Endogenous Financial Networks: Efficient Modularity and Why Shareholders Prevent It

Matthew Elliott* and Jonathon Hazell†

August 2015

Abstract

We model the implications for systemic risk in financial networks of the classic conflict of interest between debt-holders and equity-holders (Merton, 1974). Financial connections help diversify banks' idiosyncratic risk and avoid failures following small shocks, but increase the risk of contagion following large shocks. A social planner resolves this trade-off by creating a modular network structure with *fire breaks*, thereby diversifying away banks' exposures to small shocks while containing contagion. These socially efficient networks suit debt-holders' interests, but run counter to equity-holders' interests. They transfer surplus value from the equity-holders of healthy banks to the debt-holders of distressed banks. Equity-holders can profitably trade away from these networks, and so socially optimal networks are unstable. Moreover, trades are profitable for equity holders when they align counter-parties' failures with their own, creating systemic risk.

*Cambridge and California Institute of Technology

†MIT. Email: jhazell@mit.edu. Funding from Caltech's Summer Research Fellowship programme is gratefully acknowledged.

1 Introduction

During the recent financial crisis, concerns over the systemic risk generated by financial interconnections underpinned policy interventions[1]. Therefore, to design effective regulation we need a clear understanding of precisely how linkages between financial institutions contribute to systemic risk. Our comprehension of these issues has advanced rapidly. We now better understand which network structures are conducive to the spread of financial contagion, the mechanisms through which financial interconnections amplify real-economy shocks, and how to identify systemically important financial institutions. Nevertheless, our understanding of some fundamental questions is still very limited. For example, what network of interdependencies would a benevolent social planner choose? What networks should we expect shareholder value maximizing banks to generate in a decentralized system? Would these differ from socially optimal networks—and if so, why? These questions are crucial for informed contemplation of regulation, and particularly for trying to anticipate the response of market participants.

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

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

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

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

[1] See, for example, Plosser (2009).

We first consider socially efficient networks. We introduce a key trade-off in the social planner's problem, concerning small versus large shocks to asset returns. On the one hand, strong interconnections across many counterparties can prevent bankruptcies after relatively small shocks. However, they also render the system fragile to large shocks, by transmitting the shock across multiple counterparties and potentially leading to financial contagion.[2] In our first key finding, the social planner's optimal response is to partition the financial system into "clusters" of banks, with strong dependencies within each cluster, and much lower dependencies across clusters. Small shocks are absorbed without triggering bankruptcies, and large shocks only cause single clusters to fail. Intuitively, the socially efficient structures create *firebreaks*, which stop the spread of financial contagion after large shocks, but still prevent failures after small shocks. Firebreaks remain important when we allow for heterogeneous bank size. Networks somewhat resembling real-world networks, such as core-periphery structures, may be socially optimal. However these structures still require firebreaks *between* core banks—which seems unlikely to hold in real-world financial systems.

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

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

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

We next endogenise interbank relations, to consider whether individually optimising banks might arrange their contracts in a socially efficient pattern. In particular, we assume that banks trade on a bilateral basis, and aim to maximise shareholder value. Our main finding is that socially efficient networks are typically not privately stable, unless bankruptcy costs are very large. This divergence between private and social incentives is driven by the interaction between limited liability, and the absence of debtor discipline—both empirically relevant assumptions. These two assumptions create a conflict of interest between equity and debt holders, which generates a wedge between private

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

and social optimality. The social planner maximises the sum of shareholder and debt-holder value, which entails minimizing the expected number of defaults. To prevent defaults, the social planner redistributes surplus from the shareholders of healthy organizations to the debt holders of distressed organizations. Consequently the social planner's interests align with debt holders and against equity holders. To avoid this redistribution, shareholders seek to trade away from socially efficient networks. Though such trades may raise the probability of default, shareholders are protected by limited liability, meaning default is not costly to shareholders at the margin—even if the overall probability of financial contagion rises.

While the conflict of interest between equity- and debt-holders is well understood in a single-bank context (Merton, 1974; Rochet, 1992), we are unaware of other work demonstrating the consequences for systemic risk. Furthermore, our main result potentially resolves the critique of Cochrane (2014). In particular, limited liability means banks *prefer* trades which generate systemic risk, instead of idiosyncratic risk. In the states in which a bank fails at the same time as its counterparties, shareholders are protected from counterparties' bankruptcy costs, by limited liability. Instead, in these states bankruptcy costs are passed onto debt holders. Therefore, all else equal, shareholders prefer to fail in the same states as their counterparties. Thus shareholders prefer systemic risk—in the sense of preferring trades that allow banks to fail in the same states as their counterparties.

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 emphasises that interconnections can facilitate the spread of contagion. Networks can generate systemic risk by facilitating the spread of relatively large shocks (Gai and Kapadia, 2010), or by interacting with various propagation mechanisms, such as bankruptcy costs (Elliott, Golub, and Jackson, 2014); uncertainty about banks' balance sheets (Caballero and Simsek, 2013; Alvarez and Barlevy, 2014); and fire sales more generally (Cifuentes, Ferrucci, and Shin, 2005). Since these papers consider exogenous networks, they do not examine socially efficient network structures, nor whether individual banks might choose to deviate from the social optimum.

A smaller literature focuses on socially efficient network structures and the endogenous formation of financial networks, in the presence of systemic risk. Two particularly relevant papers[3] are Farboodi (2014) and Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015b). Each of these models endogenous network formation, and so examines why privately and socially optimal behaviour might differ. Acemoglu et al. (2015b) focusses on financial intermediation as the reason for network formation. Privately and socially efficient behaviour 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 Farboodi (2014), banks again 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. In equilibrium a core-periphery network can arise, whereby only banks with access to investment opportunities form the core. In this equilibrium private

[3] Other papers considering similar problems include Leitner (2005); Blume et al. (2011), Allen, Babus, and Carletti (2012); Babus (2013); Zawadowski (2013); Di Maggio and Tahbaz-Salehi (2014); Erol and Vohra (2014); and Cohen-Cole et al. (2015).

behaviour is socially inefficient because core banks behave in an excessively risky manner to capture intermediation rents. In our model, systemic risk also arises endogenously but from a different set of frictions, pertaining to shareholders' control rights and their limited liability.

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

Like us, Cabrales et al. (2014) study socially efficient networks when shocks to asset values can be both relatively small or relatively large. In their model systemic risk arises from direct claims on other banks' projects. By considering some key benchmark networks, Cabrales et al. (2014) explore the social planner's preferences over network segmentation and link density. Dense interconnections prevent failures from small shocks, but facilitate the spread of contagion from large shocks. Building on their work we are able to solve the social planner's problem without restricting attention to benchmark networks. We also include bankruptcy costs, which provide the key propagation mechanism in our setting.[4, 5]

The role of limited liability in generating financial instability has been widely discussed in the context of a single-bank framework (e.g. Merton, 1974; Rochet, 1992 and Gollier et al., 1997). However, to the best of our knowledge we are the first to examine the effect of limited liability in a *network* setting.

We structure the paper as follows. Section 2 presents our model. Section 3 examines socially efficient networks. Section 4 discusses equilibrium networks. Section 5 concludes. All proofs are contained in Appendix A.

2 The Model

2.1 Financial Institutions, Real Assets, Revenue, and Linkages

There is a set $N = \{1, \ldots, n\}$ of financial institutions, which we refer to as banks. Each bank holds a proprietary asset. This yields a return p_i for bank i, derived from i's investments in the real economy. Each bank can also hold claims on the market value of other institutions, and we denote these claims on market values as *cross-holdings* . We let $C_{ij} \geq 0$ be the *cross-holding* of bank i in bank j.[6] We can

[4] Cabrales, Gale, and Gottardi (2015) explains this modelling difference in detail vis-a-vis Elliott et al. (2014), which provides the foundation of our model.

[5] In another more technical difference, again explained in detail by Cabrales et al. (2015), we model interbank contracts somewhat differently. In our model, banks hold claims on counterparties' balance sheets, as opposed to claims on underlying assets.

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

represent $\mathbf{C}$ as a weighted, directed graph, where the banks are the nodes and the links represent their claims on each other. Equivalently, $\mathbf{C}$ is a *financial network*. After bank i has paid off all the claims on its own market value, there is a share $\widehat{C}_{ii} := 1 - \sum_{j \in N} C_{ji}$ left, assumed to be strictly positive[7]—it is the proportion of bank i's market value flowing to the outside shareholders of bank i. Setting $\widehat{C}_{ij} = 0$ for all $j \neq i$, $\widehat{\mathbf{C}}$ is a diagonal matrix.

It is worth briefly discussing the nature of these market value claims, which follow the model set out by Elliott, Golub, and Jackson (2014). In general, a financial system can be thought of as a set of claims between the firms in the system, where these claims come from a variety of sources including debt and derivative contracts. We use a linear approximation to model these claims. The linear dependencies can be motivated directly by the rationing of credit to debtors following default, or as orderly write downs. Although specific contracts between firms are often far more complex, the simple linear approximation captures a key fact, that the value of these contracts is likely to decrease with an organization's value, especially in a distressed system.[8]

The market value V_i of bank i is given by:

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

In matrix notation, this is equivalent to

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

The market value flowing to the shareholders of bank i, denoted v_i, is given by $\widehat{C}_{ii} V_i$ so that we have[9]:

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

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

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

Reassuringly, $\mathbf{A}$ is column stochastic[10], so that

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

Thus the returns from every real asset are ultimately allocated to the outside shareholders of some bank. Thus banks can effectively construct portfolios of the underlying, primitive assets by holding

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

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

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

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

claims on each others' market values—since banks all own and administer proprietary projects, claims on underlying assets cannot be held directly. Equally, the dependency matrix formulation reflects the fact that while banks negotiate contracts with one another, ultimately these contracts derive their value from returns to assets in the real economy.

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

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

This representation allows an intuitive interpretation of the relationship between the cross-holding and dependency matrices. In particular, the total claim of bank i on asset j comes from the sum of indirect claims on assets via claims on banking counterparties—who also have a series of indirect claims on assets via their own holdings on other counterparties, leading to an infinite sequence of indirect claims. Consequently, there can be a large difference between the direct exposure of bank j to bank i, given by the cross-holding C_{ij}; and bank j's ultimate exposure to bank i's asset, given by dependency matrix entry A_{ij}. In general there may be no cross-holding matrices $\mathbf{C}$ that generate a given dependency matrix $\mathbf{A}$, even when that dependency matrix is column stochastic and non-negative.[11]

2.2 External Liabilities, Bankruptcy and Debt Seniority

So far we have closely followed the model of Elliott et al. (2014). We now extend their model to include banks' external liabilities, and also their limited liability. These considerations enable us to investigate the implications of equity holders' limited liability and lack of debtor discipline, on systemic risk. It is helpful to make specific assumptions about the seniority of different claims. We choose assumptions that are hopefully simple and fairly natural, although we would expect to obtain similar results more generally.

We attempt to capture the following general ideas about debt restructuring. First, as bank i's market value falls, the value of other interbank claims on bank i also fall. This formulation is reasonably general. It could, for example, capture the market value of formal contracts that constitute interbank liabilities. Equally, banks might be willing to accept write downs on their claims to avoid the costs associated with outright bankruptcy—which is of particular interest in a distressed financial system.

Secondly, while the value of interbank claims on bank i declines as i's market value falls, the value of bank i's external liabilities do not vary continuously with market value. This might reflect the different nature of these contracts. Alternatively, bank i's external creditors may be too numerous or diverse for easy renegotiation as market value changes—if, for example, the external liabilities are bank i's wage or pension obligations. However, for sufficiently large falls in market value, external liabilities can no longer be settled in full, and so bankruptcy must occur. If a bank formally enters bankruptcy proceedings, we assume all creditors are of equal seniority. Both interbank creditors and external

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

creditors receive linearly rationed repayments. Moreover after bankruptcy proceedings, repayments are lowered by the failure costs associated with disorderly default. Overall, interbank payments are effectively junior to external liabilities if a bank does not enter bankruptcy proceedings, while in the case of bankruptcy they are of equal seniority. Finally, after bankruptcy, outside shareholders receive nothing. They are residual claimants on market value, so that their contracts are junior to the debt claims of the creditors.

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

Formally, bank i's market value is:

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

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

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

and for a single bank:

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

It is worth briefly elaborating on the nature of bankruptcy costs. These are the costs of liquidating the proprietary asset, possibly at a discount during periods of financial turmoil[12]; of inefficient allocation of resources during the bankruptcy period; the cost of bankruptcy negotiations, administration and settlement; loss of human capital after workers are fired and so forth. Default costs are typically large and variable relative to firms' market values, as in James (1991) and Davydenko et al. (2012). Bankruptcy costs are the mechanism by which financial contagion becomes socially costly—otherwise bank failures simply represent a transfer of funds between different agents in the system. Thus our model reflects the prevailing consensus on financial crises since Bernanke (1983), that they are socially costly because of the institutional disruption they cause, with consequent effects on the real economy.

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

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

equations capturing the claims of all parties:

$$\mathbf{v} = \mathbf{A}(\mathbf{p} - \mathbf{b}(\mathbf{v})) \tag{2}$$

$$\boldsymbol{\pi} = \max\{\mathbf{v} - \underline{\mathbf{v}}, \mathbf{0}\} \tag{3}$$

$$\boldsymbol{\delta} = \min\{\underline{\mathbf{v}}, \mathbf{v}\}, \tag{4}$$

where $\mathbf{0}$ is the zero vector and the max and min operators are applied entry by entry.[13]

Consequently, there is a settlement risk associated with the possibility of bank j defaulting, and thus transferring less funds to other banks, compounded by the discontinuous bankruptcy cost β_i. The dependency matrix determines how the market value generated by underlying assets flows to the owners of different banks. Bankruptcy costs are directly subtracted from these market values. Therefore bankruptcy costs end up being distributed according to the dependency matrix.[14]

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

2.3 Payment Equilibrium Existence and Multiplicity

A vector of shareholder market values $\mathbf{v}$ which solves equation 2 is a *payment equilibrium*—that is, a vector of shareholder market values which simultaneously satisfies all interbank payments. We require a payment equilibrium to always exist, so that there is always a fixed point of the system described by equation 2, for any return vector $\mathbf{p}$. As has been shown in several related settings (e.g. Eisenberg and Noe, 2001; Acemoglu et al., 2015a) at least one payment equilibrium always exists. It follows from Elliott et al. (2014) that the set of payment equilibria in our model forms a complete lattice. Thus there is always a set of mutually consistent interbank payments solving equation 2, and there may be several.

The reason for the possible multiplicity of payment equilibria is the discontinuous default costs, which has two separate effects. Firstly, for a given bank i, there may be two separate values of shareholder market value v_i satisfying equation 2—one where the bank has defaulted, so that $\pi_i = 0$ and $I_{v_i<\underline{v}_i} = 1$; and another where the bank has not failed, meaning $\pi_i \geq 0$ and $I_{v_i<\underline{v}_i} = 0$. This is essentially a form of the classic expectations-driven bank run, as in Diamond and Dybvig (1983). Secondly banks' market values are interdependent. Therefore, for any two banks i and j, there may be two shareholder market value vectors satisfying equation 2—one in which both i and j fail, and the other in which neither fail. Intuitively, two banks may require high levels of debt repayment from one another to remain solvent, meaning that it is equally possible that they both default. Again, then, we can

[13] For simplicity our formulation assumes that bank i's proportional claims on bank j are the same before and after j defaults. While i's claims on j relative to another bank k's should remain constant, more or less could flow to the outside debt holders of j post default than flowed to the equity holders of j before default. Relaxing this assumption would result in a straightforward change in the dependency matrix before and after default, which would complicate notation without yielding additional insights.

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

have multiple payment equilibria due to self-fulling expectations. Throughout, following Elliott et al. (2014), we focus on the best case equilibrium with the fewest failures. In any other equilibrium, all the organizations that fail in the best case equilibrium still fail, as well as some additional organizations. We could instead focus on the worst case equilibrium without appreciably altering our results.

2.4 Financial Contagion

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

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

Financial contagion can be driven by two complementary sources. The first is from correlated exposure to underlying asset holdings. Multiple banks may be simultaneously exposed to a large fall in underlying asset values—that is, from a large fall in value of some asset p_j. This interdependency comes from banks' indirect claims on underlying assets, via cross-holdings on counterparties' balance sheets. Second, falls in asset values can then be amplified by bankruptcy costs, further depressing the value of interbank claims by some bankruptcy cost β_j. Thus bankruptcy costs serve as a propagation mechanism internal to the banking system. It is straightforward to extend our model to include correlated holdings of underlying assets, as in Elliott et al. (2014). In one interpretation of our model, we implicitly consider a system that is already distressed—since the value of interbank claims on any bank falls as the bank's market value fall. This feature could, for example, reflect an unmodelled shock to some commonly held asset.

3 Socially Efficient Networks

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

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

[15] We conjecture that this fairness constraint does not bind in the social planner's problem and so is redundant. It does

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

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

3.1 Simplification and Shocks

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

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

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

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

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

Defining $\varepsilon^* := n(R-\underline{v})$, we let shocks $\varepsilon \in [R-\underline{v}, \varepsilon^*]$ be *small* and shocks $\varepsilon > \varepsilon^*$ be *large*. Thus, absent any cross-holdings, even small shocks are enough to cause an organization to fail. The threshold ε^* implies that at least one organization must fail following a large shock; although it is possible for no organizations to fail following a small shock. Intuitively, $n(R-\underline{v})$ is the total surplus equity value available in the system. When the shock is greater than total equity value, some defaults must occur for the system to clear and a payment equilibrium to form—conversely, for small shocks surplus equity value can always be distributed in such a way that defaults are prevented. One such network structure preventing defaults from small shocks, is when organizations arrange their interconnections with one another so that their portfolios approach the market portfolio. By contrast, if all shocks are large the empty network is socially efficient, because only a single bank fails, namely the bank whose asset is hit by the shock. Consequently, a trade-off is apparent. More interconnections will tend to insure the system against failures from small shocks, but may also cause financial contagion by exposing more banks to large shocks, meaning potentially more failures.[16]

considerably simplify the analysis.

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

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

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

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

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

The reason for this choice will be explained in detail in the following section.

3.2 Socially Efficient Networks

We next characterise the optimal network, conditional on no banks failing following a small shock. Observe that no banks fail following any small shock if and only if for all $i, j \in N$:

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

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

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

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

For the remainder of this section we abstract from the integer problems. We assume that (i) d^* is an integer; and (ii) n/d^* is an integer. Ignoring integer problems, we can partition the organizations into n/d^* groups of d^* agents. Define $G(i)$ as i's group. Let $\mathcal{A}$ be the set of feasible dependency matrices, which are non-negative and column stochastic. We then define the set of networks:

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

Thus $\mathcal{A}^*$ characterizes symmetric networks where organizations are partitioned into groups of size d^*, which have strong connections within groups, and weak connections between groups. For reasons that will become clear, we refer to the across-group links as *firebreak* links, and refer to the value of these links as the *strength* of the firebreaks. Following a large shock to the asset of a given bank, all banks within the same group fail but no banks outside the group fail. In addition, all banks perfectly insure themselves against all small shocks. In general, we say that for networks in which banks can be partitioned into groups with weak links between groups, and strong links within groups, these groups are called *clusters*. Thus the networks in $\mathcal{A}$ exhibit clusters. The network representation of dependencies for an $\mathbf{A} \in \mathcal{A}^*(6)$ is shown in Figure 1.

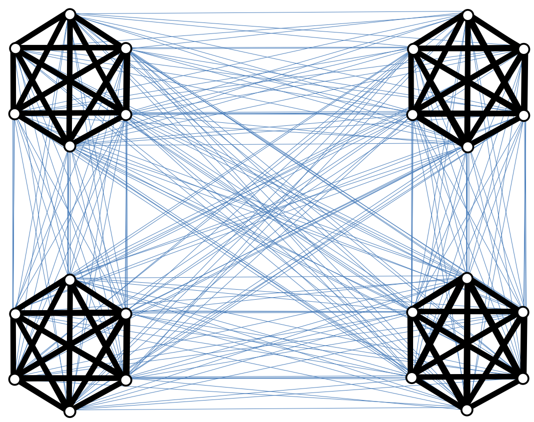

Figure 1: A network with 4 clusters and 6 banks in each cluster.

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

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

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

Hence $\widehat{\mathbf{C}} + \mathbf{C}$ is column stochastic as required.

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

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

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

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

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

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

$$0 < A^*_{ij} = \frac{R - \underline{v}}{\varepsilon_L + \beta d^*} < \frac{R - \underline{v}}{\varepsilon_S} = A^*_{jk} \quad \text{for } G(i) \neq G(j) = G(k).$$

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

It is suboptimal for clusters to be fully segmented, such that no bank holds claims to an asset from another cluster. While full segmentation would prevent any part of the large shock from being transmitted outside the cluster, it would also prevent any impact of the small shock from being diversified away from the cluster. To prevent failures following small shocks, clusters would then have to be larger, and more banks would fail following a large shock. Cabrales et al. (2014) compare the efficiency of various fully segmented network structures. Although we establish that fully segmented networks are not socially efficient, our clusters do share some important characteristics with these structures.[17]

Several observations sharpen the intuition behind proposition 1.

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

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

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

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

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

3.3 Adding Heterogeneity

So far we have only considered homogeneous banks, which are ex-ante identical. All banks are assumed to have claims on returns R, to have the same bankruptcy costs and the same failure threshold.

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

In this section we will demonstrate how networks in $\mathcal{A}^*$ can be extended to incorporate heterogeneous bank sizes, and show that they remain socially optimal.

DEFINITION **1.** Consider a financial system $(\mathbf{A}, N)$ and set of banks $M \subseteq N$, to be merged into a new bank i. The financial system $(\mathbf{A}', N')$ results from a (M, i)-merger of $(\mathbf{A}, N)$ if:

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

Therefore a bank i created from the merger of banks M inherits the cumulative claims of the banks in M. The merger preserves the overall claims of other banks and of outside shareholders. Overall bankruptcy costs, asset values and failure thresholds are also preserved. Finally, we preserve the same overall shock probability. As before, a random, firm specific shock ε hits the underlying asset of one bank. However, the probability of a given asset being hit by a shock is now proportional to the *size* of bank. We index the size of a bank by the value of its asset in the absence of a shock, given by the value of the assets belonging to the previously merged banks. Bank i is of size θ_i if its asset pays a return of $\theta_i R$ in the absence of any shocks. Hence the probability of a small shock hitting bank i is $(1-q)\left(\frac{\theta_i}{\sum_{j \in N} \theta_j}\right)$ and the probability of a large shock hitting bank i is $q\left(\frac{\theta_i}{\sum_{j \in N} \theta_j}\right)$. By construction $\mathbf{A}'$ is column stochastic, so is still a valid dependency matrix. For ease of exposition we state mergers in terms of the dependency matrix $\mathbf{A}$. These dependencies emerge naturally from merging cross-holding claims (see Appendix B).

We will consider sequences of mergers between organizations that are in the same cluster. We assume the social planner continues to take the size of organizations as given and chooses a financial network to maximise the expected sum of payments to creditors and organisations, $\mathbb{E}\left[\sum_{i \in N} \pi_i + \delta_i\right]$.

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

We can then extend our analysis of the social planner's problem to include size heterogeneity. Suppose homogeneous banks are connected by a socially optimal financial network, $\mathbf{A}^* \in \mathcal{A}^*$. 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 **2.** Consider a socially optimal financial network $\mathbf{A}^*$ connecting homogeneous banks. For any sequence of within-cluster mergers the resulting network $\mathbf{A}'$ is still socially optimal.

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

Figure 2 illustrates a socially optimal network in which there are three clusters, each with five banks. Within each cluster there is one bank that is as large as all other banks within the cluster combined. We term the relatively large organizations as *core* banks and other organizations as *periphery* banks. Each core banks could have been created by the merger of 4 periphery banks. The width of the lines represents the strength of dependencies. With the heterogeneous sizes the strongest links end up being across cluster, between the core banks—these dependencies are equal to 16 weak relationships between periphery banks from different clusters. If we focus attention on only sufficiently strong links, then the socially optimal networks associated with significant size heterogeneity closely resemble the core-periphery networks often observed in practice. However, an important property of the socially efficient networks is that there is no contagion between clusters—ensured by the relatively 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.

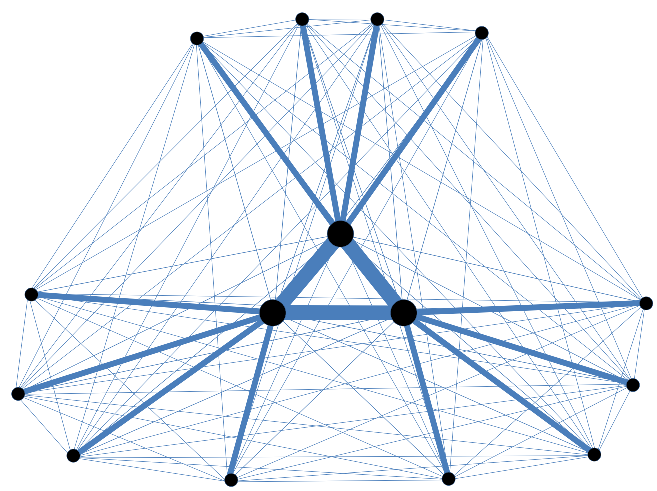

Figure 2: A socially optimal network with size heterogeneity.

In section 4 we consider whether socially optimal networks can arise in equilibrium.

3.4 Shareholders vs. Creditors

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

PROPOSITION **3.** 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 $A^* \in \mathcal{A}'$ minimises the sum of expected payments to shareholders in the class of networks $\mathcal{A}'$

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

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

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

4 Equilibrium Networks

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

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

by trading direct cross-holdings with other banks[18], where $\theta \in \Theta$ is a state of the world, and $P(\theta)$ is the probability of that state occurring.

4.1 Trading Process and Stable Networks

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

DEFINITION **2.** Crossholdings $\mathbf{C}' \in \mathcal{C}$ can be reached from $\mathbf{C}$ through a *feasible bilateral trade* between i and j if for all banks $k \in N$:

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

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

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

DEFINITION **3.** A private trading network $\mathbf{C}$ is *stable* if and only if for every pair of banks $i, j \in N$, there are no feasible pairwise trades between i and j such that:

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

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

4.2 Socially Efficient Networks and Private Instability

We now present our main result.

THEOREM **1.** There exists a $\underline{\beta} > 0$ such that for all $\beta \in (0, \underline{\beta})$ no socially efficient network with at least two clusters and at most $N/2$ clusters is stable.

Theorem 1 shows that socially efficient networks are not stable for sufficiently small β. Even with a highly restrictive set of trades—that is, only symmetric trades in direct crossholdings between a pair of banks—we show that profitable trades away from socially efficient networks can be constructed.

Before examining the intuition behind this instability result, we demonstrate that it is caused by two separate assumptions, namely shareholders' limited liability, and the absence of debtor discipline.

PROPOSITION **4.**

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

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

Under full liability, the effects of limited liability are removed. Under default probability minimisation, a bank maximises the probability that external creditors receive the face value of their debt. Therefore in this case banks act in the interests of debtholders, so that debtor discipline effects determine the stability of financial networks. Consequently, restoring *either* full liability or debtor discipline is sufficient to render socially efficient networks stable—conversely, both limited liability and a lack of debtor discipline are necessary to induce a wedge between private and social optimality.

To provide intuition for the theorem and the subsequent proposition, we consider equation 7, replicated below for convenience.

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

The equation demonstrates the key importance of limited liability and the absence of debtor discipline, in creating a divergence between private and social incentives and generating greater systemic risk. Bank i's claims on some bank j's underlying assets, given by A_{ij}, have no value to i's shareholders in states where i fails, due to shareholder limited liability. Thus if i and j fail in the same states, i's shareholders are not exposed to the impact of j's bankruptcy costs. Bank i therefore prefers claims on banks who only fail in the same states as i, because of limited liability. More generally, bank i gains more equity value from claims on banks whose failures positively correlate with it's failure, all else equal. Importantly, when a set of banks fail in the same state, so that shareholders are protected by limited liability, it is the banks' external debt-holders who lose value from the resultant bankruptcy costs—due to debt write-downs. Therefore limited liability allows shareholders to redistribute value towards themselves, and away from their external debt holders. Since there is no debtor discipline in our model, and only shareholders choose trades, external debt holders are unable to prevent shareholders from redistributing value in this way. We previously noted, in Proposition 3, that socially efficient networks act to redistribute value away from equity holders and towards external debt-holders.

Limited liability allows equity holders to reverse this process. Overall, equity holders benefit from *systemic* risk, whereby banks fails at the same time as their counterparties; as opposed to idiosyncratic risk, whereby banks and their counterparties fail in different states.

In the proof of Theorem 1, banks i and j from different clusters deviate from the social optimum by executing a trade in each others' cross-holdings, such that both banks increase the value of their claims on their own underlying assets. In return, banks reduce their claims on each others' underlying assets.[19] Bank i is more heavily exposed to the banks in its own cluster, who all default in the same states as i. These bankruptcy costs do not affect i's equity value in such states, by limited liability. Conversely, bank i weakly reduces its exposure to banks in other clusters, and so is less exposed to the failure costs of these banks, in the states where i remains solvent but banks in other clusters fail. Hence i's overall equity value increases. Concomitantly, the trade lowers the expected value of bank i's external debt, as write downs become larger due to the greater exposure of i's external debt holders to counterparty shocks and bankruptcy costs in states where i fails. Furthermore the trade significantly increases the number of failures in certain states—in particular, after the trade the entire network fails after any large shock. However, i still finds the trade optimal since it already has zero equity value in all these states prior to the trade, so that the additional failures are not costly to i's shareholders at the margin due to limited liability. Thus i and j's privately optimal trade acts to generate greater systemic risk.

Socially efficient networks can become stable when bankruptcy costs are sufficiently high. The intuition underscores the importance of the absence of debtor discipline, in generating the instability result. Bank i prefers default in the states when it already has zero equity value in the socially efficient network, so as to raise its equity value in other states, while being protected by limited liability. As discussed, this trade is optimal for i even though many more failures may occur in these states after the trade. However, such a trade entails the trading partner j absorbing more bankruptcy costs in such states—since in general j and i do not fail in the same states. For relatively small bankruptcy costs, j can absorb these defaults and still raise overall equity value after the trade. However, for large enough bankruptcy costs, the cost to j will lower j's overall equity value. Thus j will prevent i from making the trade—and vice versa. Therefore for sufficiently large bankruptcy costs, i and j will prohibit each other from making trades that excessively raise default probabilities, since these trades diminish the value of i and j's interbank contracts by a correspondingly large amount. Thus large bankruptcy costs effectively restore debtor discipline—since creditors *internal* to the financial network prevent one another from executing excessively risky trades. This intuition underscores that it is the absence of debtor discipline which allows our main instability result to go through, working in tandem with limited liability.

4.3 Trading and Systemic Risk

The key point from our analysis so far is that trades can be profitable and also increase systemic risk, specifically by increasing the correlation between different banks' failures. To further illustrate this

[19] We define bilateral trades to allow two banks to exchange cross-holdings in each other while preventing them from using these exchanges to directly increase their shareholder claims. Specifically, in the trade constructed to prove Theorem 1, we decrease i's holding in j, C_{ij}, and increase C_{ii} holding fixed $\widehat{C}_{ii}$. The alternative trade in which C_{ii} remains fixed and $\widehat{C}_{ii}$ increases is also profitable.

point, we relax our definition of feasible bilateral trades, to consider a much wider range of trades.

DEFINITION **4.** Crossholdings $\mathbf{C}' \in \mathcal{C}$ can be reached from $\mathbf{C}$ through a *relaxed bilateral trade* between i and j if:

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

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

DEFINITION **5.** A private trading network $\mathbf{C}$ is *stable with respect to relaxed bilateral trades* if and only if for every pair of banks $i, j \in N$, there are no relaxed bilateral trades between i and j such that:

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

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

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

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

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

which captures the marginal value that i's shareholders place in a cross-holding in j—given the indirect claims on various assets k that the cross-holding in j entails, and noting that the value of these claims depend on the correlation of i and k's bankruptcies. The next result establishes that a network cannot typically be stable if i has positive cross-holdings in a bank l ($C_{il} > 0$), j has positive cross-holdings in a bank k ($C_{jk} > 0$), and i values a cross-holding in l relatively more than a cross-holding in k, compared to j.

PROPOSITION **5.** Suppose that in all states of the world no bank finds itself exactly on its failure threshold, that is, $v_k(\theta) \neq \underline{v}$ for all k and all θ. Also, suppose that $C_{il} > 0$ and $C_{jk} > 0$ for $i \neq j \neq k \neq l$. Then the network is stable with respect to relaxed bilateral trades only if $\frac{p_{il}}{p_{ik}} \geq \frac{p_{jl}}{p_{jk}}$.

Crucially, all else equal, p_{il} is higher when i's failure is highly correlated with the failures of the banks in which l has large indirect claims, since limited liability means that these claims are more valuable to i. Therefore, if possible, i and j will swap cross-holdings in third parties so that i and j's own failures become more closely correlated with the failures of banks on whom they have large claims. Proposition 5 emphasizes that equity holders value exposure to organizations whose failures correlate with their own. Equity holders therefore value the systemic risk generated by certain trades, since it allows them to default in the same states as the counterparties to whom they are most heavily exposed.

An immediate corollary of Proposition 5 is that there exist relaxed bilateral trades in which only claims on third parties are exchanged, which are profitable deviations from socially efficient networks. If only claims on third parties are exchanged, then we have $C_{kl} = C'_{kl}$ for any pair k, l such that $k \in \{i, j\}$ and $l \in \{i, j\}$.

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

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

Underlying both Theorem 1 and Proposition 5 is that shareholders are indifferent between making no equity value while remaining in business, and defaulting—by definition, there is limited liability. Hence while bankruptcy is socially costly, it is not privately costly at the margin to shareholders, though ends up impacting external debt-holders. Thus situations exist in which a bank can raise its default probability *and* its expected equity value. While shareholders may sometimes prefer to raise a bank's probability of default, these trades are strictly worse for the bank's external creditors—who cannot act to prevent such trades, but will then face greater debt writedowns. Hence it is the interaction between limited liability and the absence of debtor discipline that causes private and social interests to diverge. Increasing the probability of default can therefore be privately optimal and socially suboptimal, and constitutes a conflict of interest between shareholders and bondholders. Moreover, this conflict creates incentives that favour systemic risk. We can therefore respond to the critique of Cochrane (2014), which hypothesises that banks would endogenously act to prevent financial contagion arising from "domino effects". In our model, banks do not "build buffers against dominoes", in Cochrane's words. In fact, quite the opposite—given limited liability, banks *prefer* to correlate their failures with one another, thus allowing domino effects to endogenously take place.

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

5 Conclusion

The recent financial crisis has emphasised the importance of understanding how the structure of financial networks might generate systemic risk. A key question is why privately optimal behaviour, in a decentralised financial system, might allow systemic risk to emerge. We need to understand the reasons behind any putative divergence between private and social incentives when considering regulations.

This paper seeks to explain the conflict between private and social interests by emphasising the interaction between limited liability, and the absence of debtor discipline. We start by deriving the social planner's solution, in an environment of both rare, large shocks; and common, small shocks. We show that, in general, socially efficient networks are characterised by *firebreaks*. That is, in the optimal network structure banks are partitioned into sets that we term clusters, with strong interdependencies within clusters and weak dependencies across clusters. This structure prevents failures after small shocks, and also minimises the spread of contagion after large shocks. Thus our result formalises the conjecture by Haldane (2010), that the architecture of socially efficient financial networks should exhibit *modularity*.

In our main result, we demonstrate that limited liability and the absence of debtor discipline will mean socially efficient networks are typically not stable. Limited liability prevents failure from being costly at the margin to equity holders. Banks can then make profitable trades that increase the correlation between their failures and those of their counterparties. While a counterparty failing is costly for the equity holders of a bank that is solvent, it is not costly for the equity holders of a bank that fails in the same states. Shareholders prefer systemic risk to idiosyncratic risk. In this sense, we show that banks have incentives to endogenously allow inefficient "domino effects" to occur.

Our analysis supports policy proposals that seek to break the confluence of shareholder control and limited liability, which is the underlying cause of excessive systemic risk in our model. According to our results, one set of useful policies involves reducing the hazardous effects of limited liability, by making default costly for banks at the margin[20]. Our findings also support policies that aim to prevent the breakdown of debtor discipline, by ensuring that some degree of control rights belong to creditors, especially in a distressed financial system. Certain forms of convertible debt (e.g. Bulow and Klemperer, 2015) might abet this outcome. Moreover, our work also helps justify policies that aim to rectify systemic risk directly, through our analysis of firebreaks. These arguments provide a justification for regulation that explicitly limits the size of bilateral exposures between different counterparties, in the interests of reducing systemic risk[21].

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

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

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

References

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

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

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

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

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

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

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

BERNANKE, B. S. (1983): "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression," *The American Economic Review*, 257–276.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

PLOSSER, C. I. (2009): "Redesigning financial system regulation," in *A speech at the New York University Conference Restoring Financial Stability: How to Repair a Failed System*, Citeseer.

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

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

A Omitted Proofs

A.1 Proof of Remark 1

The social planner's problem is to maximise:

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

Using equation 7 we have that:

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

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

$$\mathbb{E}\left[\sum_{i\in N}\pi_i+\delta_i\right] = nR-\varepsilon-\beta\mathbb{E}\left[\sum_{j\in N}I_{v_j<\underline{v}_j}\right] \tag{8}$$

$$= nR-\varepsilon-\beta\mathbb{E}[\textit{no. of defaults}] \tag{9}$$

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

A.2 Proof of Lemma 1

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

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

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

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

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

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

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

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

and so

$$\sum_{j \in D_i} \left[\sum_{k \in D_i} A_{jk}\beta + A_{ji}\varepsilon_L \right] \geq \varepsilon_L + d_i\beta - [n(R - \underline{v}) - d_i(R - \underline{v})]. \tag{11}$$

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

$$\sum_{j \in D_i} A_{ji} \leq \frac{d_i(R - \underline{v})}{\varepsilon_S}. \tag{12}$$

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

$$\sum_{j \in D_i} \sum_{k \in D_i} A_{jk} \leq \frac{d_i^2(R - \underline{v})}{\varepsilon_S}. \tag{13}$$

Combining inequalities (11), (12) and (13) we have:

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

Rearranging:

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

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

A.3 Proof of Remark 2

Let $\lim_{\phi\to 0} \widehat{\mathbf{C}}^*(\phi,\kappa)(\mathbf{I}-\mathbf{C}^*(\phi,\kappa))^{-1} = \mathbf{A}(\kappa)$. Any feasible dependency matrix $\mathbf{A}$ can generally be represented the Neumann series:

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

From diagonalising $\mathbf{C}^*$, it follows that if $G(i) = G(j)$, then $A_{ii}(\kappa) = A_{ij}(\kappa) = \frac{d^*\kappa+1}{d^*(1+n\kappa)}$ and if $G(i) \neq G(j)$ then $A_{ij}(\kappa) = \frac{\kappa}{1+n\kappa}$. Let $\kappa^* = \frac{\varepsilon_S - d^*(R-\underline{v})}{d^*(n(R-\underline{v})-\varepsilon_S)}$, which is always non-negative due to the upper bound on ε_S and the value of d^*. Then $A_{ii}(\kappa^*) = A^*_{ii}$ and $A_{ij}(\kappa^*) = A^*_{ij}$ for all $i, j \in N$. $\square$

A.4 Proof of Proposition 1

Lemma 1 identifies a lower bound on the number of failures when a large shock hits an organization i. This lower bound is given by the positive root of $f(d_i) = 0$, defined as d^*. This lower bound applies for all $i \in N$. We show now that there exists a network that simultaneously achieves this lower bound for all $i \in N$.

As no organization fails following a small shock to any other organization

$$A_{jk} \leq \frac{\sum_{l\in N} A_{jl}R - \underline{v}}{\varepsilon_S} \quad \text{for all } j,k \in D_i \text{ and for all } i \in N. \tag{14}$$

Inequalities (11), (12) and (13) are all used to construct $f(d_i)$, and all three must bind for the lower bound d^* to be achieved. Thus by equation 13

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

The only way for equations 14 and 15 to hold is for

$$A_{jk} = \frac{\sum_{l\in N} A_{jl}R - \underline{v}}{\varepsilon_S} \quad \text{for all } j,k \in D_i \text{ and for all } i \in N. \tag{16}$$

Then for any $j, k \in D_i$ we must have $k \in D_j$. Hence $D_j \supseteq D_i$. An equivalent analysis conditions for shocks hitting $j \in D_i$ leads to the conclusion that $D_i \subseteq D_j$. Combining set inclusions we conclude that $D_i = D_j$ for all $j \in D_i$. Hence to achieve the lower bound for all $i \in N$, the set of banks N must be partitioned into disjoint subsets such that when a large shock hits the asset of any bank in the set, all banks in the set default. Letting there be K such sets we denote these sets $\mathcal{S} = \{S_1, \ldots, S_K\}$.

The upper bound on losses absorbed by all banks $j \notin D_i$ after a large shock to asset i must also bind—and again, this must hold for all $i \in N$. Therefore by equation 10 we have;

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

Thus

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

and so

$$A_{jh} = \frac{\sum_{l \in N} A_{jl} R - \underline{v}}{\varepsilon_L + \beta d^*} \text{ for all } j \notin D_i, \text{ for all } h \in D_i. \tag{17}$$

As there are by assumption no integer problems, d^* is an integer and n/d^* is an integer. Thus $|S_k| = d^*$ for all $S_k \in \mathcal{S}$.

Next, we solve for the values of the dependency matrix. By equations 16 and 17, and substituting for the size of S_k, we have:

$$A_{ij} = \frac{R - \underline{v}}{\varepsilon_S} \text{ if } D_i = D_j$$

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

$$|D_i| = d^* \text{ for all } i \in N.$$

The above conditions define the set $\mathcal{A}^*$, and so a network minimises the expected number of failures from a large shock and prevents failures from a small shock, if and only if $\mathbf{A} \in \mathcal{A}^*$.

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

A.5 Proof of Remark 3

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

(i) Implicitly differentiating equation 18 with respect to $\beta, \varepsilon_L, \varepsilon_S$ and $R - \underline{v}$, respectively:

$$\begin{aligned}
\frac{\partial d^*}{\partial \beta} &= \frac{d^*(\varepsilon_S - d^*(R - \underline{v}))}{2d^*\beta(R - \underline{v}) + (R - \underline{v})(\varepsilon_L - \varepsilon_S) - \varepsilon_S \beta}, \\
\frac{\partial d^*}{\partial \varepsilon_L} &= \frac{\varepsilon_S - d^*(R - \underline{v})}{2d^*\beta(R - \underline{v}) + (R - \underline{v})(\varepsilon_L - \varepsilon_S) - \varepsilon_S \beta}, \\
\frac{\partial d^*}{\partial \varepsilon_S} &= \frac{d[(R - \underline{v}) + \beta] + \varepsilon_L - n(R - \underline{v})}{2d^*\beta(R - \underline{v}) + (R - \underline{v})(\varepsilon_L - \varepsilon_S) - \varepsilon_S \beta}, \\
\frac{\partial d^*}{\partial (R - \underline{v})} &= \frac{-[d^{*2}\beta + d^*(\varepsilon_L - \varepsilon_S) + \varepsilon_S n]}{2d^*\beta(R - \underline{v}) + (R - \underline{v})(\varepsilon_L - \varepsilon_S) - \varepsilon_S \beta}.
\end{aligned}$$

Recall that the socially efficient cluster size is given by the positive root of:

$$d^2(R-\underline{v})\beta + d[(R-\underline{v})(\varepsilon_L-\varepsilon_S)-\varepsilon_S\beta] + n\varepsilon_S(R-\underline{v}) - \varepsilon_S\varepsilon_L = 0 \tag{18}$$

denoted here by d^*. Rearranging equation 18 yields

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

and so

$$d^*(R-\underline{v}) < \varepsilon_S. \tag{19}$$

Since d^* is the unique positive root of equation 18, we have:

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

Rearranging yields

$$\begin{aligned} &2d^*\beta(R-\underline{v}) + (\varepsilon_L-\varepsilon_S)(R-\underline{v}) - \varepsilon_S\beta \\ &\qquad = \sqrt{[(R-\underline{v})(\varepsilon_L-\varepsilon_S)-\varepsilon_S\beta]^2 - 4(R-\underline{v})\beta\,[n\varepsilon_S(R-\underline{v})-\varepsilon_S\varepsilon_L]} > 0. \end{aligned} \tag{20}$$

Returning to the implicit differentiation, in all cases, the denominator is positive by equation 20. The numerator is positive in the second and third cases according to equation 19; and positive in the first case and negative in the fourth case by the definition of a large shock.

(ii) Substituting $\varepsilon_S = n(R-\underline{v})$ into equation 18 and dividing through by ε_S yields:

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

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

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

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

□

A.6 Proof of Proposition 3

1. Under any network $\mathcal{A}'$ where no banks fail under small shocks, the expected equity value for bank i conditional on a small shock is always $\mathbb{E}[\pi_i|\varepsilon_S] = R - \underline{v} - \frac{\varepsilon_S}{n}$. In any network $\mathcal{A}^*$, we have $\mathbb{E}[\pi_i|\varepsilon_L] = 0$, so that $\mathcal{A}^*$ achieves the lower bound on expected equity value conditional on a large shock, and therefore the lower bound on expected equity value for any network in $\mathcal{A}'$.

2. Conditional on no failures from a small shock, the minimised loss in value from a large shock is $\varepsilon_L + \beta d^*$, in any network $\mathcal{A}^*$. This loss in value is allocated between shareholders, who collectively absorb some amount in $[0, n(R - \underline{v})]$; and bondholders, who collectively absorb some amount in $[0, n\underline{v}]$. In any network A^*, conditional on a large shock shareholders absorb their maximum loss in value, $n(R - v)$. The total loss in value after a large shock is minimised, conditional on no failures from small shocks, by the definition of d^*. Therefore external creditors absorb the minimum loss in value after a large shock in network A^*, conditional on no failures from a small shock. This means expected payments to external creditors are maximised, conditional on no failures from a small shock. □

A.7 Proof of Remark 4

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

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

We have:

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

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

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

Since all but the last term is exogenously given, the social planner minimises the expected size of default costs. □

A.8 Proof of Proposition 2

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

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

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

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

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

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

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

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

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

A.9 Proof of Theorem 1

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

LEMMA **2.**

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

We now turn to proving the Theorem. The following notation will be useful:

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

Also, we denote the state in which a shock hits bank k's asset by θ_k, and the value taken by the shock ε in state θ by $\varepsilon(\theta)$.

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

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

$$\begin{aligned}
\frac{\partial A_{gh}}{\partial C_{ii}} - \frac{\partial A_{gh}}{\partial C_{ij}} + \frac{\partial A_{gh}}{\partial C_{jj}} - \frac{\partial A_{gh}}{\partial C_{ji}} &= \frac{A_{gi}A_{ih}}{\widehat{C}_{ii}} - \frac{A_{gi}A_{jh}}{\widehat{C}_{jj}} + \frac{A_{gj}A_{jh}}{\widehat{C}_{jj}} - \frac{A_{gj}A_{ih}}{\widehat{C}_{ii}} \\
&= \left(\frac{A_{ih}}{\widehat{C}_{ii}} - \frac{A_{jh}}{\widehat{C}_{jj}} \right) (A_{gi} - A_{gj}) \qquad (21)
\end{aligned}$$

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

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

Initializing the network to be socially efficient, as i is in a different cluster from j, $A_{ii} > A_{ij}$. Moreover, in the socially efficient network $\widehat{C}_{ii} = \widehat{C}_{jj}$ so i's claims on h strictly increase if and only if $A_{ih} > A_{jh}$. Thus the trade increases i's claims on all banks within i's cluster and reduces i's claims on all banks in j's cluster—by symmetric amounts, due to the symmetry of the trade. Bank i's claims on banks from other clusters remain fixed.

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

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

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

Suppose that i exchanges $\eta > 0$ units of C_{ij} with j for η units of C_{ji}. This scales up the marginal trade already considered. Let $\mathbb{E}[\pi_i(\eta)]$ be the expected shareholder value of i after a trade of size η. We can then consider the impact of a change in η on $\mathbb{E}[\pi_i(\eta)]$. Note that $\mathbb{E}[\pi_i(\eta)]$ is continuous in $v_i(\theta)$ for all θ; as there are no bankruptcy costs, $v_i(\theta)$ is continuous in A_{il} for all l; and finally A_{il} is a continuous in C_{ij}, C_{ji}, C_{ii} and C_{jj}. Thus $\mathbb{E}[\pi_i(\eta)]$ changes continuously in η. Therefore since the constructed trade is strictly profitable at the margin, there exists an $\eta > 0$ such that the proposed trade is strictly profitable for (equity holders of) i. By symmetry such a trade is also strictly profitable for j.

Finally, observe that $v_i(\theta)$, and so $\mathbb{E}[\pi_i(\eta)]$, is continuous in β. Thus there exists a $\overline{\beta} > 0$ such that the η trade constructed above is strictly profitable for all $\beta < \overline{\beta}$.

A.10 Proof of Lemma 2

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

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

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

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

A.11 Proof of Proposition 4

We begin by providing a Lemma, proved in the next subsection.

LEMMA **3.** Starting at any $\mathbf{A}^* \in \mathcal{A}^*$, for any bank $k \notin G(i), G(j)$; banks i and j cannot lower k's default probability via a symmetric bilateral trade, and can raise k's default probability only if at least one of i or j's default probability rises.

We now prove the proposition.

Part (i): Observe that since i and $k \in G(i)$ are in equivalent positions in the network, we must have that for $k \in G(i)$, $\mathbb{E}[I_{v_k<\underline{v}}]$ increases after a symmetric bilateral trade between i and j if and only if the default probability $\mathbb{E}[I_{v_i<\underline{v}}]$ increases after a trade. Also observe that after any symmetric bilateral trade, the network remains row stochastic, since the trade preserves a symmetric matrix $\mathbf{A}$, and the matrix is column stochastic.

Therefore under full liability, and considering only symmetric bilateral trades between i and j, bank i seeks to maximise

$$
\begin{aligned}
\mathbb{E}[v_i - \underline{v}] &= \mathbb{E}\left[\sum_{l\in N} A_{il}(p_l - \beta I_{v_l<\underline{v}})\right] - \underline{v} \\
&= R - \frac{q\epsilon_L + (1-q)\epsilon_S}{n} - \beta \sum_{l\in N} A_{il}\mathbb{E}[I_{v_l<\underline{v}}] - \underline{v}.
\end{aligned}
$$

We next show that, considering only symmetric bilateral trades starting from a socially efficient network, i and j cannot raise both $\mathbb{E}[v_i - \underline{v}]$ and $\mathbb{E}[v_j - \underline{v}]$ through a symmetric bilateral trade.

We demonstrate this fact by contradiction. Suppose, firstly, that i and j make a symmetric trade such that they raise both their default probabilities. Then according to Lemma 3, and given that the default probabilities of all $k \in G(i), G(j)$ must also increase, the default probability of all banks in the network weakly increases. Hence the trade cannot raise $\mathbb{E}[v_i - \underline{v}]$ or $\mathbb{E}[v_j - \underline{v}]$. Conversely, suppose that i and j make a trade that lowers both their default probabilities. Then according to Lemma 3, and given that the default probabilities of all $k \in G(i), G(j)$ must also decrease, the default probability of all banks in the network weakly decreases. This contradicts the definition of a socially efficient network, and so cannot hold.

Part (ii): Again, we show that a network $\mathbf{A}^*$ is stable by contradiction. First consider symmetric trades. Suppose that $\mathbf{A}^*$ is not stable. Then there exists i, j who can lower their default probabilities by trading with one another. By Lemma 3, the default probabilities of all $k \notin G(i), G(j)$ must remain the same. By the symmetry of the trade, the default probabilities of all $k \in G(i), G(j)$ must also fall. Therefore after the trade, $\sum_{k\in N} \mathbb{E}[I_{v_k<\underline{v}}]$ must fall, contradicting the definition of a socially efficient network. □

A.12 Proof of Lemma 3

From equation 21, we know that for any bank $k \notin G(i), G(j)$, the marginal change in A_{kh} after a symmetric marginal trade between i and j is zero. Clearly, this must also hold for all non-marginal trades, so A_{kh} does not change for $k \notin G(i), G(j)$ and all $h \in N$.

Next, note that in any network $\mathbf{A}^* \in \mathcal{A}^*$, k always fails regardless of i and j's default probabilities. Bank k only fails when a large shock hits its cluster, and always fails in these states regardless of i and

j's trades, since k has a holding $\frac{R-\underline{v}}{\epsilon_S}$ in these assets and:

$$R-\underline{v}-\frac{R-\underline{v}}{\epsilon_S}\epsilon_L<0.$$

The default probability of bank $k\notin G(i),G(j)$ only rises if i or j's default probability rises, since:

$$v_k(\theta_l)=R-A_{kl}\epsilon(\theta)-\beta\sum_{m\in N}A_{km}I_{v_m<\underline{v}}$$

which is weakly decreasing in i and j's default probabilities. □

A.13 Proof of Proposition 5

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

$$\begin{aligned}\frac{\partial A_{ig}}{\partial C_{ik}}-\frac{\partial A_{ig}}{\partial C_{jk}} &= \frac{A_{ii}A_{kg}}{\widehat{C}_{kk}}-\frac{A_{ij}A_{kg}}{\widehat{C}_{kk}} &&= (A_{ii}-A_{ij})\frac{A_{kg}}{\widehat{C}_{kk}}\\ \frac{\partial A_{ig}}{\partial C_{jl}}-\frac{\partial A_{ig}}{\partial C_{il}} &= \frac{A_{ij}A_{lg}}{\widehat{C}_{ll}}-\frac{A_{ii}A_{lg}}{\widehat{C}_{ll}} &&= -(A_{ii}-A_{ij})\frac{A_{lg}}{\widehat{C}_{ll}},\end{aligned}$$

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

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

By assumption the failure sets are unchanged by a marginal trade. The marginal value to i of the trade is then:

$$\sum_{\theta\in\Theta}P(\theta)\left(\sum_{g\in N}\left[(A_{ii}-A_{ij})\left(\lambda\frac{A_{kg}}{\widehat{C}_{kk}}-\frac{A_{lg}}{\widehat{C}_{ll}}\right)\right](p_g(\theta)-\beta_g I_{v_g(\theta)<\underline{v}})\right)I_{v_i(\theta)<\underline{v}}. \tag{22}$$

Simplifying equation (22), the trade is profitable to i if:

$$(A_{ii}-A_{ij})\sum_g\left(\lambda\frac{A_{kg}}{\widehat{C}_{kk}}-\frac{A_{lg}}{\widehat{C}_{ll}}\right)\nu_{ig}>0. \tag{23}$$

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

$$(A_{jj}-A_{ji})\sum_g\left(\frac{A_{lg}}{\widehat{C}_{ll}}-\lambda\frac{A_{kg}}{\widehat{C}_{kk}}\right)\nu_{jg}>0. \tag{24}$$

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

$$\begin{aligned}\frac{\sum_g A_{lg}\nu_{jg}}{\sum_g A_{kg}\nu_{jg}} &> \frac{\sum_g A_{lg}\nu_{ig}}{\sum_g A_{kg}\nu_{ig}}\\ \frac{p_{jl}}{p_{jk}} &> \frac{p_{il}}{p_{ik}}\end{aligned}$$

A.14 Proof of Corollary 1

As the proof closely mirrors that of Theorem 1 we only provide an outline here. First, it easily verified that at for a socially efficient network with at least two clusters and at most $N/2$ clusters:

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

when $i \neq k$ are in the same cluster and $j \neq l$ are in a different cluster. By the proof of Proposition 5 there then exists a strictly profitable marginal trade in which i exchanges marginal claims on l for j's marginal claims on k. Following the proof strategy in Theorem 1 this strictly profitable marginal trade can then be scaled up to generate a strictly profitable non-marginal trade. Finally, bankruptcy costs can be increased a little without making this trade unprofitable as equity values are continuous in bankruptcy costs.

B Mergers

In this section we show how mergers that combine cross-holdings generate the dependencies claim in section 3.3. As a primative, we assume that mergers change cross-holdings in the following way:

DEFINITION **6.** Consider a financial system $(\mathbf{C}, N)$ and set of banks $M \subseteq N$, to be merged into a new bank i. The financial system $(\mathbf{C}', N')$ results from a (M, i)-merger of $(\mathbf{C}, N)$ if:

- $C'_{ij} = \sum_{k\in M} C_{kj}$ and $C'_{ji} = \sum_{k\in M} C_{jk}$ for all $j \notin i$
- $\widehat{C}'_{ii} = \sum_{k\in M} \widehat{C}_{kk}$
- $C_{ii} = \sum_{k,k'\in M} C_{kk'}$
- $C'_{jk} = C_{jk}$ for $j, k \notin M$
- $\beta_i = \sum_{m\in M} \beta_m \quad \underline{v}_i = \sum_{m\in M} \underline{v}_m \quad p_i = \sum_{m\in M} p_m$
- $N' = N \setminus M \cup \{i\}$

Note that this system includes positive self cross-holdings. After the merger bank i has positive dependencies on itself. This corresponds to dependencies being preserved among the constituent parts of the created bank and allows us to capture the dependency of one division in a bank on another.

To show that a merger generates the dependencies claimed in section 3.3 we utilize the Neumann series. Before the merger:

$$\mathbf{A} = \widehat{\mathbf{C}} \sum_{p=0}^{\infty} \mathbf{C}^p \qquad \text{and after the merger} \qquad \mathbf{A}' = \widehat{\mathbf{C}}' \sum_{p=0}^{\infty} \mathbf{C}'^p.$$

Sufficient conditions for generating the $\mathbf{A}'$ in our claim is that for all integers $p \geq 0$:

(i) $\left(\widehat{\mathbf{C}}' \sum_{p=0}^{\infty} \mathbf{C}'^p\right)_{ij} = \sum_{k \in M} \left(\widehat{\mathbf{C}} \sum_{p=0}^{\infty} \mathbf{C}^p\right)_{kj}$ for all $j \in N$

(ii) $\left(\widehat{\mathbf{C}}' \sum_{p=0}^{\infty} \mathbf{C}'^p\right)_{ji} = \sum_{k \in M} \left(\widehat{\mathbf{C}} \sum_{p=0}^{\infty} \mathbf{C}^p\right)_{jk}$ for all $j \in N$

(iii) $\left(\widehat{\mathbf{C}}' \sum_{p=0}^{\infty} \mathbf{C}'^p\right)_{jk} = \left(\widehat{\mathbf{C}} \sum_{p=0}^{\infty} \mathbf{C}^p\right)_{jk}$ for $j, k \notin M$

Proceeding inductively it can be verified that these conditions hold. We omit the formal proof as it does not add any intuition. Instead, we provide some intuition. The Neumann series captures the flow of asset returns in the system. Consider a return R generated by a bank j. In the first step of the flow the outside shareholders of j receive $\widehat{C}_{jj}R$ and $C_{kj}R$ flows to k, for all $k \in N$. In the second step of the flow, k's outside shareholders receive $\widehat{C}_{kk}C_{kj}R$ and $\sum_{k \in N} C_{k'k}C_{kj}R$ flows to k' for all $k' \in N$. And so on. The Neumann series just sums these flows to outside shareholders to find the ultimate claims that the outside shareholders of a given bank have on another bank. The key property of a merger is that it does not affect any of these flows. The returns from assets flows through the system in exactly the same way as before and the merger just relabels the banks. Maintaining this relabeling generates dependencies $\mathbf{A}'$ from $\mathbf{A}$.