)'s project falls. As \( C \) has a 1/12th claim on \( B \)'s project, and the costs of financial distress are 1, this lowers \( C \)'s equity value by 1/12. Overall, with the trade, \( C \)'s equity value is higher when a small shock hits \( A \)'s project, while \( A \)'s equity value is unchanged. Similarly, \( A \)'s equity value is higher with the trade when a small shock hits \( C \)'s project, while \( C \)'s equity value is unchanged. When a small shock hits any other bank, equity value is unchanged.

When a large shock hits \( A \) or \( B \), \( A \) and \( B \) both fail with and without the trade. Thus \( A \)'s equity value remains unchanged. However, \( C \)'s equity value is higher with the trade as \( C \) then has no claims on \( A \)'s project. Indeed, the only losses that \( C \) suffers are because \( B \) fails, leaving \( C \) with an equity value of 11/12. When a large shock hits \( B \), \( A \) and \( B \) both also fail with and without the trade. While without the trade \( C \) suffers losses from three sources—the initial shock, the failure of \( B \) and the failure of \( A \)—with the trade \( C \) no longer suffers losses from the failure of \( A \). This leaves \( C \) with an equity value of 1/12. Similarly, after a shock to \( C \), in comparison to without the trade, \( C \)'s equity value is unchanged while \( A \)'s increases substantially; and after a shock to \( D \), \( C \)'s equity value remains unchanged, while \( A \) equity
value increases a little.

Overall the trade increases the expected equity values of both \( A \) and \( C \), while substantially lowering the social planner’s objective by causing multiple failures to now occur from a small shock to either \( A \) or \( C \).

We now adjust parameters to show that high financial distress costs can induce stability. When financial distress costs are sufficiently large, bank \( i \) will prevent its counterparty from generating more financial contagion. The resulting contagion would lower \( i \)'s equity value, in the states in which it does not fail. We increase financial distress costs threefold, so that \( \beta = 3 \). To hold constant the socially optimal cluster size of \( d^* = 2 \), we decrease the size of the small shock to \( \varepsilon_S = 8/3 \). The strength of within group links in the socially optimal networks are now \( 3/8 \), while the strength of across group links in the socially optimal networks are now \( 1/16 \). However socially optimal networks are now stable. For example, consider again a trade between banks \( A \) and \( C \) in which \( A \) and \( C \) reduce their claims on each other’s projects to 0. Thus \( A'_{AC} = A'_{CA} = 0 \) and \( A'_{AA} = A'_{CC} = 7/16 \). Table 3 shows the equity values of banks \( A \) and \( C \) before and after this trade in the different possible states of the world.

<table>
<thead>
<tr>
<th></th>
<th>No Shock</th>
<th>Small Shock</th>
<th>Large Shock</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>Before Trade ( \pi_A )</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \pi_C )</td>
<td>1</td>
<td>( \frac{5}{6} )</td>
<td>( \frac{5}{6} )</td>
</tr>
<tr>
<td>After Trade ( \pi'_A )</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \pi'_C )</td>
<td>1</td>
<td>( \frac{13}{16} )</td>
<td>( \frac{5}{6} )</td>
</tr>
</tbody>
</table>

Table 3: Equity values of banks \( A \) and \( C \) before and after the trade between banks \( A \) and \( C \), in different states of the world. Failures are indicated in bold and red.

From table 3 it can be seen that the trade now decreases the equity value of \( A \) following a small shock to \( C \). Although \( A \) is not directly exposed to \( C \)'s project with the trade, the failure of \( C \) causes \( D \) to fail and \( A \)'s claims on \( D \)'s project fall in value. Bank \( A \)'s losses from \( D \)'s failure with trade are greater than the losses \( A \) would incur from the direct shock to \( C \)'s project without trade. Hence \( A \)'s equity value is lower with the trade when there is a small shock to \( C \). Although \( A \)'s equity value is higher with the trade following a large shock to either \( C \) or \( D \), large shocks are relatively rare and this effect is not enough to compensate \( A \)'s shareholders for the losses they would suffer following a small shock to \( C \). As a result the overall effect of the trade is to reduce the expected equity value of \( A \). Thus the shareholders of \( A \) would not be willing to enter into the trade because of the systemic risk it generates. The prospect of \( C \) failing and causing a cascade of failures that lowers \( A \)'s claims on other projects is enough to dissuade \( A \)'s shareholders from undertaking the trade. Indeed, for this new parameterization, with larger costs of financial distress, the efficient network is stable.

### 4.4 Shareholders’ Incentives

In the previous section we observed that socially efficient networks may or may not be stable. With low financial distress costs, banks optimally traded away from the socially efficient network, despite

\[20\] An equivalent argument shows that the trade also decreases \( C \)'s expected equity value.
the trade generating more failures and contagion in certain states. However with high financial distress costs, banks do not find such a deviation optimal. In this section we explore banks’ incentives further. By correlating their failures with their largest counterparties, banks use risk-shifting to raise their equity value, thus generating systemic risk. Equally, banks seek to prevent their counterparties from risk-shifting, because it raises the risk of contagion and reduces equity value in the states where the bank is solvent. As financial distress costs becomes large, the second force tends to dominate, potentially aligning shareholder incentives with the social planner.

**Remark 3.** Bank $i$’s equity value can be rewritten as

$$
\mathbb{E}\pi_i = P(v_i \geq \underline{v}) \left( \sum_{j \in N} A_{ij} \mathbb{E} [p_j | v_i \geq \underline{v}] - \underline{v} \right) - \beta P(v_i \geq \underline{v}) \left( \sum_{j \in N} A_{ij} P(v_j \leq \underline{v}) \right) + \beta \sum_{j \in N} A_{ij} \text{Cov} \left[ I_{v_j < \underline{v}}, I_{v_i < \underline{v}} \right].
$$

The first term of equation (12) is simply the value of bank $i$’s investment cashflows when $i$ is solvent. If the costs of financial distress were zero, this would be the only term in equation (12). This term captures the risk-shifting incentives of bank $i$. Bank $i$ would like to increase the value of its claims on projects when it does not fail, but does not care about the value of its claims when it does fail. The second term shows that the equity value of $i$ is reduced by the failure of its counterparties, assuming that these failures are uncorrelated with $i$’s failure. The third term in equation (12) accounts for the correlation of failures between $i$ and its counterparties. A positive correlation increases the equity value of bank $i$ and so captures bank $i$’s incentives to positively correlate its failures with its counterparties’ failures. Bank $i$ can increase its equity value by raising the covariance of its failures with its largest counterparties, and lowering the covariance with its smallest counterparties. When bank $i$ fails in the same states as its largest counterparties, its equity holders do not face these counterparties’ financial distress costs—due to limited liability. In the states when bank $i$ does not fail, it is exposed to financial distress costs only from small counterparties, limiting the impact.

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 financial distress costs. Thus shareholders can redistribute value away from external debt holders towards themselves by increasing claims on the counterparties with whom their failures are more correlated and reducing their claims on counterparties with whom their failures are less correlated. External debt holders do not control these decisions, so are unable to prevent shareholders from redistributing value in this way, leading to risk-shifting. Remark 3 shows that in financial networks, limited liability and an absence of debtor discipline create incentives for risk-shifting by equity holders, and that this creates incentives to take on systemic risk.

An opposing force may attenuate risk-shifting. Consider the second term of equation (12). Suppose that $i$’s counterparties fail in states in which $i$ does not fail. Then $i$’s equity value falls. All else equal, $i$ prefers its counterparties not to fail when $i$, itself, is solvent. If the costs of counterparty failure are sufficiently strong—that is, if $\beta$ is large—then $i$ may prevent its counterparties from taking on excess risk limiting risk-shifting.
The network structure allows this force to operate. When \( i \) and \( j \) make a trade, they could, as in the Example from Section 4.3, eliminate their direct exposure to one another. Nevertheless, \( i \) might still be exposed to \( j \)’s counterparties. Thus if a trade with \( j \) cause contagion, \( i \) may still be affected through a series of cascading defaults—which may cause \( i \) to refuse to enter into such a trade. Precisely this force allows the second network in subsection 4.3 to be stable. In the trade \( A \) and \( C \) attempt, \( A \) is not directly exposed to \( C \)’s financial distress costs afterwards. However, \( A \) is still exposed to \( C \)’s counterparty, \( D \). After the trade, systemic risk increases, and \( C \) fails in more states. \( C \)’s financial distress costs causes \( D \) to also fall in more states. \( A \) is still exposed to \( D \), and suffers substantial falls in equity value, since \( \beta \) is relatively large. Thus \( A \) prevents the trade.

Overall, shareholders want their own banks to engage in risk-shifting. However, they prefer counterparties to avoid risk-shifting, lest it generates a contagion that causes many of the banks’ counterparties to default. Moreover, as financial distress costs increase, banks have stronger incentives to prevent their counterparties from risk-shifting. In equilibrium, this force might prevent risk-shifting altogether. At the system-wide level, which effect dominates? When will we observe a socially inefficient level of systemic risk? We turn to these questions in the next section.

4.5 Socially Efficient Networks and Private Instability

We now characterise the stability of socially efficient networks. We find that socially efficient networks are stable if and only if the costs of financial distress are above a certain threshold.

**Proposition 3.** Define \( \bar{\beta} = \frac{q^L + (1-q)L}{(1-q)d - (d-1)q} \). Then

(i) For all \( \beta < \bar{\beta} \), if socially efficient networks have at least two clusters and at most \( \frac{N}{2} \) clusters, then no socially efficient network is stable.

(ii) For all \( \beta \geq \bar{\beta} \) all socially efficient networks are stable.

To illustrate Proposition 3, consider again the two examples from Section 4.3. In the first parameterization we have \( \beta = 0.1 \) and \( \bar{\beta} = 0.334 \) and all socially optimal networks are unstable. In the second parameterization, \( \beta = 0.3 \) and \( \bar{\beta} = 0.299 \), and all socially optimal network are stable.

Proposition 3 shows that socially efficient networks are unstable when financial distress costs are sufficiently small, but become stable after exceeding a threshold level. Instability at lower financial distress costs comes from banks’ incentives to take on systemic risk and correlate their failures, as in Remark 3. However, sufficiently high financial distress costs make contagion so costly for banks that they prevent their counterparties from generating systemic risk.

First, we elaborate on the causes behind the lack of stability when financial distress costs are sufficiently small. Proposition 2 shows that socially efficient networks act to redistribute value away from equity holders and towards external debt-holders. Further, by Remark 3, if banks can raise the correlation of their failures with their largest counterparties, thereby increasing systemic risk, they can increase their equity values. Such trades will redistribute value towards the shareholders of the banks and away from external debt holders. Starting from the socially efficient network, the opportunity for such trades exists. Thus the network is not stable. Given that our stability concept assumes relatively little coordination amongst banks, our finding of instability is relatively general. If we permitted more
trades in our definition of a stable network, the socially efficient networks would remain unstable and part (i) of Proposition 3 would be unaffected.

As we formally show in the Appendix, Section B, if shareholders cannot engage in risk-shifting, then socially efficient networks are stable. If shareholders no longer have limited liability, or debt holders can exert direct control over cashflows, agents do not deviate from the social optimum.

Risk-shifting is typically studied at the individual firm level. We study the implications for the financial system, beyond an individual firm. Risk-shifting generates incentives for firms to increase systemic risk, by increasing the correlation of their failures with their largest counterparties\textsuperscript{21}. Not only does the probability of bankruptcy by individual firms rise, but the likelihood of financial contagion also increases.

We now elaborate on the case of sufficiently large financial distress costs. While the shareholders of banks are indifferent about whether their own failure precipitates a cascade of failures or not, they can suffer when the failure of another bank precipitates a cascade of failures. When $\beta > \bar{\beta}$, the potential for contagion causes banks to endogenously prevent each other from engaging in risk-shifting away from the social network, because these trades increase the threat of contagion.

Cochrane (2014) argues that each bank might have endogenous incentives to self-regulate and prevent its own failure, as part of a cascade or otherwise. Our analysis clarifies these incentives in the presence of limited liability and a lack of debtor discipline. The network structure of financial independencies creates opportunities for shareholders to engage in risk-shifting and banks will not individually self-regulate. In fact, quite the opposite—banks prefer to correlate their failures with their largest counterparties, allowing domino effects to endogenously occur. Nevertheless, the financial system as a whole can sometimes self-regulate—a bank can be unwilling to take the other side of a trade that creates substantial systemic risk. In our model, the threat of contagion from financial distress costs, if they are sufficiently large, means banks prevent their counterparties from engaging in risk-shifting\textsuperscript{22}.

The key determinant of whether contagion might arise in equilibrium, is whether financial distress costs are low enough. While our results are for a stylized contract space, we expect similar findings to hold more generally. Banks seek to correlate failures with their largest counterparties because the value of a bank falls most when its largest counterparty fails. Under limited liability, these financial distress costs are passed on to external debt holders and onto other banks in the network. A cascade of failures, through the costs of financial distress, amplifies and propagates these costs. It seems likely that these effects should operate with other contracts.

The cost of debt, given by the debt payment $v$, may include the higher likelihood of default due to risk-shifting. Nevertheless, as in standard risk-shifting arguments, our results hold regardless of whether equity holders absorb these costs ex ante. The key assumption is that equity holders cannot commit to avoiding risk-shifting, and there are no mechanisms or contracts by which debt holders can control cashflows directly or discipline shareholders’ decisions\textsuperscript{23}.

\begin{footnotesize}
\begin{itemize}
\item[\textsuperscript{21}] In a non-network setting, Acharya (2009) uncovers a similar force, which he terms ‘systemic risk-shifting’.
\item[\textsuperscript{22}] This point seems more general than our model. Rogers and Veraart (2013) shows that banks may want to bail each other out when the costs of financial distress are large, while Erol and Vohra (2014) also reaches some related conclusions.
\item[\textsuperscript{23}] Much theoretical work discusses contracts—such as covenants, convertible debt, varying debt maturity or altering managers compensation—that might prevent risk-shifting. For example, Bulow and Klemperer (2014) propose an instrument
\end{itemize}
\end{footnotesize}
That financial distress costs determine the stability of socially efficient networks is potentially important for policy. On the one hand, diminishing financial distress costs reduce the threat of contagion, and allow the social planner to construct more resilient networks. However, regulatory policies that reduce the cost of counterparty exposures may be hazardous. They may undermine banks’ incentives to endogenously regulate each other and create resilient network structures. In the next section, we introduce bailouts to formalize this tradeoff.

4.6 Bailouts

We now consider a bailout policy. After a bank fails, we let bailouts reduce counterparties’ financial distress costs. Therefore by our previous analysis, they diminish the strength of contagion, but potentially at the cost of rendering socially efficient networks unstable. We show that bailouts engender an important tradeoff, distinct to a networked context. When financial distress costs are initially high, the social planner can lower the expected number of failures by introducing a partial bailout. However, sufficiently large bailouts can dampen banks’ incentives to prevent one another from risk-shifting, and render socially efficient networks unstable. Indeed when financial distress costs are low to begin with, the social planner may prefer to increase financial distress costs in the network.

To model bailouts, we assume the social planner has access to an extra tool. After a bank fails, she can reduce the effect of the failure on the bank’s counterparties by reducing the financial distress cost from $\beta$ to $\beta' \geq 0$. Thus bailouts partially insulate banks from counterparty failures. For simplicity we let the social planner finance bailouts with non-distortionary taxes from outside the network. Moreover, as a bailout can help avoid fire sales and other costly activities associated with failure, we assume it costs the government only $(\beta - \beta')\zeta$ to reduce financial distress costs by $\beta - \beta'$, where $\zeta \in (0, 1)$. Thus all else equal, bailouts are efficiency enhancing.\(^{24}\)

The social planner now jointly chooses network structure and bailouts. She commits to providing some financial support to a failed banks’ counterparties before shocks and bankruptcies are realised, and aims to maximise expected value in the network subject to individually optimal behaviour. The network chosen by the social planner must be stable, otherwise it would be infeasible to implement. Formally, the social planner solves

$$
\sup_{A, \beta'} E \left[ \sum_{i \in N} v_i(A, \beta') + I_{v_i(A, \beta') < 0}(\beta - \beta')(1 - \zeta) \right]
$$

subject to the stability of $A$ and $\sum_{j \in N} A_{ij} = \sum_{j \in N} A_{ji} = 1$ for all $i \in N$.

**Proposition 4.**

(i) If $\beta > \bar{\beta}$ the social planner can increase her objective by providing financial assistance to the counterparties of a failed bank and setting $\beta' \in [\bar{\beta}, \beta]$.

(ii) There exists some $\tilde{\beta} \in (0, \bar{\beta})$ such that for $\beta \in (\tilde{\beta}, \bar{\beta})$, the social planner can strictly increase her objective by increasing financial distress costs and setting $\beta' = \tilde{\beta} > \beta$ (in comparison to setting resembling convertible debt to deal with risk-shifting by banks.\(^{25}\)The analysis would be similar if the social planner levied taxes inside the network.

---

\[^{24}\text{The analysis would be similar if the social planner levied taxes inside the network.}\]
\[ \beta' = \beta \).

Proposition 4 shows that both positive and negative bailouts can sometimes increase the social planner’s objective. All else equal, bailouts are efficiency enhancing, because the government reduces financial distress costs by more than the cost of the bailout. Nevertheless, as shown in part (ii), regulators may instead benefit from increasing financial distress costs. This occurs because the efficiency losses from raising financial distress costs increase continuously with the amount these costs are raised, while the benefits of being able to select a socially optimal network, unconstrained by the stability requirement, are discontinuous.

To illustrate Proposition 4 consider again the second example from Section 4.3 in which there are six banks; equity values \( \pi_i = 1 \) for all banks \( i \); shock sizes \( \varepsilon_L = 10 \) and \( \varepsilon_S = 8/3 \); bankruptcy costs \( \beta = 3 \); probability of a small shock 0.0098; and probability of a large shock of 0.0002. As we have seen, the socially optimal network is stable. Moreover, as \( \beta > \overline{\beta} \), the network will remain stable for a marginal decrease in the costs of financial distress. Thus, at the margin, reducing the costs of financial distress from their initial level increases the planner’s objective. Fixing the network structure, in general decreasing bailout costs weakly decreases the number of banks that fail in each state,\(^{25}\) and by assumption the cost to the planner of providing support is less than the value of the support to the banks.

Suppose now that large shocks become marginally more likely, relative to small shocks. While the overall probability of a shock remains fixed at \( r = 0.01 \), let the probability of a large shock increases from \( 1/5000 \) to \( 1/4900 \). For these parameters, \( \beta = \overline{\beta} = 0.3 \). Thus the socially efficient network is stable and can be selected by the planner. However, for any reduction in \( \beta \), so for any \( \beta' < \beta \), the planner cannot implement the socially optimal network, since it is no longer stable. Thus marginally lowering \( \beta \) causes discontinuous losses. Likewise, suppose the socially efficient network is just unstable such that \( \beta < \overline{\beta} \). For \( |\overline{\beta} - \beta| \) but sufficiently small, there will then be large gains to the social planner from slightly increasing financial distress costs and setting \( \beta' = \overline{\beta} \), so that the socially efficient network becomes stable and can be implemented.

Bailouts introduce a key tradeoff. For a given exogenous network, they prevent contagion from spreading. However, by undermining banks’ incentives to endogenously prevent contagion, they may be counterproductive.

5 Firebreaks and Robustness

A key argument in our paper is that efficient social networks exhibit 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 \( \varepsilon \) has only two points of support, and that all banks are of identical size. In this section, we argue that the importance of firebreaks is more general.

Recall that for the social planner to satisfy the banks’ participation constraints she has to choose a doubly stochastic network \( A \), such that \( \sum_{j \in N} A_{ij} = 1 \) for all \( i \in N \), in the limit as

\(^{25}\) Although starting from a socially efficient network a marginal reduction in the costs of financial distress leaves failure sets unchanged.
the probability of a shock becomes small \((r \to 0)\). In this section it will be helpful to compare the relative value of the social planner’s objective function for different networks and shock distributions. However, as \(r \to 0\), these differences become small. Indeed, for \(r = 0\), the planner is indifferent about all feasible networks. To avoid this problem, in this section we therefore fix \(r > 0\), but for tractability continue to constrain the social planner to choose a doubly stochastic network.

### 5.1 Shock Size Uncertainty

First 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 \([\varepsilon_L, \varepsilon_L]\), the small shock is drawn from a distribution \(F_S\) with support \([\varepsilon_S, \varepsilon_S]\) and the probability of a large shock rather than a small shock hitting is drawn from a distribution \(F_q\) with support \([0, \bar{q}]\) for some \(\bar{q} < \frac{1}{n^2}\). As before, each shock hits a single bank’s underlying investments 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 \(\varepsilon_L\) for sure, the small shock were \(\varepsilon_S\) for sure and the probability of the large shock were \(\bar{q}\). We abuse notation by writing \(A^*(\varepsilon_S, \varepsilon_L, q) = A^*(\varepsilon_S, \varepsilon_L)\) to represent the set of networks \(A^* (d^*(\varepsilon_S, \varepsilon_L, q)) = A^* (d^*(\varepsilon_S, \varepsilon_L))\).

**Proposition 5.** For all \(\bar{q} < \frac{1}{n^2}\), 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 \(A^* \in A^*(\varepsilon_S, \varepsilon_L, \bar{q})\) is bounded from above by:

\[
\phi(F_S, F_L, F_q) := \frac{\bar{q} \beta (d^*(\varepsilon_S, \varepsilon_L) - d^*(\varepsilon_S, \varepsilon_L))}{nR - r(1 - \bar{q})\varepsilon_S - r\bar{q} (\varepsilon_L + \beta d^*(\varepsilon_S, \varepsilon_L))}.
\]

Proposition 5 bounds the percentage losses incurred by a social planner implementing a network \(A^* \in A^*(\varepsilon_S, \varepsilon_L, \bar{q})\) instead of an optimal network, for distributions \(F_L, F_S\) and \(F_q\). Moreover, the upper bound on the losses from imposing network \(A^* \in A^*(\varepsilon_S, \varepsilon_L, \bar{q})\) is tight if and only if the large shock has size \(\varepsilon_L\) for sure, and the small shock has size \(\varepsilon_S\) for sure. When there is uncertainty over shock sizes, that is \(F_S(\varepsilon_S), F_L(\varepsilon_L) < 1\), the social planner loses strictly less than the bound on losses established in proposition 5, by choosing a network \(A^* \in A^*(\varepsilon_S, \varepsilon_L, \bar{q})\). This implies that the value of social planner’s objective function from choosing \(A^* \in A^*(\varepsilon_S, \varepsilon_L, \bar{q})\) converges smoothly to its optimum at \(\varepsilon_S = \varepsilon_S\) and \(\varepsilon_L = \varepsilon_L\), as the supports of \(F_S\) and \(F_L\) converge to \(\varepsilon_S\) and \(\varepsilon_L\) respectively. 26 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 tradeoff facing the social planner faces remains relevant beyond the very simple two point support of the shocks we analyse, 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 that \(F_L \in F_L, F_S \in F_S\) and \(F_q \in F_q\). Abusing notation, let \(\varepsilon_L\) be the maximum size a large shock can take in any distribution \(F_L \in F_L\) and suppose the degenerate distribution in

\[26\] In contrast, if \(F_S\) and \(F_L\) are atomless, the value of planner’s objective function from choosing \(A^* \in A^*(\varepsilon_S, \varepsilon_L, \bar{q})\) will not converge smoothly to the value of the planner’s optimal network at \(\varepsilon_S = \varepsilon_S\) and \(\varepsilon_L = \varepsilon_L\), as the support of \(F_S\) and \(F_L\) converge to \(\varepsilon_S\) and \(\varepsilon_L\).
which the large shock has size \( \varepsilon_L \) for sure, is included in \( \mathcal{F}_L \). Similarly, let \( \varepsilon_S \) be the maximum value a small shock can take in any distribution \( F_S \in \mathcal{F}_S \), and suppose the degenerate distribution in which the small shock has size \( \varepsilon_S \) for sure is included in \( \mathcal{F}_S \). Finally, for any \( F_q \in \mathcal{F}_q \), let the maximum value \( q \) can take, denoted by \( \bar{q} \), be less than \( 1/n^2 \). We model the ambiguity aversion of the social planner, by adjusting the planner’s objective so that she maximizes the minimum possible social surplus (Gilboa and Schmeidler, 1989), and solves

\[
\max_{A \in \mathcal{A}} \min_{F_S \in \mathcal{F}_S, F_L \in \mathcal{F}_L, F_q \in \mathcal{F}_q} \mathbb{E} \left[ \sum_{i \in N} \pi_i(A) + \delta_i(A) \right].
\]

**Corollary 1.** A network \( A \) is optimal for an ambiguity averse social planner if and only if \( A \in \mathcal{A}^*(\varepsilon_S, \varepsilon_L) \).

Corollary 1 reinforces the key insights from Proposition 1. Indeed, ambiguity aversion might provide a reasonable approximation of the objective functions of policy-makers such as central banks, particularly during times of market stress.

### 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 investment cashflows \( R \), to have the same financial distress 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. To show this result, we return to our baseline model so that the shock size again takes only two values.

**Definition 3.** Consider a financial system \((A, N)\) and set of banks \( M \subseteq N \), to be merged into a new bank \( i \). The financial system \((A', N')\) results from a \((M, i)\)-merger of \((A, N)\) 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 \( j, k \notin M \).
- \( \beta_i = \sum_{m \in M} \beta_m \), \( v_i = \sum_{m \in M} v_m \), \( p_i = \sum_{m \in M} p_m \).
- \( N' = (N \cup \{i\}) \setminus M \).

Therefore a bank \( i \) created from the merger of banks \( M \) inherits the cumulative claims of the banks in \( M \). The also merger preserves the overall claims of other banks and of outside shareholders, insofar as the value of these claims remains unchanged conditional on no shock hitting. Overall, total financial distress costs, investment cashflows and failure thresholds are also preserved. Finally, we preserve the same overall shock probability. As before, a random, firm specific shock \( \varepsilon \) hits the underlying investment of one bank. However, the probability of a given investment 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 investment in the absence of a shock, given by the value of the investments belonging to the previously merged banks. Bank \( i \) is of size \( \sigma_i \) if its investment yields a cashflow of \( \sigma_i R \) in the absence of any shocks. Hence the probability of a small shock hitting bank \( i \) is \( r(1 - q) \left( \frac{\sigma_i}{\sum_{j \in N} \sigma_j} \right) \) and the probability of a large shock
hitting bank $i$ is $rq \left( \sum_{j \in \mathcal{N}} \sigma_j \right)$. By construction $A'$ is column stochastic, so is still a valid dependency matrix. By construction the mergers also make shareholders, external debt holders and counterparties with the financial system no worse or better off as the probability of a shock goes to zero ($r \to 0$).

We define the **generalized social planner’s problem** by letting the social planner choose a network structure and sequence of mergers. Thus the social planner chooses from among the set of financial networks, the nodes of which are the set of post-merger banks, such that the network can be reached by selecting a network $A \in \mathcal{A}$, and then implementing some sequence of mergers. As before, the social planner selects a network to maximize the expected sum of payments to debt holders and shareholders, $\mathbb{E} \left[ \sum_{i \in \mathcal{N}} \pi_i + \delta_i \right]$, subject to the participation constraints.

**Remark 4.** In the presence of mergers, the social planner’s objective is equivalent to minimizing the expected sum of default costs rather than simply the number of organizations that fail.

We need one more definition before stating the main result for this subsection. Starting from a $d^*$-clustered network $A \in \mathcal{A}^*(d^*)$, 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.** Under the maintained assumptions, namely integer conditions (i) and (ii) above and $q < \frac{1}{n^2}$, a network $A^*$ solves the generalized social planner’s problem if it can reached by a sequence of within-cluster mergers starting from a network $A \in \mathcal{A}^*(d^*)$.

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 $A^*$ and did this with all organizations, the optimal cluster size would double to $2d^*$. Repeatedly this exercise $k$ times, we would have $2^kd^*$ organizations in each cluster. We could then further apply Proposition 6 to create different sized organizations with much more variation than is 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 6 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 bank could have been created by the merger of 4 periphery banks. The width of the lines represents the strength of inter-bank claims. With the heterogeneous sizes the strongest links end up being across cluster, between the core banks—these inter-bank claims 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 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.

6 Conclusion

Regulation needs to account for the systemic risk caused by counterparty relations. We need to understand what features of the financial system will enhance its resilience. We must also model banks’ incentives. Will banks endogenously produce resilient networks? Can regulation encourage this process? Could policy even inadvertently prevent resilient networks from forming?

A key characteristic that allows financial systems to minimise the spread of contagion is firebreaks. In general, there is a tradeoff between minimising contagion after a large shock, and maximising risk sharing after small shocks. In our model, the system that optimally resolves this tradeoff partitions banks into groups with strong inter-bank claims within group, and weak claims between groups. When organized in this way, we term the weak claims between clusters as firebreaks, since they halt the flow of contagion.

We also explore when and why financial systems endogenously generate such features. Two contrasting forces are at work. Shareholders aim to engage in risk-shifting. In a financial networks setting, risk-shifting causes firms to correlate their failures with their largest counterparties, thus raising the likelihood of contagion. Thus risk-shifting undermines incentives for banks to endogenously create firebreaks. However, an opposing force is also at work. Shareholders want their own bank to engage in risk-shifting, but prefer counterparties to take on minimal risk. In this way, the bank minimises its exposure to a costly contagion, in states of the world in which it does not fail. Thus a bank might prevent its counterparties from executing trades that undermine firebreaks, and increase risk-shifting.

In practice, which of these two forces are likely to dominate? We argue that the size of counterparties’ costs of financial distress is the key determinant. When these costs are above a certain threshold, contagion is very harmful to equity, so that firms dissuade one another from risk-shifting. Conversely,
when these costs are below this threshold, banks will engage in risk-shifting, and will not endoge-
nously form firebreaks.

It is well understood that risk-shifting causes institutions to take on excess risk, at the individual
level. We show that risk-shifting causes banks to generate excess systemic risk. Therefore policies that
reduce risk-shifting in the financial system might diminish the likelihood of contagion. For example,
certain forms of convertible debt (e.g. Bulow and Klemperer, 2014) might ameliorate risk-shifting,
and so reduce systemic risk. Equally, regulation that reduces counterparty financial distress costs is
potentially hazardous. Bailouts, for example, can reduce the malign effects of financial distress costs,
in propagating contagion. However, they may undermine banks’ incentives to endogenously generate
resilient network structures, and lead to less resilient financial systems in equilibrium.
References


33


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

A Omitted Proofs

A.1 Proof of Remark 1

The social planner’s problem is to maximize:

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

Using equation (11) we have that:

$$
E \left[ \sum_{i \in N} \pi_i + \delta_i \right] = E \left[ \sum_{i \in N} \max\{v_i - v_j, 0\} + \min\{v_i, v_j\} \right]
$$

$$
= \sum_{i \in N} \sum_{j \in N} A_{ij} E[p_j] - \beta \sum_{i \in N} \sum_{j \in N} A_{ij} E[I_{v_j < v_i}]
$$

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

$$
E \left[ \sum_{i \in N} \pi_i + \delta_i \right] = nR - \epsilon - \beta E \left[ \sum_{j \in N} I_{v_j < v_i} \right]
$$

$$
= nR - \epsilon - \beta E[\text{no. of defaults}]
$$

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

A.2 Proof of Proposition 1

We first prove Lemma 1 from Section 3.3. This Lemma states that: For all doubly stochastic network structures $A$ such that no banks fail following any small shock, at least $\lceil d^* \rceil$ banks fail following a large shock to $i$.

For any network $A$ there is a minimum set of organizations $D_i(A)$ that will fail following a large shock to $i$. Letting $d_i(A) = |D_i(A)|$, 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.

P1: $\min_{A \in A} d_i(A)$ subject to $A_{jk} \epsilon_S \leq R - \underline{v}$ for all $j, k \in N$

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

$$
A_{ji} \epsilon_L + \sum_{k \in D_i} A_{jk} \beta \leq R - \underline{v}.
$$

Thus an upper bound on the losses absorbed by banks not in $D_i$ after a large shock to $i$ is
\[(n - d_i)(R - v).\]

The remaining losses must be absorbed by the remaining banks. Thus collectively organizations in \(D_i\) incur losses of at least
\[
\varepsilon_L + d_i\beta - [n(R - v) - d_i(R - v)],
\]
and so
\[
\sum_{j \in D_i} \left[ \sum_{k \in D_i} A_{jk}\beta + A_{ji}\varepsilon_L \right] \geq \varepsilon_L + d_i\beta - [n(R - v) - d_i(R - v)]. \tag{14}
\]

No organization fails following a small shock to \(i\), so by the constraints in P1:
\[
\sum_{j \in D_i} A_{ji} \leq \frac{d_i(R - v)}{\varepsilon_S}. \tag{15}
\]
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
\[
\sum_{j \in D_i} \sum_{k \in D_i} A_{jk} \leq \frac{d_i^2(R - v)}{\varepsilon_S}. \tag{16}
\]
Combining inequalities (14), (15) and (16) we have:
\[
\varepsilon_L + d_i\beta - [n(R - v) - d_i(R - v)] \leq \frac{d_i^2(R - v)}{\varepsilon_S} \beta + \sum_{j \in D_i} A_{ji}\varepsilon_L \leq \frac{d_i^2(R - v)}{\varepsilon_S} \beta + \frac{d_i(R - v)}{\varepsilon_S} \varepsilon_L.
\]
Rearranging:
\[
f(d_i) := d_i^2(R - v)\beta + d_i((R - v)(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta) + \varepsilon_S(n(R - v) - \varepsilon_L) \geq 0. \tag{17}
\]

The constant term of the quadratic equation is always negative, according to the definition of \(\varepsilon_L\); 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 \([d_i^*]\) banks fail when bank \(i\) is hit by a large shock.

We now prove Proposition 1.

We will first show that there can be no failures following a small shock in a socially optimal network and that the participation constraints mean that the social planner must choose a row stochastic network in the limit as \(r \to 0\). Then, using these results, we’ll show that all \(d^*\)-clustered networks are socially optimal and finally we’ll show that all socially optimal networks are \(d^*\)-clustered.
Under complete network in which $A_{ij} = 1/n$ for all $i$ and all $j$, $v_i(A) = R - qr(\varepsilon_L + n\beta) - (1-q)r\varepsilon_S$. The maximum possible value per bank that can be achieved by a network with at least one failure following a small shock is $R - r(1-q)\varepsilon_S + rq\varepsilon_L + r\beta/n$. Thus, for it never be optimal for the social planner to choose a network in which any bank fails following a small shock it is sufficient for the complete network to be preferred, requiring $q < 1/n^2$ which holds by assumption.

We now show that in the limit, as $r \to 0$, the planner will choose a row stochastic network. We do so in two steps. First we show that there exist row stochastic networks satisfying the participation constraints for all $r \in [0, 1]$. We then show that as $r \to 0$, any network that is not row stochastic will violate a participation constraint.

It is straightforward to verify that any network $A \in \mathcal{A}^*$ satisfies the participation constraints for all $r \in [0, 1]$. The participation constraints require that

$$v_i(A) \geq R - \frac{r}{n} (q\varepsilon_L + (1-q)\varepsilon_S + \beta).$$

This is satisfied for any network $A \in \mathcal{A}^*$ because $q < 1/d^*$. For any $A \in \mathcal{A}^*$,

$$v_i(A) = R - \frac{r}{n} (q\varepsilon_L + (1-q)\varepsilon_S + qd^*\beta).$$

In the limit as $r \to 0$, the participation constraints require that $\lim_{r \to 0} v_i(A) = \sum_j A_{ij} R \geq R$ for all $i$. This is equivalent to $\sum_j A_{ij} \geq 1$ for all $i$. As, by column stochasticity, $\sum_i \sum_j A_{ij} = n$, if there is any $i$ such that $\sum_j A_{ij} \neq 1$ there must exist a $k$ such that $\sum_j A_{kj} < 1$ implying that a participation constraint will be violated in the limit. Thus the planner must choose a row stochastic network $A$ in the limit as $r \to 0$.

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 (17) and assuming 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 $A \in \mathcal{A}^*$ achieve this bound.

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

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

(ii) $A_{ji} = \frac{R-v}{\varepsilon_L+\beta d}$ 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 $R - v - A_{ji}\varepsilon_S$. 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 $R - v - A_{ji}\varepsilon_L - A_{ji}d^*\beta$, which is weakly positive and so $j$ does not fail. Thus, in expectation there are $qd^*$ failures and all networks $A \in \mathcal{A}^*$ are socially optimal.
We now show that only networks $A \in A^*$ 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

$$A_{jk} \leq \frac{R - v}{\varepsilon_S} \text{ for all } j, k \in D_i \text{ and for all } i \in N. \tag{18}$$

Inequalities (14), (15) and (16) are all used to construct $f(d_i)$, and all three must bind for the lower bound $d^*$ to be achieved. Thus by equation (16)

$$\sum_{j \in D_i} \sum_{k \in D_i} A_{jk} = \frac{d^2_i (R - v)}{\varepsilon_S}. \tag{19}$$

The only way for equations (18) and (19) to hold is for

$$A_{jk} = \frac{R - v}{\varepsilon_S} \text{ for all } j, k \in D_i \text{ and for all } i \in N. \tag{20}$$

Thus, for any $j, k \in D_i$, we have $A_{jk} = \frac{R - v}{\varepsilon_S}$ 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 investment 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 investment $i$ must also bind—and again, this must hold for all $i \in N$. Therefore by equation (13) we have;

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

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

$$A_{jh} = \frac{R - v}{\varepsilon_L + \beta d^*} \text{ for all } j \notin D_i, \text{ for all } h \in D_i. \tag{21}$$

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 $A$ that achieves the lower bound must satisfy:

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

(ii) $A_{ij} = \frac{R - v}{\varepsilon_L + \beta d^*}$ 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 $A^*$ with $D_i = G_i$ for all $i$. Thus all socially optimal networks are members of $A^*$. 

}$\square$
A.3 Proof of Remark 2

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:

\[
d^2(R - \nu)\beta + d[(R - \nu)(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta] + \nu\varepsilon_S(R - \nu) - \varepsilon_S\varepsilon_L = 0
\]

Implicitly differentiating equation (22) with respect to \( \beta, \varepsilon_L, \varepsilon_S \), yields:

\[
\frac{\partial d^*}{\partial \beta} = \frac{d^*(\varepsilon_S - d^*(R - \nu))}{2d^*\beta(R - \nu) + (R - \nu)(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta},
\]

\[
\frac{\partial d^*}{\partial \varepsilon_L} = \frac{\varepsilon_S - d^*(R - \nu)}{2d^*\beta(R - \nu) + (R - \nu)(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta},
\]

\[
\frac{\partial d^*}{\partial \varepsilon_S} = \frac{d[(R - \nu) + \beta] + \varepsilon_L - n(R - \nu)}{2d^*\beta(R - \nu) + (R - \nu)(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta},
\]

\[
\frac{\partial d^*}{\partial (R - \nu)} = \frac{-[d^*\beta + d^*(\varepsilon_L - \varepsilon_S) + \varepsilon_S\beta]}{2d^*\beta(R - \nu) + (R - \nu)(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta}.
\]

We will sign these expressions. Rearranging equation (22) yields

\[
d^*(R - \nu) = \varepsilon_S - (n - d)\frac{R - \nu}{\varepsilon_L + \beta d^*},
\]

and so

\[
d^*(R - \nu) < \varepsilon_S.
\]

Since \( d^* \) is the unique positive root of equation (22), applying the quadratic formula:

\[
d^* = \frac{-[(R - \nu)(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta] + \sqrt{[(R - \nu)(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta]^2 - 4(R - \nu)\beta(\varepsilon_S(R - \nu) - \varepsilon_S\varepsilon_L)}}{(2R - \nu)\beta}.
\]

Rearranging yields

\[
2d^*\beta(R - \nu) + (\varepsilon_L - \varepsilon_S)(R - \nu) - \varepsilon_S\beta
\]

\[
= \sqrt{[(R - \nu)(\varepsilon_L - \varepsilon_S) - \varepsilon_S\beta]^2 - 4(R - \nu)\beta(\varepsilon_S(R - \nu) - \varepsilon_S\varepsilon_L)} > 0. \quad (28)
\]

We can now sign equations (23), (24), (25) and (26). In all cases the denominator is positive by equation (28). The numerator is positive in equations (23) and (24) according to equation (27); and positive in equation (25) and negative in equation (26) by the definition of a large shock.

(ii) Substituting \( \varepsilon_S = n(R - \nu) \) into equation (22) and dividing through by \( \varepsilon_S \) yields:

\[
\frac{d^2 \beta}{n} + d\left[\frac{\varepsilon_L - n(R - \nu)}{n} - \beta\right] + n(R - \nu) - \varepsilon_L = 0
\]
which has \( d^* = n \) as its (weakly) positive root.

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

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

\[\square\]

### A.4 Proof of Proposition 2

1. For any \( A \in \mathcal{A}' \) there are, by definition, no failures following a small shock. Thus the sum of all banks’ equity values is \( n(R-v) - \varepsilon_S \). After a large shock, by limited liability, the sum of all banks’ equity values is weakly positive. We show now that for all \( A^* \in \mathcal{A}^* \subset \mathcal{A}' \) 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 \( A^* \in \mathcal{A}^* \) 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 \( \max \{ 0, R-v - A_{ji} \varepsilon_L - \sum_{k \in G_i} A_{jk} \beta \} \). By the definition of the class of networks \( \mathcal{A}^* \) this simplifies to

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

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: \( E \left[ \sum_i (\pi_i + \delta_i) \right] \). Thus all networks \( A^* \in \mathcal{A}^* \) maximize \( E \left[ \sum_i (\pi_i + \delta_i) \right] \). As we have already shown that all networks in \( \mathcal{A}^* \) minimize \( E \left[ \sum_i \pi_i \right] \) within the class of networks \( \mathcal{A}' \), thus all networks in \( \mathcal{A}^* \) must maximize \( E \left[ \sum_i \delta_i \right] \) within the class of networks \( \mathcal{A}' \). \[\square\]
\section{Proof of Remark 3}

We have

\[
\mathbb{E}[\pi_i] = \sum_{\theta \in \Theta} P(\theta) \left( \sum_{j \in N} A_{ij} \left( p_j(\theta) - \beta I_{v_j(\theta) < \underline{v}} \right) - \underline{v} \right) (1 - I_{v_i(\theta) < \underline{v}}) \\
= \sum_{j \in N} A_{ij} \left[ \mathbb{E} [p_j | v_i \geq \underline{v}] P(v_i \geq \underline{v}) - \beta (P(v_j < \underline{v}) - \mathbb{E} [I_{v_j < \underline{v}} I_{v_i < \underline{v}}]) \right] - \underline{v} \mathbb{P}(v_i \geq \underline{v}) \\
= \sum_{j \in N} A_{ij} \left[ \mathbb{E} [p_j | v_i \geq \underline{v}] P(v_i \geq \underline{v}) - \beta (P(v_j < \underline{v})(1 - P(v_i < \underline{v})) - \text{Cov} [I_{v_j < \underline{v}}, I_{v_i < \underline{v}}]) \right] \\
- \underline{v} \mathbb{P}(v_i \geq \underline{v}) \\
= P(v_i \geq \underline{v}) \left( \sum_{j \in N} A_{ij} [\mathbb{E} [p_j | v_i \geq \underline{v}] - \beta P(v_j < \underline{v})] - \underline{v} \right) + \beta \sum_{j \in N} A_{ij} \text{Cov} [I_{v_j < \underline{v}}, I_{v_i < \underline{v}}].
\]

\qed

\section{Proof of Proposition 3}

Throughout the proof we use prime superscripts to denote the value of variable with the trade in question, and the absence of a prime superscript to denote the value of variables without the trade in question.

First we argue that as \( r \to 0 \), we can restrict our attention to symmetric trades, such that \( A'_{ii} - A_{ii} = A'_{jj} - A_{jj} = A_{ij} - A'_{ij} = A_{ji} - A'_{ji} \). Suppose not, and instead consider an asymmetric trade. Without loss of generality, we let \( A'_{ii} - A_{ii} = A'_{jj} - A_{jj} - \chi = A_{ij} - A'_{ij} - \chi = A_{ji} - A'_{ji} \) for \( \chi > 0 \), where \( \chi \) captures the magnitude of the asymmetry. With the trade, \( i \)'s expected value is

\[
\mathbb{E}[\pi_i'] = (1 - r) \max [(1 - \chi) R - \underline{v}, 0] + r \pi_{i,\text{shock}},
\]

where \( \pi_{i,\text{shock}} \) is the expected profit of \( i \) conditional on some shock realising. Bank \( i \)'s profit before the trade is

\[
\mathbb{E}[\pi_i] = (1 - r) (R - \underline{v}) + r \pi_{i,\text{shock}},
\]

where again, \( \pi_{i,\text{shock}} \) is the expected profit conditional on some shock realising. Note that both \( \pi_{i,\text{shock}} \) and \( \pi_{i,\text{shock}}' \) are bounded from above by \( R \). Fixing \( \chi > 0 \), there exists a \( r > 0 \) such that for all \( r < \underline{r} \), we have \( \mathbb{E}[\pi_i'] < \mathbb{E}[\pi_i] \). Thus, all asymmetric trades reduce \( i \)'s equity value for \( r \) sufficiently small. We thus restrict attention to symmetric trades.

Now consider the four possible symmetric trades between \( i \) and \( j \). The trades are:

- **Trade (1):** \( A'_{ii} - A_{ii} = A'_{jj} - A_{jj} = A_{ij} - A'_{ij} = A_{ji} - A'_{ji} = \tau > 0, j \in D_i \)
- **Trade (2):** \( A'_{ij} - A_{ij} = A'_{ji} - A_{ji} = A_{ii} - A'_{ii} = A_{jj} - A'_{jj} = \tau > 0, j \in D_i \)
- **Trade (3):** \( A'_{ii} - A_{ii} = A'_{jj} - A_{jj} = A_{ij} - A'_{ij} = A_{ji} - A'_{ji} = \tau > 0, j \notin D_i \)
- **Trade (4):** \( A'_{ij} - A_{ij} = A'_{ji} - A_{ji} = A_{ii} - A'_{ii} = A_{jj} - A'_{jj} = \tau > 0, j \notin D_i \)
We provide conditions for when, under each trade, the expected equity values of $i$ and $j$ are higher with the trade than without it, so that $\mathbb{E}[\pi_i'] > \mathbb{E}[\pi_i]$ and $\mathbb{E}[\pi_j'] > \mathbb{E}[\pi_j]$. However, as trades are symmetric, $\mathbb{E}[\pi_i'] > \mathbb{E}[\pi_i]$ if and only if $\mathbb{E}[\pi_j'] > \mathbb{E}[\pi_j]$, and so we restrict attention to $\mathbb{E}[\pi_i']$ and $\mathbb{E}[\pi_i]$.

Some additional notation will be helpful. We let $\theta_i^L$ and $\theta_i^S$ represent the states of the world in which a large and small shock hit bank $i$, respectively. By Proposition 1, all efficient networks are $d^*$-clustered. Thus, absent trade, the network will be $d^*$-clustered and we let $G_i$ denote the group of banks within the same cluster as $i$ in the initial network.

- **Trade (1).** In any state of the world $\theta$ in which a shock hits a bank other than $i$ or $j$, banks $i$’s equity value with and without the trade is the same ($\pi_i(\theta) = \pi_i'(\theta)$). When a large or small shock hits bank $i$, we have $\pi_i'(\theta_i^S) = \pi_i'(\theta_i^L) = \pi_i(\theta_i^S) = \pi_i(\theta_i^L) = 0$ for all $\tau > 0$. When a shock of size $\varepsilon_S$ hits $j$ we have

$$\pi_i'(\theta_i^S) = \max \left[ R - v - \left( \frac{R - v}{\varepsilon_S} \right) \beta(d - 1) - \left( \frac{R - v}{\varepsilon_S} - \tau \right) \varepsilon_S, 0 \right].$$

As $\pi_i'(\theta_i^S) \geq \pi_i'(\theta_i^L) \geq \pi_i(\theta_i^S) = \pi_i(\theta_i^L) = 0$, trade (1) is profitable for $i$ if and only if $\pi_i'(\theta_i^S) > 0$ for some trade size $\tau$. The equity values $\pi_i'(\theta_i^S)$ is increasing in $\tau$, and so is maximised by $\tau = \frac{R - v}{\varepsilon_S}$.

Hence trade (1) is profitable if and only if

$$R - v - \left( \frac{R - v}{\varepsilon_S} \right) \beta(d - 1) > 0.$$

This condition holds if and only if $\beta < \frac{\varepsilon_S}{d\tau}$.

- **Trade (2).** Trade (2) is profitable if and only if trade (1) is profitable, since $A_{ik} = A_{jk}$ for all $k \in N$.

- **Trade (3).** The trade is feasible for $\tau \in \left( 0, \frac{R - v}{\varepsilon_L + d\beta} \right]$. All banks in $G_i$ including $i$, now fail when a large or small shock hits $i$. When a small shock hits $i$, its value, conditional on no other failures, is now

$$v_i(\theta_i^S) = R - \left( \frac{R - v}{\varepsilon_S} + \tau \right) \varepsilon_S < v_i.$$

so $i$ fails after a small or large shock and $\pi_i'(\theta_i^S) = \pi_i'(\theta_i^L) = 0$. In both cases, the failure of $i$ causes banks $k \in G_i \setminus \{ i \}$ to also fail. Even though absent trade these banks do not fail, they have zero equity value $\pi_k(\theta_i^S) = 0$, and with the trade $i$ fails causes additional losses and their failures.

By symmetry, when a large or small shock hits $j$, all banks in the set $G_j$ fail. No other banks fail. Given that $i$ does not fail, none of the other banks will fail, and $i$ does not fail given that no other banks fail. Indeed, when a small shock hits $j$, $i$ has positive equity value

$$\pi_i'(\theta_j^S) = R - \left( \frac{R - v}{\varepsilon_L + d\beta} \right) ((d - 1)\beta) - \left( \frac{R - v}{\varepsilon_L + d\beta} - \tau \right) (\varepsilon_S + \beta)$$

$$= \left( \frac{R - v}{\varepsilon_L + d\beta} \right) (\varepsilon_L - \varepsilon_S + \tau (\varepsilon_S + \beta) > 0.$$

In comparison, without the trade,
\[
\pi_i(\theta_j^S) = R - v - \left( \frac{R - v}{\varepsilon_L + d\beta} \right) \varepsilon_S = \left( \frac{R - v}{\varepsilon_L + d\beta} \right) (\varepsilon_L - \varepsilon_S + d\beta) > 0.
\]

When a large shock hits a bank \( j \), bank \( i \) again has positive equity value
\[
\pi_i(\theta_j^L) = R - \left( \frac{R - v}{\varepsilon_L + d\beta} \right) (d - 1)\beta - \left( \frac{R - v}{\varepsilon_L + d\beta} - \tau \right) (\varepsilon_L + \beta) = \tau(\varepsilon_L + \beta) > \pi_i(\theta_j^L) = 0.
\]

When a large shock hits a bank \( k \in G_j \setminus \{j\} \), bank \( i \) again has positive equity value
\[
\pi_i(\theta_k^L) = R - \left( \frac{R - v}{\varepsilon_L + d\beta} \right) (\varepsilon_L + (d - 1)\beta) - \left( \frac{R - v}{\varepsilon_L + d\beta} - \tau \right) \beta = \tau\beta > \pi_i(\theta_k^L) = 0.
\]

Combining these changes, the change in expected equity value is
\[
E[\pi_i' - \pi_i] = \frac{r}{n} \left[ q\tau(\beta + \varepsilon_L) + (1 - q) \left( \tau(\beta + \varepsilon_S) - \left( \frac{R - v}{\varepsilon_L + \beta_d} \right) \beta d \right) \right] + \left( \frac{d - 1}{n} \right) qr\tau\beta
\]
\[
= \frac{r}{n} \left[ \beta (1 + (d - 1)q) + (q\varepsilon_L + (1 - q)\varepsilon_S) \right] - \frac{r}{n} (1 - q) \left( \frac{R - v}{\varepsilon_L + \beta d} \right) \beta d
\]
\[
(29)
\]

As equation (29) is increasing in \( \tau \), and \( \tau \in \left[ 0, \frac{R - v}{\varepsilon_L + \beta d} \right] \), there exists a profitable trade of type 3 if and only if equation (29) is positive for \( \tau = \frac{R - v}{\varepsilon_L + \beta d} \), which simplifies to\(^{27}\)
\[
\beta < \frac{q\varepsilon_L + (1 - q)\varepsilon_S}{(1 - q)d - 1 - (d - 1)q}.
\]

- **Trade (4).** We have already seen that there are profitable trades for \( \beta \leq \frac{\varepsilon_S}{d - 1} \). We will show that there are no trades of type (4) that are profitable for \( \beta \geq \frac{\varepsilon_S}{d - 1} \).

We will consider the equity values of bank \( i \) in the different possible states of the world. Suppose first that a small shock hits \( i \). Without the trade, \( j \) does not fail. If \( \tau \leq \frac{R - v}{\varepsilon_S} - \frac{R - v}{\varepsilon_L + \beta d} \), then with the trade \( j \) also does not fail. For this range of \( \tau \), with the trade, the value of \( j \) is
\[
v'_j(\theta_i^S) = R - \left( \frac{R - v}{\varepsilon_L + \beta d} + \tau \right) \varepsilon_S.
\]

Thus \( v'_j(\theta_i^S) \geq v \) if and only if
\[
\tau \leq \frac{R - v}{\varepsilon_S} - \frac{R - v}{\varepsilon_L + \beta d}.
\]

This shows that \( j \) does not fail following the small shock to \( i \). An immediate consequence is that \( i \)'s value is higher with the trade than without it, and all other banks have the same value with or without the trade (and so do not fail in either case). Indeed,

\(^{27}\)Note that denominator in the following expression is positive: \((1 - q)d - 1 - (d - 1)q = q + d - 1 - 2qd > d - 1 - \frac{1}{d^2}(2d - 1) > d - 2 \geq 0\), where in the first inequality we use \( q < \frac{1}{d^2} < \frac{1}{d} \), since it is the probability of the large shock; and in the third inequality we use \( d \geq 2 \).
\[ \pi'_i(\theta^S_i) = (R - v) - \left( \frac{R - v}{\varepsilon_S - \tau} \right) \varepsilon_S = \tau \varepsilon_S > \pi_i(\theta^S_i) = 0. \]

If instead \( \tau > \frac{R - v}{\varepsilon_S} - \frac{R - v}{\varepsilon_L + \beta d} \), then \( j \) does fail following a small shock to \( i \). As \( j \) fails, banks \( G_i \setminus \{i\} \) fail too. Without the trade these banks have values \( v_j \), and so the additional losses they suffer from \( j \) failing cause them to fail too. We then have

\[
v'_i(\theta^S_i) \leq R - \left( \frac{R - v}{\varepsilon_S - \tau} \right) \varepsilon_S - \left( \frac{R - v}{\varepsilon_S} \right) (d - 1)\beta - \left( \frac{R - v}{\varepsilon_L + \beta d} + \tau \right) \beta < 0.
\]

The first inequality comes from the assumption that, other than \( i \), only banks in \( G_j \) fail. The second inequality follows from \( \beta \geq \frac{\varepsilon_S}{d - 1} \). As \( i \) fails, \( \pi'_i(\theta^S_i) = \pi_i(\theta^S_i) = 0 \).

Suppose now that a large shock hits \( i \). Banks in \( G_i \) continue to fail with the trade. In addition, for all \( \tau > 0 \), bank \( j \) now fails. Hence \( i \)'s value can be bounded from above by assuming that only these banks fail and

\[
v'_i(\theta^L_i) \leq R - \left( \frac{R - v}{\varepsilon_S - \tau} \right) \varepsilon_L - \left( \frac{R - v}{\varepsilon_S} \right) (d - 1)\beta - \left( \frac{R - v}{\varepsilon_L + \beta d} + \tau \right) \beta < 0.
\]

Thus \( \pi'_i(\theta^L_i) = \pi_i(\theta^L_i) = 0 \).

Suppose now that a small shock hits \( j \). We saw that no bank, including \( j \), fails following a small shock to \( i \) if and only if \( \tau \leq \frac{R - v}{\varepsilon_S} - \frac{R - v}{\varepsilon_L + \beta d} \). Thus, by symmetry, for this range of \( \tau \) bank \( i \) does not fail following a small shock to \( j \) and the equity value of bank \( i \) is

\[
\pi'_i(\theta^S_j) = (R - v) - \left( \frac{R - v}{\varepsilon_L + \beta d} + \tau \right) \varepsilon_S \geq 0.
\]

Suppose instead that \( \tau > \frac{R - v}{\varepsilon_S} - \frac{R - v}{\varepsilon_L + \beta d} \). Then, with the trade, bank \( i \) will fail and \( \pi'_i(\theta^S_j) = 0 \).

In comparison, without the trade, \( \pi_i(\theta^S_j) = (R - v) - \left( \frac{R - v}{\varepsilon_L + \beta d} \right) \varepsilon_S. \) Thus, when \( \tau \leq \frac{R - v}{\varepsilon_S} - \frac{R - v}{\varepsilon_L + \beta d} \),

\[
\pi'_i(\theta^S_j) - \pi_i(\theta^S_j) = - \tau \varepsilon_S.
\]

When \( \tau > \frac{R - v}{\varepsilon_S} - \frac{R - v}{\varepsilon_L + \beta d} \),

\[
\pi'_i(\theta^S_j) - \pi_i(\theta^S_j) = -(R - v) \left( 1 - \frac{\varepsilon_S}{\varepsilon_L + \beta d} \right).
\]

Suppose now a large shock hits \( j \). We have seen that with the trade \( i \) the banks \( G_i \) fail when a large shock hit \( i \). By symmetry, the banks \( G_j \) fails when a large shock hits \( j \). Thus

\[
v'_i(\theta^L_j) = R - \left( \frac{R - v}{\varepsilon_L + \beta d} + \tau \right) (\varepsilon_L + \beta) - \left( \frac{R - v}{\varepsilon_L + \beta d} \right) (d - 1)\beta = - \tau (\varepsilon_L + \beta) < 0
\]
Hence \( i \) fails and \( \pi'_i(\theta_j^L) = \pi_i(\theta_j^L) = 0 \).

Consider a large shock to bank \( k \in G_j \setminus \{ j \} \). With (and without) the trade, banks \( G_j \) fail. The equity value of bank \( i \) is then

\[
v'_i(\theta_k^L) = R - \left( \frac{R - v}{\varepsilon_L + \beta d} \right) (\varepsilon_L + (d - 1)\beta) - \left( \frac{R - v}{\varepsilon_L + \beta d} + \tau \right) \beta = \tau \beta < v.
\]

Bank \( i \) now fails, but (by limited liability) \( i \)'s equity value is still \( \pi'_i(\theta_k^L) = \pi_i(\theta_k^L) = 0 \).

In all other states of the world \( \theta \), the value of bank \( i \) after the trade is the same as the value of bank \( i \) before the trade and so \( \pi'_i(\theta) = \pi_i(\theta) \).

Combining these changes, the expected equity value of bank \( i \) weakly decreases for \( \tau \leq \frac{R - v}{\varepsilon_S} - \frac{R - v}{\varepsilon_L + \beta q} \) and strictly decreases for \( \tau > \frac{R - v}{\varepsilon_S} - \frac{R - v}{\varepsilon_L + \beta q} \).

When \( \tau \leq \frac{R - v}{\varepsilon_S} - \frac{R - v}{\varepsilon_L + \beta d} \), the equity value of \( i \) differs with and without the trade in just two states of the world. When a small shock hits \( i \), \( i \)'s equity value is greater with the trade and the trade increases \( i \)'s equity value by an amount \( \tau \varepsilon_S \). When a small shock hits \( j \), \( i \)'s equity value with the trade is lower than without it, and the trade decreases \( i \)'s equity value by an amount \( \tau \varepsilon_S \). Thus overall, \( i \)'s expected equity value does not change.

When \( \tau > \frac{R - v}{\varepsilon_S} - \frac{R - v}{\varepsilon_L + \beta d} \), bank \( i \)'s equity value is weakly lower with the trade than without it in all states of the world, and strictly lower in some states of the world.

Summarising, there exists a trade of type (1) or (2) that is profitable for some \( \tau \) if and only if

\[
\beta < \frac{\varepsilon_S}{d - 1}.
\]

There exists a profitable trade of type (3) for some \( \tau \) if and only if

\[
\beta < \frac{q \varepsilon_L + (1 - q) \varepsilon_S}{(1 - q) d - 1 - (d - 1) q}.
\]

There is never a profitable trade of type (4) for any \( \tau \) if

\[
\beta \geq \frac{\varepsilon_S}{d - 1}.
\]

Therefore one of trades (1)-(4) is profitable for some \( \tau \), if and only if

\[
\beta < \max \left[ \frac{\varepsilon_S}{d - 1}, \frac{q \varepsilon_L + (1 - q) \varepsilon_S}{(1 - q) d - 1 - (d - 1) q} \right].
\]

By inspection,\(^{28}\) we have that

\[
\beta := \frac{q \varepsilon_L + (1 - q) \varepsilon_S}{(1 - q) d - 1 - (d - 1) q} > \frac{\varepsilon_S}{d - 1}.
\]

\(^{28}\) Note that \( \frac{q \varepsilon_L + (1 - q) \varepsilon_S}{(1 - q) d - 1 - (d - 1) q} = \frac{q (\varepsilon_L - \varepsilon_S) + \varepsilon_S}{d - 1 - (2d - 1)q} \) and so

\[
\frac{d}{\varepsilon_S} \left[ \frac{q (\varepsilon_L - \varepsilon_S) + \varepsilon_S}{d - 1 - (2d - 1)q} \right] = \frac{(\varepsilon_L - \varepsilon_S)(d - 1 - (2d - 1)q) + q(\varepsilon_L - \varepsilon_S) + \varepsilon_S)(2d - 1)}{(d - 1 - (2d - 1)q)q} > 0,
\]

since \( d \) does not depend on \( q \). We also have

\[
\frac{d}{\varepsilon_S} \left[ \frac{q (\varepsilon_L - \varepsilon_S) + \varepsilon_S}{d - 1 - (2d - 1)q} \right] = \frac{\varepsilon_S}{d - 1 - (2d - 1)q} > \frac{\varepsilon_S}{d - 1}\]
We conclude that there exists a profitable trade, of type (1)-(4), for some feasible \( \tau \), if and only if \( \beta \leq \bar{\beta} \). This completes the proof. We have shown there is no asymmetric trade that is profitable for both \( i \) and \( j \) as long as the probability of a shock, \( r \), is sufficiently small. Trades (1)-(4) encompass all trades of symmetric size. And so there exists a profitable trade in the limit as \( r \to 0 \) if and only if \( \beta < \bar{\beta} \). \( \square \)

A.7 Proof of Proposition 4

The social planner’s objective can be written as

\[
\max_{A, \beta' \geq 0} n(R - \nu) - rq\varepsilon_L - r(1 - q)\varepsilon_S - \mathbb{E} \left[ \sum_{i \in N} I_{v_i(A, \beta') < v}(\beta' + (\beta - \beta')\zeta) \right],
\]

subject to choosing a feasible and doubly stochastic network \( A \) that is stable. Note that the social planner can only influence the value of the last term. This last term is always negative.

Suppose \( \beta \geq \bar{\beta} \). Choosing any \( \beta' \in [\bar{\beta}, \beta) \), the social planner can select a socially efficient network given \( \beta' \) as, by Proposition 3, it will be stable. Doing so increases the social planner’s objective function in two ways. First, the planner directly benefits by an amount \( (\beta - \beta')(1 - \zeta) \) whenever there is a failure. Second, by Remark 2, \( d^*(\beta) \) is increasing in \( \beta \). Thus choosing \( \beta' < \beta \) reduces the cluster size in the socially efficient networks. This is beneficial for the planner as it reduces the expected number of failures from \( rqd^*(\beta) \) to \( rqd^*(\beta') \).

Let \( V(\beta', A) \) be the value of the social planner’s objective before cost savings \( \mathbb{E} \left[ \sum_{i \in N} b_i(v)(1 - \zeta) \right] \), at a given \( \beta' \) and given \( A \). Let \( \tilde{V}(\beta') \) be the supremum, over the class of stable networks, of the social planner’s objective before the cost savings at a given \( \beta' \). Recall that the set of socially efficient networks for failure costs \( \beta' \), ignoring stability, is denoted \( \mathcal{A}^*(\beta') \). We have for all \( \mathcal{A}^*(\beta') \in \mathcal{A}^*(\beta') \) and for all \( \mathcal{A} \notin \mathcal{A}^*(\beta') \)

\[
V(\beta', \mathcal{A}^*) > V(\beta', A),
\]

which follows from Proposition 1. Furthermore,

\[
V(\beta', \mathcal{A}^*(\beta')) = n(R - \nu) - rq\varepsilon_L - r(1 - q)\varepsilon_S - rq\beta'd^*(\beta')
\]

and \( V(\beta', \mathcal{A}^*(\beta')) \) is continuous and decreasing in \( \beta' \) by Remark 2. Thus, if \( \beta' \geq \bar{\beta} \)

\[\tilde{V}(\bar{\beta}) = V(\bar{\beta}, \mathcal{A}^*(\bar{\beta}))\]

by Proposition 3. We also have by Proposition 1 that only networks in \( \mathcal{A}^*(\beta') \) minimize the expected number of failures. Thus for all \( \beta' < \beta \) the stability constraint binds and

\[\tilde{V}(\beta') < V(\beta', \mathcal{A}^*(\beta')).\]

Thus

\[
\lim_{\beta' \to \bar{\beta}^-} \tilde{V}(\bar{\beta}) < \lim_{\beta' \to \bar{\beta}^-} V(\beta', \mathcal{A}^*(\beta')) = V(\bar{\beta}, \mathcal{A}^*(\bar{\beta})) = \tilde{V}(\bar{\beta}).
\]

(30)
This shows that $\hat{V}(\bar{\beta})$ is discontinuous and increasing at $\bar{\beta}$. Given initial bankruptcy costs $\beta < \bar{\beta}$ but sufficiently close to $\bar{\beta}$, a social planner can increase her objective by increasing the costs of financial distress to $\beta' = \bar{\beta}$ if

$$\hat{V}(\bar{\beta}) - \hat{V}(\beta) > rq(1 - \zeta)(\bar{\beta} - \beta)$$

(31)

For $\beta$ sufficiently close to $\bar{\beta}$ the right hand side of the inequality is an upper bound on the costs of increasing $\beta$ to $\beta' = \bar{\beta}$. As $\hat{V}(\bar{\beta})$ is discontinuous and increasing at $\bar{\beta}$, the number of failures in the socially optimal stable network must decrease at $\beta = \bar{\beta}$. Thus, for $|\bar{\beta} - \beta|$ sufficiently small, the additional costs of increasing $\beta$ to $\beta'$ will be strictly less than $\bar{\beta} - \beta$, multiplied by the expected number of failures when $\beta = \bar{\beta}$, which is $rq$, weighted by the cost saving $(1 - \zeta)$.

Note that $\lim_{\beta \rightarrow \bar{\beta}} rq(1 - \zeta)(\bar{\beta} - \beta) = 0$, but $\lim_{\beta \rightarrow \bar{\beta}} \hat{V}(\bar{\beta}) - \hat{V}(\beta) > 0$, so equation (31) holds for an initial $\beta$ sufficiently close to $\bar{\beta}$. Specifically, there exists a $\bar{\beta} < \bar{\beta}$ such that for all $\beta \in (\bar{\beta}, \bar{\beta})$ the social planner can increase her objective by increasing post bailout failure costs to $\bar{\beta}$.

A.8 Proof of Proposition 5

Fix any network $A$ and any $q \in (0, \bar{q}]$. By equation (5), 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 $\varepsilon$. Thus, the expected number of failures is smallest when $F_L(\varepsilon_L) = 1$ and $F_S(\varepsilon_S) = 1$.

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

$$S := nR - r(1 - q) \int_{\varepsilon_S}^{\varepsilon_L} \varepsilon S dF_{\varepsilon_S} - rq \left( \int_{\varepsilon_L}^{\varepsilon_S} \varepsilon L dF_{\varepsilon_L} + \beta d^*(\varepsilon_S, \varepsilon_L) \right).$$

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

$$S := nR - r(1 - q) \int_{\varepsilon_S}^{\varepsilon_L} \varepsilon S dF_{\varepsilon_S} - rq \left( \int_{\varepsilon_L}^{\varepsilon_S} \varepsilon L dF_{\varepsilon_L} + \beta d^*(\varepsilon_S, \varepsilon_L) \right).$$

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

$$\frac{S - S}{S} = \frac{rq\beta (d^*(\varepsilon_S, \varepsilon_L) - d^*(\varepsilon_S, \varepsilon_L))}{nR - r(1 - q) \int_{\varepsilon_S}^{\varepsilon_L} \varepsilon S dF_{\varepsilon_S} - rq \left( \int_{\varepsilon_L}^{\varepsilon_S} \varepsilon L dF_{\varepsilon_L} + \beta d^*(\varepsilon_S, \varepsilon_L) \right)}.$$
Finally, note that

\[
\frac{rq\beta \left( d^*_{xz} - d^*_{x}\right)}{nR - r(1 - q) \int_{\mathbb{E}} \varepsilon_{x} dF_{x} - rq \left( f_{x} \varepsilon_{x} dF_{x} + \beta d^*_{x}\right)} \leq \frac{rq\beta \left( d^*_{xz} - d^*_{x}\right)}{nR - r(1 - \bar{q}) \varepsilon_{x} - rq \left( \varepsilon_{x} + \beta d^*_{x}\right)},
\]

which completes the proof.

A.9 Proof of Corollary 1

As shown in the proof of Proposition 5, fixing for any network A, and fixing any \( q \in (0, \bar{q}] \), for any \( F_x \) and any \( F_L \), the social surplus is minimized by a shock distribution in which a shock \( \varepsilon_L \) occurs with probability \( 1 - q \) and a shock \( \varepsilon_S \) occurs with probability \( q \). Thus applying Proposition 1 the optimal network for the social planner is \( A^*_{xz} \).

A.10 Proof of Remark 4

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

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

We have:

\[
E \left[ \sum_{i \in N} \pi_i + \delta_i \right] = E \left[ \sum_{i \in N} v_i \right] = E \left[ \sum_{i \in N} \sum_{j \in N} A_{ij} (p_j - \beta \sigma_j I_{v_j < u}) \right] = \sum_{j \in N} E[p_j] - \beta \sum_{j \in N} \sigma_j E[I_{v_j < u}].
\]

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

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

Since all but the last term is exogenously given, the social planner minimizes the expected size of financial distress costs.

A.11 Proof of Proposition 6

Any feasible network \( A'' \) can be reached by starting from a network \( A \in \mathcal{A} \) (among homogeneously sized banks) and implementing a sequence of mergers. For a set of disjoint mergers \( M \), let \( \mathcal{A}' \subset \mathcal{A} \) be the subset of the initially feasible networks among \( n \) banks, such that \( A'_{k,j} = A''_{k,j} \) for all \( k, l \in M \), and all \( M \in \mathcal{M} \).
First we will show that network $A''$ with nodes $N'' \subseteq N$, can be reached by a sequence of disjoint mergers from an initial network $A' \in \mathcal{A}'$, even when the actual initial network $A \not\in \mathcal{A}'$. Take any bank $i$ in the post-merger network $A''$. This bank must comprise some subset of the initial banks $N$. Denote by $M_i$ this subset of banks. Define $A'_{kj} = \frac{\sum_{k \in M_i} A_{kj}}{|M_i|}$ for all banks $k \in M_i$ and all $j \in N \setminus M_i$. Repeating for all banks $i$ in the post-merger network $A''$ defines a network $A'$. Let $\mathcal{M}'' = \{M_i\}_{i \in N''}$ be the minimal set of of (disjoint) mergers associated with the network $A''$. For the set of disjoint mergers $\mathcal{M}''$, network $A' \in \mathcal{A}'$.

Observe that:

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

(ii) Post-merger, in a given state of the world, $v''_i = \sum_{j \in N'} \sum_{k \in M} A'_{kj} (p_j - \beta_j I_{v_j < 0})$.

By property (i), in the network $A'$ in all states of the world either all banks $k \in M$ fail, or else no bank $k \in M$ fails. Moreover, by property (ii), post merger, bank $i$ fails in $A''$ if and only if all banks $k \in M$ would have failed in network $A'$. Hence the same financial distress costs are incurred in every state of the world in the pre-merger network $A'$ as in the post merger network $A''$.

We have shown that every feasible post-merger network can be mapped into a feasible pre-merger network in $A' \in \mathcal{A}'$, such that there are the same financial distress costs in the post-merger network as in the feasible pre-merger network $A' \in \mathcal{A}'$. As $\mathcal{A}' \subseteq \mathcal{A}$ the social planner must be able to do weakly better when choosing from the set of feasible networks among the initial $n$ banks. We conclude that enabling the social planner to implement mergers, as in the generalized social planner's problem, does not help the social planner do better and achieve lower expected default costs.

We have shown that the optimized expected default costs obtained without any mergers are a lower bound on the expected default costs a social planner can achieve with mergers. We now show that this lower bound can be achieved by the networks that result from taking a socially optimal network $A^*$ and merging banks within cluster. Hence we conclude that these networks also solve the generalized social planner's problem.

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 and all $l$. Take a socially optimal network $A^*$ and consider a sequence of within-cluster mergers. Without loss of generality, let these mergers be among disjoint sets of banks and denote this set of mergers by $\mathcal{M}$. Denote the resulting network by $A''$. As argued above, this network generates the same default costs in all states of the world as the network $A'$, where $A'_{kj} = \frac{\sum_{k \in M} A_{kj}}{|M|}$ for all $k \in M_i$ and all $i \in N''$. However, as for all $M \in \mathcal{M}$, all banks in $M$ are from the same cluster in the initial network $A^*$. Thus $A' = A^*$ and the network $A''$ achieves the same expected default costs as the socially optimal network $A^*$, and so is also socially optimal.

\[\Box\]

### B  Risk-Shifting and Socially Efficient Networks

The instability of socially efficient networks is caused by two separate assumptions, namely shareholders’ limited liability, and the absence of debtor control over cashflows, which together cause risk-shifting.
Proposition 7.

(i) Suppose that shareholders have full liability so that bank $i$ seeks to maximize $E[v_i - v]$. Then all socially efficient networks are stable.

(ii) Suppose that bank $i$ seeks to minimize its probability of default $E[I_{v_i < v}]$. Then all socially efficient networks are stable.

Under full liability, the effects of limited liability are removed. Under default probability minimisation, a bank maximizes the probability that external creditors receive the face value of their debt. Therefore in this case banks act in the interests of debt holders, so that debtors effectively determine the stability of financial networks. Consequently, allowing either full liability or debtor control is sufficient to render socially efficient networks stable. Since these two assumptions are necessary for risk-shifting, preventing risk-shifting restores stability to socially efficient networks.

Proof. Part (i): After any feasible bilateral trade in which $i$ and $j$ both increase their equity value, since $r \to 0$, matrix $A'$ remains row stochastic. Suppose not, and instead consider an asymmetric trade in which $j$ receives the large side. In particular, consider a trade $A'_{ii} - A_{ii} = A'_{jj} - A_{jj} - \chi = A_{ij} - A'_{ij} - \chi = A_{ji} - A'_{ji}$ for $\chi > 0$, where $\chi$ captures the magnitude of the asymmetry. Then $i$'s expected value after the trade is

$$E[\pi'_i] = (1 - r)(1 - \chi)R - v + r\pi'_{i, shock},$$

where $\pi'_{i, shock}$ is the expected equity value conditional on some shock realising. $i$'s equity value before the trade is

$$E[\pi_i] = (1 - r)(R - v) + r\pi_{i, shock},$$

where again, $\pi_{i, shock}$ is the expected equity value conditional on some shock realising. For $r$ sufficiently small, we have $E[\pi'_i] < E[\pi_i]$, and so any asymmetric trade reduces $i$'s equity value.

Therefore under full liability, bank $i$ will seek a trade among the set of feasible bilateral trades with $j$ that increases:

$$E[v_i - v] = E\left[\sum_{l \in N} A_{il}(p_l - \beta I_{v_l < v})\right] - v = R - r\frac{q \epsilon L + (1 - q) \epsilon S}{n} - \beta \sum_{l \in N} A_{il}E[I_{v_l < v}] - v. \quad (32)$$

We will show that $i$ and $j$ cannot raise both $E[v_i - v]$ and $E[v_j - v]$ through a feasible bilateral trade. By the symmetry of the efficient network and the symmetry of any trade that increases equity value for both counterparties, a pairwise bilateral trade between $i$ and $j$ must either weakly increase both their expected default probabilities or else decrease them both.

If the trade weakly increases both $i$ and $j$'s default probabilities, all other banks’ default probabilities weakly increase. Thus $E[v_i - v]$ weakly falls by equation (32), so that the trade is unprofitable. If the trade lowers both default probabilities, then all other banks’ default probabilities must also weakly
fall. If all banks’ default probabilities fall after the trade, the initial network cannot have been socially efficient—which is a contradiction. Thus the efficient network is stable under full liability.

**Part (ii):** The result follows immediately from the proof of part (i). If a pairwise bilateral trade strictly decreases both $i$ and $j$’s default probabilities, it also weakly decreases all other banks’ default probabilities, contradicting the optimality of the efficient network.

\[\square\]

### C Payment Equilibria

**Proposition 8.** For any financial network and vector of returns, at least one payment equilibrium always exists, and the set of equilibria forms a complete lattice.

**Proof.** We need to show that there is always a vector of market value flowing to shareholders, $v$ satisfying

$$v = A(p - b(v))$$

where $p$ is any permissible return vector, therefore constituting a fixed point of the equation system. Let $L = \prod_{i=1}^{n} [0, \sum_{j \in N} A_{ij} p_j]$. Hence any possible vector of shareholder market value is an element in $L$.

Suppose that elements in $L$ are element-wise ordered, i.e. for any two vectors $v_1, v_2 \in L$:

$$v_1 \leq v_2 \text{ if and only if } v_{i1} \leq v_{i2} \text{ for all } i = \{1, \ldots, n\}$$

We can easily show that $L$ must be a complete lattice. Define any subset $S \subset L$ by $S = \{v_p\}_{p=1}^{q}$, for $v_p \in L$. $S$ has an infimum given by:

$$\inf S = v_1 \land \ldots \land v_q,$$

and a supremum given by:

$$\sup S = v_1 \lor \ldots \lor v_q,$$

which are both elements in $L$. Hence $L$ is a complete lattice.

Define the mapping $\Phi : L \to L$ by:

$$\Phi(v) = A(p - b(v)).$$

Then $\Phi(v)$ is an order-preserving function. This is because failures are complements, so that:

$$v_1 \leq v_2 \implies \beta_i I_{v^i_1 < v^i} \geq \beta_i I_{v^i_2 < v^i} \implies b(v_1) \geq b(v_2) \implies \Phi(v_1) \leq \Phi(v_2),$$

as required.
Thus $L$ is a complete lattice, and $\Phi : L \to L$ is an order-preserving function. Therefore by the Knaster-Tarski fixed point theorem, $\Phi$ always has a fixed point in $L$, and the set of fixed points is a complete lattice.