

# Cascades and Fluctuations in an Economy with an Endogenous Production Network

Mathieu Taschereau-Dumouchel\*

The Wharton School of the University of Pennsylvania

June 18, 2017

## Abstract

This paper proposes a simple theory of production in which the network of input-output linkages is endogenously determined by the firms' decisions to operate. Since producers benefit from having multiple suppliers, the economy features complementarities between the operating decisions of nearby firms. As a result, tightly connected clusters of producers emerge around productive firms. In addition, after a firm is hit by a severe shock, a cascade of shutdowns might spread from neighbor to neighbor as the network reorganizes itself. While well-connected firms are better able to withstand shocks, they trigger larger cascades upon shut down. The theory also predicts how the shape of the network interacts with the business cycle. As in the U.S. data, periods of low economic activity feature fewer well-connected producers and less clustering among firms. Once calibrated, the model predicts that allowing the network to reorganize itself in response to shocks leads to substantially smaller variations in aggregate output, suggesting that endogenous changes in the shape of the production network have a significant impact on the aggregation of microeconomic shocks into macroeconomic fluctuations.

**JEL Classifications:** E23, D85, C67

---

\*Finance Department, The Wharton School of the University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104 (email: mathieu@wharton.upenn.edu).

# 1 Introduction

Production in modern economies involves a complex network of specialized producers, each using inputs from suppliers and providing their own output to downstream production units. The shape of this network is constantly changing in response to shocks. Cascades of firm shutdowns are a salient example of this process. If a severe shock pushes a producer to shut down, its neighbors, now missing a useful input or a valuable customer, might also decide to stop production, thereby pushing their own neighbors to do the same, and so on. A single shock can thus trigger a cascade of shutdowns that spreads through the economy.<sup>1</sup> While these cascades have a damaging impact on many firms, they also free economic resources that can now be used by other producers to resume production. Through this creative destruction process, the firms' decisions to operate therefore change the shape of the production network. At the same time, the shape of the network plays a crucial role in determining how microeconomic shocks propagate through the network to create macroeconomic fluctuation. To properly understand the origin of these fluctuations, it is therefore important to have a joint theory of production and endogenous network formation.

This paper proposes a simple theory to understand these economic forces. In the model, a large number of firms produce differentiated goods using labor and a set of inputs from other producers. Since production requires the payment of a fixed cost, firms operate or not as a function of economic conditions. When a firm operates, it makes an additional input available to all of its customers, thereby creating new input-output links. Together these operation decisions therefore determine the shape of the production network.

Firms benefit from using a greater variety of inputs into production. This feature of the environment creates complementarities between the decisions of nearby producers to operate and has several implications for the organization of production in the efficient allocation. First, firms that have multiple suppliers or multiple customers are more likely to operate. Second, the allocation features tightly connected clusters of producers. This way, each firm benefits from the inputs provided by the cluster and provides a valuable good in return, thereby taking advantage of the complementarities created by the network. Because of this tendency for clustering, a lot of economic activity occurs in the neighborhoods around the most productive firms while the rest of the network remains mostly inactive. Through this process, the network organizes itself to magnify the impact of high-productivity firms on the economy while limiting that of their low-productivity counterparts. Third, a small change in the environment can trigger a large reorganization of the network. For

---

<sup>1</sup>Carvalho (2014) mentions the aftermath of the Great East Japan Earthquake of 2011 as an important example of this dynamics. Carvalho et al. (2014) find that firms inside the affected zone that stopped production because of the catastrophe had a significant negative impact on their customers and suppliers that were outside of the affected zone. In contrast, firms inside the affected zone that did not stop production had an insignificant impact on their customers outside the affected area. Another example of these cascades concerns the 2008 bailout of U.S. car manufacturers. At that time, the chief executive officer of Ford advocated for the bailout of its competitors, General Motors and Chrysler. The rationale was that the bankruptcy of these carmakers would push their suppliers to shutter operations which would adversely affect Ford since suppliers are shared in the industry (Carvalho, 2014).

instance, a small drop in the productivity of an influential firm can lead to the shutdown of its whole neighborhood while activity moves to another part of the network.

Two features of the environment make the efficient allocation in this economy challenging to solve for. First, since the decision to operate a unit is binary, the optimization problem has a non-convex feasible set and belongs to the class of mixed integer nonlinear problems — a family of problems that are notoriously difficult to solve (Garey and Johnson, 1990). Second, the complementarities between producers make the objective function non-concave. As a result, usual numerical algorithms are ineffective and the Karush-Kuhn-Tucker conditions are not sufficient to characterize a solution. I propose instead a novel approach that involves *reshaping* the original optimization problem, away from potential solutions, such that 1) the reshaped problem can be solved easily, and 2) the solutions to the reshaped and the original problems coincide. This approach is expected to work well if the network of potential connections between firms is densely connected. To verify its robustness, I extensively test the approach on sparsely connected networks and find that it performs very well in practice.

I calibrate the model to the United States economy. To understand how the forces of the economy operate to shape the organization of production, it is useful to compare the optimal network to a completely random one. In contrast to the random network, the optimal network features 1) more highly connected firms, 2) a higher degree of clustering, 3) a smaller average distance between firms, and 4) a positive correlation between a firm’s number of customers and its number of suppliers. The optimal allocation therefore features a more tightly connected network of producers centered around a few central and highly productive firms. The forces at work in the economy also affect the distributions of firm-level outcomes. For instance, since the optimal network organizes production to magnify the impact of high productivity firms and mitigate that of low productivity ones, the distributions of employment and output show significant positive skewness.

I use the calibrated economy to investigate how cascades of firm shutdowns, resulting from a single individual shock, propagate through the network. I find that while more highly connected firms are more resilient to shocks, they also trigger more important cascades when they shutdown. In the calibrated economy, these cascades move mostly downstream, from customer to customer. This result however depends on the elasticity of substitution between intermediate goods entering production. When inputs are poor substitutes, significant upstream cascades can also occur. These cascades are the manifestation of the network optimally reorganizing itself in the face of shocks. As a result, preventing them from occurring leads to losses in aggregate output that are, at the median, point 35% larger.

As the economy faces many local shocks at the same time, multiple overlapping cascades spread through the production network, thereby changing its shape and creating aggregate fluctuations. As a result, the shape of the network differs between recessions and expansions in several ways. In particular, recessions are periods in which, 1) fewer highly connected firms produce and 2)

the amount of clustering in the network is lower. These correlations between the shape of the production network and the business cycle are also visible in the U.S. data. Overall, expansions are periods in which production can be efficiently organized in tightly connected groups of producers centered around influential firms. In contrast, recessions are periods in which creating these tight clusters is inefficient, perhaps because a highly connected firm is facing a bad shock.

To evaluate the impact of the optimal organization of the network on the size of fluctuations, I once again compare the optimal network to the random one and find that the standard deviation of output is more than 35% larger in the random network. This finding highlights the importance of considering how the network structure of production adapts to shocks to properly understand the microeconomic origin of aggregate fluctuations.

## Literature review

A large empirical literature documents that losing a supplier is disruptive of a firm's operation. For instance, [Hendricks and Singhal \(2005\)](#) find that firms facing supply chain disturbances face an abnormal stock return of  $-40\%$ . They further find that the impact of the disturbance is long lasting. [Wagner and Bode \(2008\)](#) survey business executives in Germany who report that supply side disturbances, including supplier suddenly stopping production, were responsible for significant disturbances in production. The [Zurich Insurance Group \(2015\)](#) also conducted a global survey of executives in small and medium enterprises. Of all the respondent, 39% report that losing their main supplier would adversely affect their operation and 14% report that they would need to significantly downsize their business, require emergency support or that they would shut down.

This paper relates to a literature that studies how shocks to interconnected sectors contribute to aggregate fluctuations in exogenous networks ([Long and Plosser, 1983](#); [Horvath, 1998](#); [Dopor, 1999](#); [Atalay, 2015](#)).<sup>2</sup> [Acemoglu et al. \(2012\)](#) find that sectoral shocks can lead to large aggregate fluctuations if there is enough asymmetry in the way sectors supply to each other. [Acemoglu et al. \(2015a\)](#) further show that inter-sectoral linkages can generate tail-risks in aggregate output. This literature emphasizes the importance of the shape of the network in transmitting idiosyncratic shocks. In contrast, the current paper studies how endogenizing the shape of the network affects the size of aggregate fluctuations.

This paper also contributes to a recent literature on endogenous network formation. [Atalay et al. \(2011\)](#) consider a model in which links between firms are created through random and preferential attachment. They find that this model is better able to match the properties of the U.S. data. [Oberfield \(2013\)](#) considers an economy in which producers optimally choose one input from a randomly evolving set of suppliers, thereby creating a production network. While the networks in both papers are formed endogenously, the present work considers how the operat-

---

<sup>2</sup>See [Carvalho \(2014\)](#) for an overview of the literature on production networks. [Bigio and La'O \(2016\)](#) study how an economy's network structure of production influences its response to financial shocks.

ing status of firms shapes the production network, so that cascades of shutdowns arise and their impact on aggregate fluctuations can be studied. [Carvalho and Voigtländer \(2014\)](#) consider how inputs diffuses in a network through input adoption decisions. To do so, they build a model in which producers direct their search for new inputs to the suppliers of their current neighbors in the network. [Lim \(2016\)](#) considers how frictions in the creation and destruction of firm-to-firm linkages affect aggregate output and trade. To my knowledge, the current paper is the first to consider how the shape of the production network changes over the business cycle.

[Baqaee \(2016\)](#) also studies the impact of cascades of firm shutdowns on the macroeconomy. To do so, the paper considers an exogenous sectoral network in which the mass of firm in each sector can vary. This adjustment margin can lead to further amplification of sectoral shocks in the presence of external economies of scale. In contrast, the current paper considers an economy without external effects, so that the allocation is efficient, and in which the network describes connections between individual producers directly. Shutting down a firm therefore removes a node in the network thereby endogenously changing its shape.

Most of the literature on networks and fluctuations considers economies in which adjustments to changes in the environment are continuous rather than discrete. Like the present work, some papers consider how discrete adjustment margins can lead to cascading effects. In the finance literature, [Elliott et al. \(2014\)](#) and [Acemoglu et al. \(2015b\)](#) study cascades of failures in a financial network. When a neighboring institution defaults, the connections to its former trading partners are severed, which might trigger their default in turn.<sup>3</sup> [Bak et al. \(1993\)](#) propose a model of production and inventories with non-convex interactions between neighbors that can generate large fluctuations from idiosyncratic shocks.

The next section introduces the model and the solution approach. Section 3 highlights some of the features of the optimal allocation. Section 4 contains a quantitative exploration of the forces of the model. The last section concludes. All proofs are in the appendix.

## 2 Model and Solution Approach

There are  $n$  units of production, also referred to as firms, indexed by  $j \in \{1, \dots, n\}$ . Each of these units, produces an intermediate good that can be used in the production of a final good or as an input in the production of other intermediate goods. A final good producer uses a CES production function with elasticity of substitution  $\sigma > 1$ ,

$$Y = \left( \sum_{j=1}^n (y_j^0)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

---

<sup>3</sup>See also [Blume et al. \(2011a\)](#) and [Blume et al. \(2011b\)](#) for studies of cascading failures and network formation.

to convert intermediate goods  $\{y_j^0\}_{j=1}^n$  into aggregate output  $Y$ . A representative household consumes the final good and supplies  $L$  units of labor inelastically.

Firm  $j$  has access to a production technology that converts a vector  $x_j$  of intermediate inputs and  $l_j$  units of labor into  $y_j$  units of good  $j$  according to the production function

$$y_j = \frac{A}{\alpha^\alpha (1-\alpha)^{1-\alpha}} z_j \left( \sum_{i=1}^n x_{ij}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}\alpha} l_j^{1-\alpha}$$

where  $x_{ij}$  denotes the intermediate input from firm  $i$ ,  $0 < 1-\alpha < 1$  is the labor intensity,  $A > 0$  is aggregate productivity and  $\bar{z} > z_j \geq \underline{z} > 0$  is firm-specific total factor productivity. The elasticity of substitution  $\epsilon > 1$  determines the benefit from producing with a wider variety of inputs.

Firm  $j$  can use good  $i$  only if there exists a *connection* from firm  $i$  to firm  $j$ . The set of all connections is fully described by an  $n \times n$  matrix  $\Omega$  such that  $\Omega_{ij} = 1$  if firm  $j$  can use inputs from  $i$  and  $\Omega_{ij} = 0$  otherwise. Without loss of generality, assume that  $\sum_{i=1}^n \Omega_{ij} \geq 1$  otherwise firm  $j$  cannot produce and we can redefine the economy without it. As it describes the set of inputs that are usable by each firm,  $\Omega$  is part of the production possibility frontier of the economy.

A firm can only produce if it is *operating*, which requires the payment of a fixed cost  $f \geq 0$ , in units of labor. The vector  $\theta$  denotes the operating status of the firms, such that  $\theta_j = 1$  if firm  $j$  is operating and  $\theta_j = 0$  otherwise. This fixed cost captures overhead labor, such as managers and other non-production workers, that is necessary for production. A broader interpretation could include the cost of fixed structures or any other factor that does not scale with production.<sup>4</sup>

Together,  $\Omega$  and  $\theta$  describe the network structure of production in this economy. While the set of *possible* connections  $\Omega$  is taken as exogenous, the set of *active* connections evolves endogenously as economic conditions affect which firms operate. Figure 1a provides an example of the set of possible connections  $\Omega$  in a simple economy with six firms. Each arrow represents a connection  $\Omega_{ij} = 1$ , with the direction of the arrow showing the movement of goods between the firms. The set of active connections, in blue in Figure 1b, is selected endogenously and depends on the set of active firms, also in blue.<sup>5</sup>

---

<sup>4</sup>See, for instance, [Bresnahan and Ramey \(1994\)](#) and [Hall \(2000\)](#) for a discussion of the importance of these costs in explaining plant-level production behavior.

<sup>5</sup>While the model is kept simple for tractability, it can be extended easily to accommodate heterogeneity in input shares  $\alpha$ , elasticities of substitution  $\epsilon$  and fixed costs  $f$ . With these extensions, the model could feature endogenous links formation. A link, in this setup, would simply be a firm with an input share close to 1. The operating cost of this firm-link could be interpreted as the cost of creating a link between two standard firms. In addition, the model can use permanent shocks to  $z$  to capture permanent entry and exit of firms.

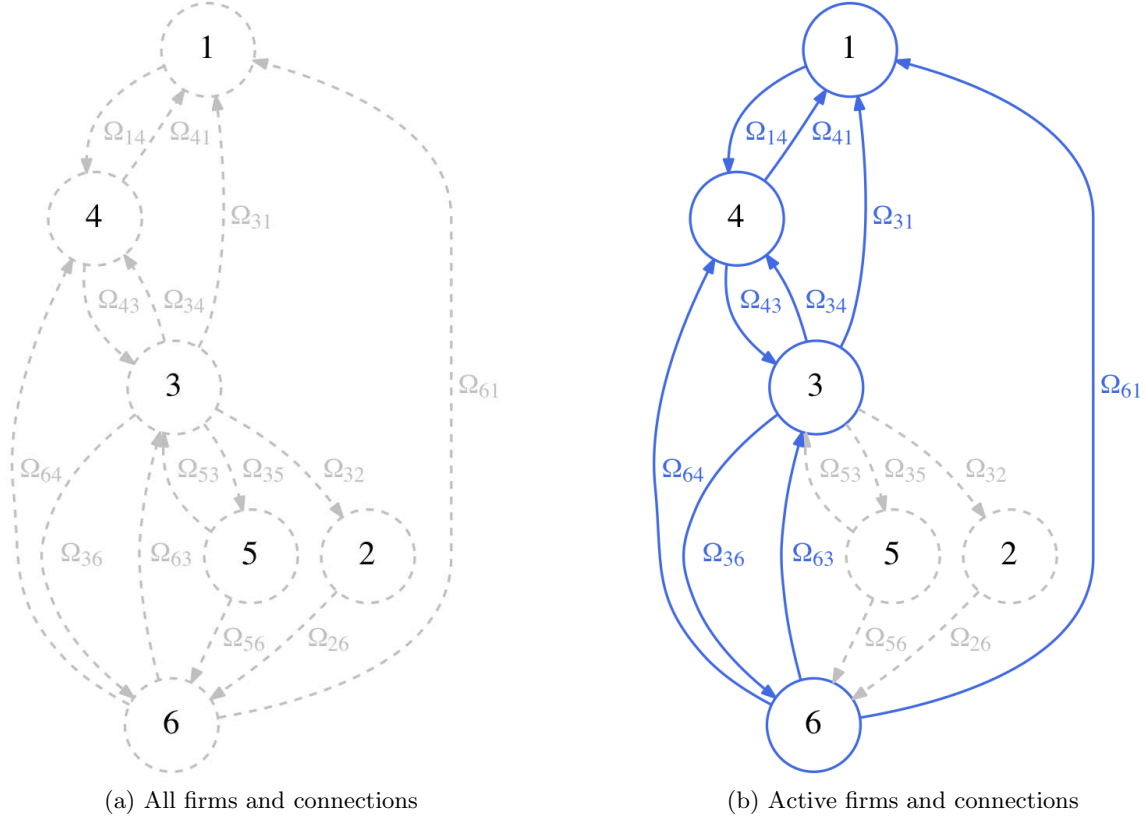


Figure 1: Operating decisions determine the shape of the network endogenously

## 2.1 The Planner's Problem

Consider the problem  $\mathcal{P}_{SP}$  of a social planner in this economy.<sup>6</sup> The planner maximizes the utility of the representative household

$$\max_{\substack{y^0 \geq 0, x \geq 0, l \geq 0 \\ \theta \in \{0,1\}^n}} \left( \sum_{j=1}^n (y_j^0)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \quad (1)$$

subject to a resource constraint for each intermediate good

$$y_j^0 + \sum_{k=1}^n x_{jk} \leq \frac{A}{\alpha^\alpha (1-\alpha)^{1-\alpha}} z_j \left( \sum_{i=1}^n x_{ij}^{\frac{\epsilon-1}{\epsilon}} \right)^{\alpha \frac{\epsilon}{\epsilon-1}} l_j^{1-\alpha}, \quad \forall j \in \{1, \dots, n\}, \quad (2)$$

a resource constraint for labor

$$\sum_{j=1}^n l_j + f \sum_{j=1}^n \theta_j \leq L, \quad (3)$$

<sup>6</sup>I focus on the efficient allocation for a few reasons. First, there is no obvious natural notion of a decentralized equilibrium in this economy. In addition, for some standard notions of equilibrium, multiple equilibria exist. [Oberfield \(2013\)](#) shows that the efficient allocation in a model of network formation can be sustained as a pair-wise stable equilibrium.

a set of operation constraints

$$\{\theta_j = 0\} \Rightarrow \{l_j = 0\}, \quad \forall j \in \{1, \dots, n\}, \quad (4)$$

such that a firm can only operate if the fixed cost  $f$  is paid, and a set of connection constraints

$$\{\Omega_{ij} = 0\} \Rightarrow \{x_{ij} = 0\}, \quad \forall i, j \in \{1, \dots, n\}^2 \quad (5)$$

so that there must be a connection from firm  $i$  to firm  $j$  for good  $i$  to be used as input by  $j$ .

It is convenient to integrate the operation constraints (4) and the connection constraints (5) directly into the goods resource constraints (2) such that they become

$$y_j^0 + \sum_{k=1}^n x_{jk} \leq \theta_j \frac{A}{\alpha^\alpha (1-\alpha)^{1-\alpha}} z_j \left( \sum_{i=1}^n \Omega_{ij} x_{ij}^{\frac{\epsilon-1}{\epsilon}} \right)^{\alpha \frac{\epsilon}{\epsilon-1}} l_j^{1-\alpha}, \quad \forall j \in \{1, \dots, n\}. \quad (6)$$

Notice the additions of the terms  $\theta_j$  and  $\Omega_{ij}$  in the production function. The term  $\theta_j$  imposes that firm  $j$  will optimally only produce goods if  $\theta_j = 1$ . Otherwise it would be using valuable resources to generate no output. Similarly, the addition of  $\Omega_{ij}$  imposes that firm  $j$  can only use goods  $x_{ij}$  if a connection  $\Omega_{ij} = 1$  exists. Notice that since  $\theta_j$  and  $\Omega_{ij}$  only take values in  $\{0, 1\}$ , these new terms have no impact on the solution of the problem.

## 2.2 Solution Approach

To isolate the difficulties that arise in solving  $\mathcal{P}_{SP}$ , it is useful to first fix the set of operating firms  $\theta$  and to solve for the optimal allocation of resources in the exogenous network.

### Planner's problem with exogenous firm status

For a given vector  $\theta$ , maximizing (1) subject to the constraints (3) and (6) is a convex maximization problem and the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize its solution. Denote by  $w$  the Lagrange multiplier on the labor resource constraint (3) and by  $\lambda_j$  the Lagrange multiplier on the resource constraint for the  $j$ th good (6). Combining the first-order conditions with the production function yields the following Lemma.<sup>7</sup>

**Proposition 1.** *The Lagrange multipliers associated with the efficient allocation satisfy*

$$q_j = z_j \theta_j A \left( \sum_{i=1}^n \Omega_{ij} q_i^{\epsilon-1} \right)^{\frac{\alpha}{\epsilon-1}}, \quad \forall j \in \{1, \dots, n\} \quad (7)$$

---

<sup>7</sup>I follow Oberfield (2013) in recasting the optimization problem in the  $q$  space.



where  $q_j = w/\lambda_j$ . Furthermore, there is a unique vector  $q$  that satisfies (7) such that  $q_j > 0$  if firm  $j$  is active and part of a strongly connected set of active firms.

Since, from the first-condition on  $l_j$ , output can be written as  $(1 - \alpha)y_j = q_j l_j$ , it is natural to interpret  $q_j$  as a measure of labor productivity.<sup>8</sup> Equation 7 highlights the importance of the production network for the labor productivity  $q_j$  of the firms. In particular,  $q_j$  increases with the number and the productivity of firm  $j$ 's suppliers. As a result, a productive firm has a positive effect on all firms that are downstream from it in the production chain.

Equation 7 can be solved by iterating on the mapping. The solution  $q$  can then be used to fully characterize the optimal allocation of resources in the economy. For instance, aggregate output  $Y$  can be computed directly from  $q$ .

**Lemma 1.** *Aggregate output in the optimal allocation is given by*

$$Y = Q \left( L - f \sum_{j=1}^n \theta_j \right) \quad (8)$$

where  $Q \equiv \left( \sum_{j=1}^n q_j^{\sigma-1} \right)^{\frac{1}{\sigma-1}}$ .

This Lemma shows that aggregate output is equal to the product of aggregate productivity  $Q$  and the amount of labor available after fixed costs have been paid.

Similarly, the vector of labor  $l$  used by each firm satisfies

$$l = (1 - \alpha) [I_n - \alpha \Gamma]^{-1} \left( \frac{q}{Q} \right)^{\circ(\sigma-1)} \left( L - f \sum_{j=1}^n \theta_j \right) \quad (9)$$

where  $\circ$  denotes the Hadamard exponent (each element of  $q$  raised to the power  $\sigma - 1$ ),  $I_n$  denotes the  $n \times n$  identity matrix and where  $\Gamma$  is an  $n \times n$  matrix with elements  $\Gamma_{jk} = \frac{\Omega_{jk} q_j^{\epsilon-1}}{\sum_{i=1}^n \Omega_{ik} q_i^{\epsilon-1}}$ .<sup>9</sup>  $\Gamma_{jk}$  captures the relative importance of firm  $j$  as a supplier to firm  $k$ . Since the Leontief inverse  $[I_n - \alpha \Gamma]^{-1}$  can be written as  $I_n + \alpha \Gamma + (\alpha \Gamma)^2 + \dots$ , we see that the planner allocates a large fraction of the labor available after fixed costs have been paid,  $L - f \sum_{j=1}^n \theta_j$ , to firms that are important suppliers in the production network.

---

<sup>8</sup>Note that since  $y_j$  is in units of goods  $j$ , this notion of labor productivity does not correspond to revenue (measured in units of the final good) divided by labor input.

<sup>9</sup>See the proof of 1 for the derivation of (9).

## Planner's problem with endogenous firm status

We now turn to the full planner's problem in which the operational status  $\theta$  of the firms is a choice variable. Combining (7) and (8) we can rewrite  $\mathcal{P}_{SP}$  as

$$\mathcal{P}_{SP}: \max_{\theta \in \{0,1\}^n} Q \left( L - f \sum_{j=1}^n \theta_j \right)$$

subject to

$$q_j = z_j \theta_j A \left( \sum_{i=1}^n \Omega_{ij} q_i^{\epsilon-1} \right)^{\frac{\alpha}{\epsilon-1}}, \quad \forall j \in \{1, \dots, n\} \quad (7)$$

When deciding whether to operate firm  $j$  or not the planner compares the cost of the decision with its benefit. On the cost side, operating  $j$  takes away  $f$  units of labor that could be used as input by other firms. On the benefit side, operating  $j$  increases  $Q$  and therefore increases the efficiency with which the production network transforms labor into final goods.

Two features of  $\mathcal{P}_{SP}$  make it hard to solve. First, each variable  $\theta_j$  is limited to the binary set  $\{0,1\}$ . Second, the objective function is not (quasi) concave. As a result, the standard Karush-Kuhn-Tucker conditions cannot be used to solve this problem. Using the standard classification of optimization problems,  $\mathcal{P}_{SP}$  belongs to the class of Mixed Integer Nonlinear Problems (MINLP). Their combinatorial nature makes these problems notoriously challenging to solve and they are, from the perspective of computational complexity theory, NP-Hard (Karp, 1972).

One way to solve  $\mathcal{P}_{SP}$  is to try each one of the  $2^n$  possible vectors  $\theta$ . For each of them,  $q$  can be found by iterating on (7) and the objective function can then be computed easily. This *exhaustive* approach is guaranteed to find the correct solution but is limited to economies with only a few firms. Indeed, for  $n$  larger than a few dozens, the huge number of possible vectors  $\theta$  becomes impossible to handle.

Instead of relying on this computationally intensive approach, I propose to solve  $\mathcal{P}_{SP}$  by first relaxing and then reshaping it in a way that makes it easy to solve. To do so, consider the following alternative *relaxed* and *reshaped* version of  $\mathcal{P}_{SP}$  which we denote by  $\mathcal{P}_{RR}$ ,

$$\mathcal{P}_{RR}: \max_{\theta \in [0,1]^n} Q \left( L - f \sum_{j=1}^n \theta_j \right)$$

where the vector  $q$  solves

$$q_j = z_j \theta_j^a A \left( \sum_{i=1}^n \theta_i^b \Omega_{ij} q_i^{\epsilon-1} \right)^{\frac{\alpha}{\epsilon-1}}. \quad (10)$$

$\mathcal{P}_{RR}$  differs from  $\mathcal{P}_{SP}$  in two important ways. First, the binarity constraint  $\theta \in \{0,1\}^n$  is now relaxed to the convex set  $\theta \in [0,1]^n$ . Second, (7) now becomes (10) and includes the *reshaping*

parameters  $a > 0$  and  $b \geq 0$ . These parameters modify the shape of the objective function, but only away from potential solutions of  $\mathcal{P}_{SP}$ . To see this, remember that any solution of  $\mathcal{P}_{SP}$  is such that  $\theta_j \in \{0, 1\}$  for all  $j$ , so that  $\theta_j^a = \theta_j$ . Similarly, for the  $b$  parameter, if  $\theta_i = 0$  then  $q_i = 0$ , and if  $\theta_i^b = 1$  then  $\theta_i^b q_i^{\epsilon-1} = q_i^{\epsilon-1}$ , such that (7) and (10), and therefore the objective functions of  $\mathcal{P}_{SP}$  and  $\mathcal{P}_{RR}$ , coincide over the set  $\theta \in \{0, 1\}^n$ .

The relaxed and reshaped problem  $\mathcal{P}_{RR}$  is only useful if it satisfies two properties. First, and unlike  $\mathcal{P}_{SP}$ , it should be easy to solve. Second, a solution to  $\mathcal{P}_{RR}$  should also be a solution to  $\mathcal{P}_{SP}$  — the problem we are actually interested in solving. It turns out that both of these properties can be satisfied when the reshaping parameters take on the values

$$a = \frac{1}{\sigma - 1} \quad \text{and} \quad b = 1 - \frac{\epsilon - 1}{\sigma - 1}. \quad (\star)$$

The following proposition establishes that  $\mathcal{P}_{RR}$  is easy to solve when  $\Omega$  belongs to a particular family of networks.

**Proposition 2.** *If the network of potential connections is such that  $\Omega_{ij} = c_i d_j$  for some vectors  $c$  and  $d$  then the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution to  $\mathcal{P}_{RR}$ .*

For more general networks, a similar result can be established as long as the network is sufficiently connected.

**Proposition 3.** *Let  $\sigma = \epsilon$  and suppose that the fixed cost  $f > 0$  and the dispersion in productivity  $\bar{z} - \underline{z} > 0$  are not too big. If the network  $\Omega$  is sufficiently connected with  $\Omega$  invertible, then the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution to  $\mathcal{P}_{RR}$ .*

Propositions 2 and 3 both show that, under certain conditions, any feasible point  $\theta^*$  that satisfies the first-order conditions and the complementary slackness conditions is a global maximum of the objective function and therefore solves  $\mathcal{P}_{RR}$ . As a result, standard numerical algorithms, such as gradient ascent, can be used to easily find the solution to  $\mathcal{P}_{RR}$ . It is also important to note that Propositions 2 and 3 only provide *sufficient* conditions for the result. In particular, section 2.3 below shows that, in practice, the relaxation and reshaping approach finds the correct solution to the planner's problem even in economies with sparse networks.

While Propositions 2 and 3 show that it is easy to solve  $\mathcal{P}_{RR}$ , the following proposition shows how to use the solution to  $\mathcal{P}_{RR}$  to solve  $\mathcal{P}_{SP}$ .

**Proposition 4.** *If  $\theta^*$  solves  $\mathcal{P}_{RR}$  and that  $\theta_j^* \in \{0, 1\}$  for all  $j$ , then  $\theta^*$  also solves  $\mathcal{P}_{SP}$ .*

The result follows directly from the fact that the feasible set of  $\mathcal{P}_{RR}$  contains the feasible set of  $\mathcal{P}_{SP}$  and that both of their objective functions coincide when  $\theta_j \in \{0, 1\}$  for all  $j$ .

Together, Propositions 2, 3 and 4 offer a convenient way to solve  $\mathcal{P}_{SP}$ . One can simply solve the relaxed and reshaped problem  $\mathcal{P}_{RR}$  and look at the solution  $\theta^*$ . If  $\theta^*$  is such that, for all  $j$ ,  $\theta_j^*$  hits either 0 or 1, then  $\theta^*$  also solves  $\mathcal{P}_{SP}$ . If not, then  $\theta^*$  does not solve  $\mathcal{P}_{SP}$ . It is therefore important for the approach to work that the solutions to  $\mathcal{P}_{RR}$  are such that  $\theta^* \in \{0, 1\}^n$ . Fortunately, this can be achieved in large connected networks when the  $(\star)$  condition is satisfied.

To see why, it is convenient to consider the first-order conditions of the equivalent problem  $\mathcal{P}'_{RR}$  in which (10) is treated as an inequality constraint

$$\mathcal{P}'_{RR}: \max_{\theta \in [0,1]^n, q} \left( \sum_{j=1}^n q_j^{\sigma-1} \right)^{\frac{1}{\sigma-1}} \left( L - f \sum_{j=1}^n \theta_j \right)$$

$$q_j \leq A z_j \theta_j^a B_j^\alpha$$

where  $B_j = \left( \sum_{i=1}^n \theta_i^b \Omega_{ij} q_i^{\epsilon-1} \right)^{\frac{1}{\epsilon-1}}$  is the composite input used by firm  $j$ .<sup>10</sup>

The first-order conditions yield

$$\frac{\partial q_k}{\partial \theta_k} \frac{\partial Q}{\partial q_k} \left( L - f \sum_{j=1}^n \theta_j \right) - fQ + \sum_{j=1}^n \beta_j \left( \frac{\partial q_k}{\partial \theta_k} \frac{\partial B_j}{\partial q_k} + \frac{\partial B_j}{\partial \theta_k} \right) \frac{\partial q_j}{\partial B_j} = \bar{\mu}_k - \underline{\mu}_k \quad (11)$$

where  $\beta_j$  is the Lagrange multiplier on the  $j$ th inequality constraint and where  $\underline{\mu}_k$  and  $\bar{\mu}_k$  are the Lagrange multipliers on  $\theta_k \geq 0$  and  $\theta_k \leq 1$ .<sup>11</sup> We see from the equation that  $\theta_k$  influences the objective function through its *direct* and its *indirect* impact on the objective function. The direct impact captures the increase in productivity of firm  $q_k$  which leads to an increase in aggregate productivity  $Q$  (first term in 11). Increasing  $\theta_k$  also lead to higher costs of operation (second term in 11). The indirect impact of  $\theta_k$  operates through its action on the other firms. First, its impact on  $q_k$  leads to an increase in the labor productivity of all of  $k$ 's customers (first term in the summed parenthesis). Second, increasing  $\theta_k$  also increases the composite input of all of  $k$ 's customers through the reshaping term  $\theta_k^b$  (second term in the summed parenthesis).

Equation 11 determines the operational status  $\theta_k$  of firm  $k$ . By the complementary slackness condition, if its left-hand side is larger than 0 the firm is active ( $\theta_k = 1$ ) and, similarly, if it is smaller than 0 the firm is inactive ( $\theta_k = 0$ ). The key to the reshaping approach is to make the left-hand side of (11) independent of any firm-specific endogenous variables. The following lemma shows how this can be done.

<sup>10</sup> $\mathcal{P}'_{RR}$  is equivalent to  $\mathcal{P}_{RR}$  since the inequality constraint always binds at the optimum.

<sup>11</sup>The partial derivatives of  $q_j$  are to be understood for the binding inequality constraint, i.e. such that  $\frac{\partial q_j}{\partial \theta_j} = A z_j a \theta_j^{a-1} A B_j^\alpha$  and  $\frac{\partial q_j}{\partial B_j} = A z_j \theta_j^a A \alpha B_j^{\alpha-1}$ .

**Lemma 2.** *The first-order condition (11) can be written as*

$$z_k^{\sigma-1} \theta_k^{a(\sigma-1)-1} F_1(Q, B_k) \left( L - f \sum_{j=1}^n \theta_j \right) - fQ + z_k^{\epsilon-1} \theta_k^{b-1+a(\epsilon-1)} \sum_{j=1}^n \beta_j \Omega_{kj} F_2(B_k, B_j) \frac{\partial q_j}{\partial B_j} = \bar{\mu}_k - \underline{\mu}_k \quad (12)$$

where  $F_1$  and  $F_2$  are functions that only depend on the aggregates  $B$ 's and  $Q$ .

To understand how the reshaping approach operates, notice that under the reshaping parameters  $(\star)$ , the  $\theta_k$ 's cancel out from the first-order condition (12). The only remaining impact of  $\theta_k$  on the equation is through the variables  $B$ 's and  $Q$ . But these variables are aggregate, in the sense that they are summations over a large number of firms. Therefore, as the number of firms increases, and the network becomes increasingly connected, they become increasingly independent of  $\theta_k$  and, with them, so does equation (12). In the limit, the marginal benefit of increasing  $\theta_k$ , the left-hand side of (12), is therefore independent of  $\theta_k$  itself and depends only on the aggregate state of the network and on the firm  $k$ 's exogenous characteristics. If this marginal benefit is positive, the planner increases  $\theta_k$  until it reaches 1. If it is negative, the planner decreases  $\theta_k$  until it reaches 0. The solution to  $\mathcal{P}_{RR}$  is therefore such that  $\theta_k \in \{0, 1\}$  for all  $k$ . Using Proposition 4 we can then conclude that the solution to  $\mathcal{P}_{RR}$  also solves  $\mathcal{P}_{SP}$ .

### Example with $n = 2$

To better understand how the reshaping approach operates, it is useful to consider a simple economy with two firms and a complete set of potential connections  $\Omega$  between them. Consider first the relaxed problem without any reshaping ( $a = 1, b = 0$ ). The contour plot of its objective function  $V_R(\theta)$  is shown in Figure 2a. We see that  $V_R$  is shaped like a saddle with local maxima at  $(1, 0)$  and  $(0, 1)$  and local minima at  $(0, 0)$  and  $(1, 1)$ . The planner therefore prefers to have a single firm operating than to have both, or neither, firms operating. The global maximum is at  $\theta = (1, 0)$ .

Since  $V_R$  is not concave, attempts at finding the global maximum will encounter two related problems. First, since the first-order conditions are not sufficient to characterize the optimum, finding a solution to the FOCs does not guarantee that this solution is the global maximum. Second, this problem cannot be solved reliably with standard numerical algorithms. Starting from an initial point, these algorithms generally operate by following the steepest slope, which can easily lead to the local minimum at  $(0, 1)$ .

Consider instead the objective function  $V_{RR}$  of the same optimization problem but reshaped with the parameters  $\star$ . Figure 2b shows the contour plot of  $V_{RR}$ . Notice first that the reshaping has not changed the value of the objective function in the corners  $\theta \in \{0, 1\}^2$ . As a result, the corner

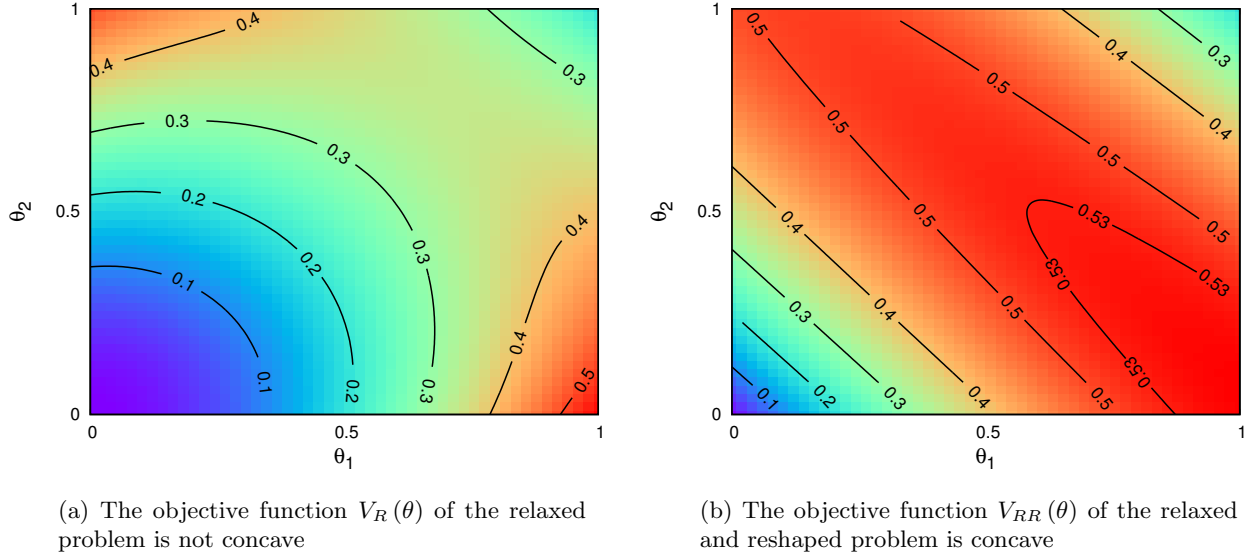


Figure 2: Example: reshaping the objective function

with the highest utility according to  $V_{RR}$  is also the corner with the highest utility according to  $V_R$ . Second, we see that the reshaping basically stretched the objective function between the corners so as to transform the non-concave function  $V_R$  into the concave function  $V_{RR}$ . The first-order conditions are now sufficient to characterize a solution and, starting from any point  $(\theta_1, \theta_2) \in [0, 1]^2$ , a simple numerical algorithm like gradient ascent will converge to the global maximum at  $(0, 1)$ . Since the reshaping preserves the ranking of the corners,  $(0, 1)$  is also the global maximum of  $V_R$  and therefore of  $\mathcal{P}_{SP}$ .

From the example of Figure 2, one can get some intuition about how the reshaping approach makes the problem easier to solve. Let us first consider the role played by the parameter  $a$ . Suppose that we begin the optimization at the local maximum  $\theta = (1, 0)$  in Figure 2a and are now looking for alternative allocations in its neighborhood to improve the objective function. Moving towards the global maximum at  $\theta = (1, 1)$  incurs a marginal cost proportional to  $fQ$ . The marginal benefit of this move is, however, proportional to  $\partial Q / \partial \theta_2 \propto \theta_2^{\sigma-2}$  without the deformation. As a result, for  $\theta_2$  small, the marginal benefit is close to 0 and it declines with the elasticity  $\sigma$  — operating a new firm provides little benefit if the goods are highly substitutable. Increasing  $\theta_2$  is thus not *locally* advantageous and the optimization converges erroneously to the local maximum. Including the reshaping parameter  $a$  changes the incentives to increase  $\theta_2$  around  $\theta_2$ . In particular, the marginal benefit is now proportional to  $\partial Q / \partial \theta_2 \propto \theta_k^{a(\sigma-1)-1}$ . Under the parameter values  $(\star)$ , this benefit is constant and of the same order of magnitude as the cost  $fQ$ , which helps the optimization algorithm to move towards the global maximum. A similar intuition applies for the reshaping parameter  $b$ , but the argument involves the addition of an inputs into the production of intermediate goods.

## 2.3 Testing the Solution Approach

The theoretical results of the last section show that the reshaping approach finds the correct solution to  $\mathcal{P}_{SP}$  when the number of firms is large and the network is sufficiently connected. To verify that it also works well when these assumptions are relaxed, I randomly generate a large set of networks with various parameters. I then compare the allocation found by the reshaping approach to the correct solution given by exhaustively evaluating the utility provided by the  $2^n$  possible networks. As explained earlier, this exhaustive search becomes impossible to conduct for  $n$  large so that the comparisons are limited to relatively small networks.<sup>12</sup> Nevertheless, since small and sparse networks are the ones for which we expect the reshaping approach to fail, the exercise provides a useful test of its performance.

The results are presented in Table 1. We see that the relaxing and reshaping approach works remarkably well even for small and sparsely connected network. The algorithm attributes the correct operational status to more than 99.8% of the firms. The approach also finds output values that are, on average, within 0.002% of the correct value. In contrast, attempting to solve for the efficient allocation by only relaxing the problem yields more than 15% of the firms with the wrong operational status and the average error in output can be close to 1%, a large number when considering business cycle fluctuations.<sup>13</sup>

	Number of firms $n$			
	8	10	12	14
A. With reshaping				
Firms with correct $\theta_j$	99.9%	99.9%	99.9%	99.8%
Error in output $Y$	0.00039%	0.00081%	0.00174%	0.00171%
B. Without reshaping				
Firms with correct $\theta_j$	84.3%	83.2%	82.3%	81.3%
Error in output $Y$	0.84%	0.89%	0.93%	0.98%

Notes: I construct networks with parameters  $f \in \{0.05/n, 0.1/n, 0.15/n\}$ ,  $\sigma_z \in \{0.34, 0.39, 0.44\}$ ,  $\alpha \in \{0.45, 0.5, 0.55\}$ ,  $\sigma \in \{4, 6, 8\}$  and  $\epsilon \in \{4, 6, 8\}$ . For each possible combination of the parameters, 1000 different economies are created. For each economy, productivity is drawn from  $\log(z_k) \sim \text{iid } \mathcal{N}(0, \sigma_z)$  and  $\Omega$  is drawn randomly such that a firm has on average five possible incoming connections. As a result, the networks  $\Omega$  have different degrees of sparsity. All draws are independent. The network is redrawn until it is strongly connected. Appendix C.1 describes the algorithm used to create  $\Omega$ . A network is kept in the sample only if the first-order conditions converge to a solution for which  $\theta$  hits the bounds.

Table 1: Testing the reshaping approach on small networks

It is also possible to test the relaxing and reshaping approach on large network. While the true solution to the planner's problem is unknown in this case, we can verify whether there exist benefits from deviating from the allocation given by the reshaping approach. In particular, I verify whether changing the status of a single firm improves the utility of the planner. I keep repeating

<sup>12</sup>Appendix C.2 describes the algorithms used to solved for the efficient allocation.

<sup>13</sup>Since the problem is solved numerically, any algorithm would be expected to converge on the wrong solution from time to time. For instance, the algorithm might converge to a point that is not a solution but for which the first-order conditions are satisfied within numerical tolerance.

this procedure as long as there are deviations to be found and then compare this deviation-free solution to the original one given by the reshaping approach.

Since this procedure is computationally costly, I consider only large economies that satisfy the calibration presented in Section 4.2. The results are presented in Table 2. Again, the reshaping approach performs very well. After all the possible deviations are accounted for, 99.8% of the firms have kept their status and aggregate output has changed by a negligible amount. In contrast, without reshaping the planner’s objective function, the deviations reveal that almost 28% of the firms were assigned the wrong operational status and the error in aggregate output amounts to about 0.7%. While this test does not guarantee that the reshaping approach finds the correct efficient allocation, it provides a good indication that there are no obvious mistakes in its solution.

	With reshaping	Without reshaping
Firms with correct $\theta_j$	99.8%	72.1%
Error in output $Y$	0.00028%	0.69647%

Notes: I simulate 200 different networks  $\Omega$  and productivity vectors  $z$  that satisfy the properties of the calibrated economy (see Section 4.2). I run the procedure described in the text on each of them and report average results.

Table 2: Testing the reshaping approach on large networks

### 3 Economic Forces at Work

We now explore how the economic forces at work in the model shape the network structure of production and how this shaping, in turns, interacts with firm-level outcomes.

#### 3.1 Selection of firms

The problem  $\mathcal{P}_{SP}$  highlights the two main forces that push the planner to operate a firm. First, from the expression for aggregate output (8), we see that the planner is likely to operate high- $q$  firms. Doing so contributes directly to aggregate output and therefore to the welfare of the household. Second, from the recursive expression determining  $q$  (7), we see that the planner is likely to operate a firm if doing so increases the  $q$ ’s of other firms, thereby indirectly increasing the household’s utility. Because of these two motives, the planner tends to operate firms with high total factor productivity  $z$ . Not only do these firms produce high- $q$  goods for the production of the final good, they also provide valuable inputs to their customers, which in turn increases these firms’ contribution to aggregate output and benefits their own customers, and their customers’ customers and so on.

In addition, the planner also operates firms that are advantageously positioned in the network. For instance, operating a firm with many *outgoing* active connections benefits a large number of other firms and these firms customers, etc. Similarly, we know from (7) that a firm with many



*incoming* connections is likely to have a high  $q$ , making it attractive to operate. As a result of these two channels, the planner is likely to operate firms that are well-connected, either in terms of suppliers or customers, even if their total factor productivity  $z$  is low.

### 3.2 Complementarities and clustering of activity

Since the benefit of operating a firm is larger if it is well-connected, the economy features complementarities in operation decisions between nearby firms. To maximize the benefit from these complementarities, the planner tends to organize economic activity in tightly connected clusters of operating firms. In this configuration, each firm  $j$  benefits from the presence of operating neighbors that, in turn, benefit from a more productive firm  $j$ . This recursive process reinforces the labor productivity  $q$  of all the members of the cluster and makes this way of organizing production very efficient.

A simple example is useful to understand the economic forces that make clustering more or less desirable. Suppose that the planner wants to operate  $n$  firms and can decide to do so by operating clusters of  $m$  fully-connected firms, where each producer supplies to the  $m - 1$  other firms.<sup>14</sup> Which amount of clustering  $m$  is preferred? We can easily compute the productivity  $q$  of each firm in an  $m$ -cluster. Assuming that firms share the same  $z = 1$  for simplicity, symmetry imposes that all firms have the same  $q$ . From (7), we find that  $q = (m - 1)^{\frac{\alpha}{1-\alpha} \frac{1}{\epsilon-1}}$  and from (8) we can compute total output

$$Y = n^{\frac{1}{\sigma-1}} (m - 1)^{\frac{\alpha}{1-\alpha} \frac{1}{\epsilon-1}} (L - fn).$$

Since this expression is increasing in  $m$ , the planner prefers to operate a unique tightly connected group of firms instead of a large number of connected pairs of firms. This preference for clustering is also more important when the share  $\alpha$  of intermediate inputs in production is large and when the elasticity of substitution  $\epsilon$  between these inputs is low. Under these conditions, the planner values additional connections more, which leads to a heightened desire for clustering.

When we introduce shocks to  $z$ , the optimal network tends to cluster economic activity around the most productive firms. Doing so allows the planner to magnify the impact of high- $z$  firms. Figure 3 presents an example of this process in an economy in which the network  $\Omega$  consists of two groups of highly connected firms linked together by a pair of connections. The figure shows the same economy under four different total factor productivity vectors  $z$ . We see that, as the previous example showed, economic activity is concentrated in either the top or the bottom group of firms, the one with the highest- $z$  firm, while it leaves the other group inactive.

---

<sup>14</sup>For simplicity, let us assume that  $n$  is large enough and that it is divisible by  $m$ . While the current example might seem abstract it fits in the framework outlined in the last section. The whole economy would be formed of groups of  $n$  firms which various degree of clustering. The exercise would then be about which of these groups the planner would prefer to operate.

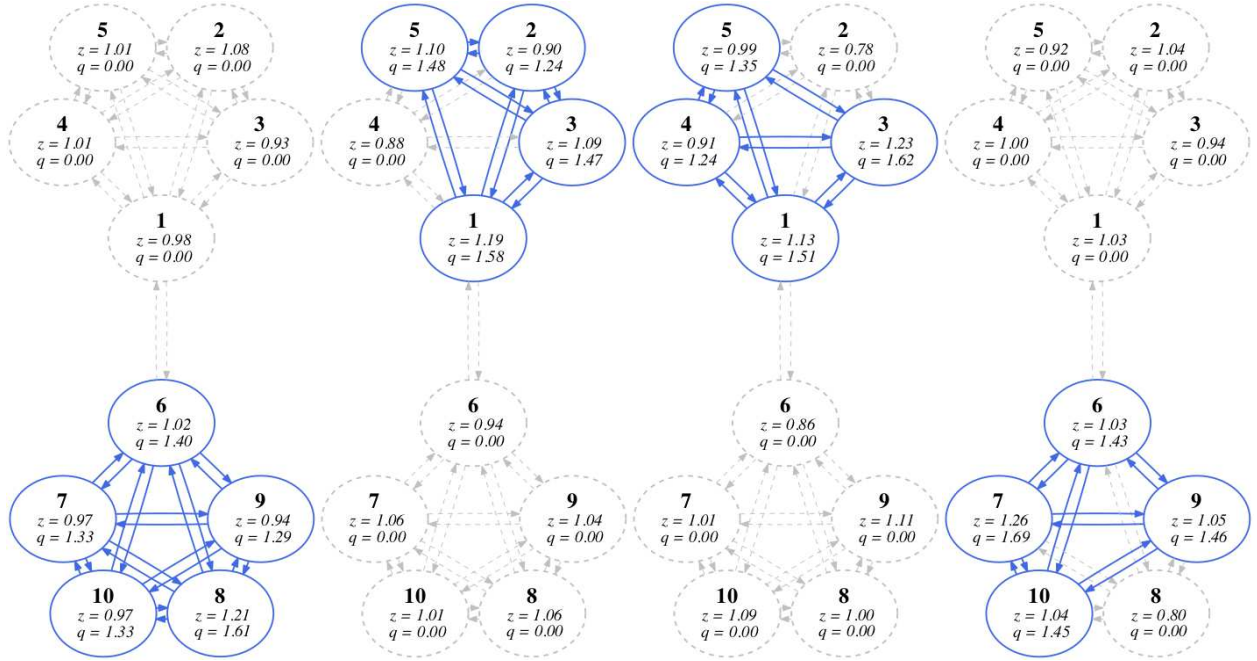


Figure 3: The optimal network features clusters of economic activity

### 3.3 Small shock and large reorganization

In the efficient allocation, a small variation in the environment can lead to a large reorganization of the network. This unusual feature arises because of the non-convex nature of the planner's problem. To determine the optimal allocation, the planner essentially compares the  $2^n$  potential network configurations and selects the one providing the highest utility. As the environment changes, say a firm's  $z$  increases, a previously optimal configuration is replaced by a new one that can be substantially different.

Figure 4 shows an example of this mechanism. The primitives of the two networks are identical except that in the economy on the left the central node is slightly more productive than in the economy on the right. While the drop in productivity from left to right is negligible, it triggers a large reorganization of the network structure of production. On the left, the planner prefers to make goods transit through the productive central firm while on the right it creates a loop of active firms to avoid the slightly less productive central firm.

While the shape of the network changes substantially, aggregate output is barely affected by the drop in productivity. Indeed, the planner reorganizes the network precisely so that the impact of the bad shock is as small as possible. Even though it leaves output mostly unaffected, the reorganization of the network can have a large impact on the distributions of firm-level variables. In this example, for instance, the dispersion in labor productivity, output and employment collapses after the shock. A negligible shock can therefore have a large impact on cross-producer distributions.

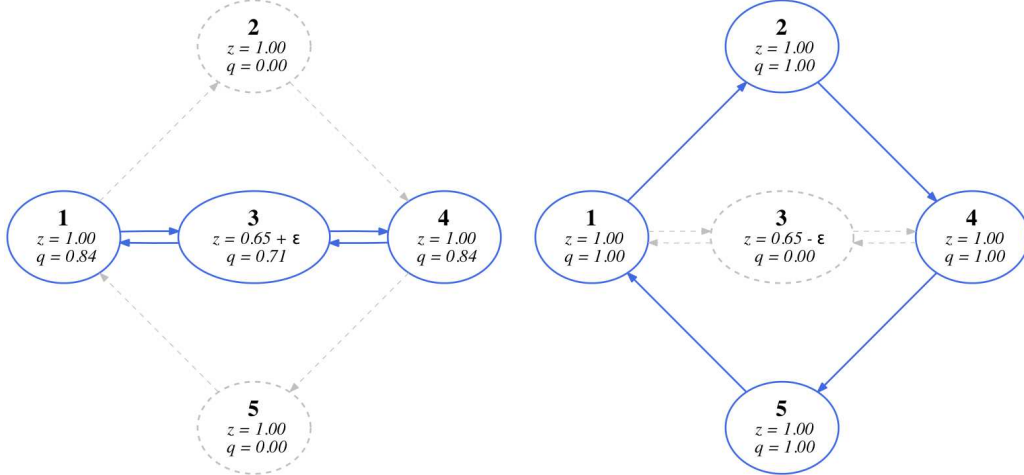


Figure 4: A small shock triggers a large reorganization of the economy

## 4 Quantitative Exploration

To further investigate the mechanisms at work in the economy, I calibrate the model to the United States. I then characterize the shape of the production network and explain how its shape is influenced by various economic forces. I then consider how the optimal design of the production network affects firm-level outcomes, and how large cascades of firm shutdowns can arise after a severe shock to an important producer. Finally, I investigate how the optimal adjustment of the network in response to shocks influence aggregate fluctuations.

### 4.1 Network data

Bringing the model to the data requires taking a stand on what constitutes a unit of production in this economy. While a natural interpretation is that a producer corresponds to a plant, I am unaware of readily available data about input-output linkages at the plant level in the U.S. There is however firm-level data available through the Compustat Segment database. These data were built from public firms' financial disclosures and span the period 1979-2014. They contain firms' own reports of their major customers.<sup>15</sup> Cohen and Frazzini (2008) and Atalay et al. (2011) use these data to construct the network structure of production in the U.S.<sup>16</sup> I use the outcome of their analyses to calibrate the model and verify some predictions of the theory.

Information about the network structure of production is also available for Japan. The mar-

<sup>15</sup>Firms report these customers in accordance with Financial Accounting Standards No. 131. A major customer is defined as a firm responsible for purchasing at least 10% of the seller's revenue. Some firms sometimes report smaller customers.

<sup>16</sup>I thank the authors for sharing their data. Their samples cover the periods 1980 to 2004 and 1976 to 2009, respectively. Since the customer information reported by the firms contains many typos and various abbreviations, both papers use a combination of automatic algorithms and manual matching to create the production networks. As these approaches are imperfect, I use both datasets to verify the predictions of the theory. The Compustat Segment data is also used by Kelly et al. (2013) and Wu (2016) among others.

keting firm Tokyo Shoko Research has built a large dataset containing supplier-customer linkages for almost all Japanese firms with more than 4 employees. I also refer to some properties of this network, as reported by [Saito et al. \(2007\)](#) and [Bernard et al. \(2015\)](#).

## 4.2 Parametrization

I normalize  $A = 1$  and  $L = 1$ . For the share of intermediate goods in the production function, I follow [Jorgenson et al. \(1987\)](#) and sets  $\alpha = 0.5$ . [Jones \(2011\)](#) surveys the literature on the share of intermediate goods in different countries and also suggests  $\alpha = 0.5$  as a good benchmark.

[Broda and Weinstein \(2006\)](#) uses disaggregated trade data for the U.S. to estimate the elasticity of substitution  $\sigma$  between product varieties. Their estimates vary greatly across sectors, levels of aggregation as well as time periods. At the highest level of disaggregation (10-digit level), the mean estimate for the 1990-2001 period is 12.6 while the median is 3.1. At the 5-digit level, the mean estimate is 6.6 while the median is 2.7. I set  $\sigma = 6$  as an average of these estimates. The empirical literature provides little guidance about the elasticity of substitution between intermediate inputs at the firm level. I therefore set  $\epsilon = \sigma$  as default and provide robustness analysis along the way.<sup>17</sup>

I assume that the log of the productivities  $z_{it}$  are drawn in an i.i.d. way from a normal distribution with mean zero and standard deviation  $\sigma_z$ . Many studies document a very large dispersion in producer productivity. For instance, [Bartelsman et al. \(2013\)](#) measure the dispersion in firm-level physical productivity in a number of countries and find  $\sigma_z = 0.39$  for the U.S. I adopt this number for the calibration. Importantly for our purpose, their estimation technique controls for the usage of intermediate inputs.

There is no consensus in the literature about the value of overhead labor  $f$ . Since employment in management occupations account for about 5% of total employment in the U.S., I set  $f$  so that  $f \times n = 5\%$ .<sup>18</sup>

We are left with the *unobserved* set of potential links  $\Omega$  to parametrize. So that the results shown below do not rest on one specific network, the simulations involve multiple networks  $\Omega$ . They are all drawn randomly so that the distribution of the number of potential suppliers follows a power law with a shape parameter of  $\alpha_\Omega = 1$  — the celebrated Zipf’s law.<sup>19</sup> By parametrizing  $\Omega$  in this way, the *observed* in-degree distribution (the number of suppliers) in the efficient allocation is close to a power law with a shape parameter that is near the one observed in the data, as we

---

<sup>17</sup>While no estimate is available at the firm level, [Atalay \(2015\)](#) finds that the elasticity of substitution between inputs at the sector level is likely smaller than one so that inputs are complement. This elasticity is too extreme for the current model — losing a single supplier would lead to no production at all. Nonetheless I consider an economy with a low elasticity  $\epsilon$  as a robustness exercise.

<sup>18</sup>These fixed costs have often been estimated using models of firm turnover. For instance, [Bartelsman et al. \(2013\)](#) estimate a model of firm entry and exit on the manufacturing sector and estimate that overhead labor accounts for 14% of total employment in the industry. This number is likely to be lower for firms outside of manufacturing. Higher fixed costs do not affect the conclusions of this section significantly but require solving economies with more firms, which is numerically costly.

<sup>19</sup>The results of this section are similar if the network is built with the approach of [Atalay et al. \(2011\)](#) instead.

will see below. In addition, Zipf’s law arises from a wide variety of processes. For instance, in the network literature, it emerges as a limiting case of the Simon model (Simon, 1955) and the Price model (de Solla Price, 1965).<sup>20</sup> For these reasons, Zipf’s law seems to be a natural choice for the shape of  $\Omega$ .<sup>21</sup> Finally, I set  $n = 3000$  to roughly match the average number of firms in a year in the Atalay et al. (2011) dataset.

Table 3 displays the parameters of the calibrated economy. Using this parameterization, I create 32 matrices  $\Omega$ .<sup>22</sup> For each of them, I then draw 400 productivity vectors  $z$  and solve for the efficient allocation in each case. The next sections describe various features of these simulated economies.<sup>23</sup>

Parameter	Value
Average productivity	$A = 1$
Labor supply	$L = 1$
Size of the network	$n = 3000$
Share of intermediate goods	$\alpha = 0.5$
Elasticity of substitution for final goods	$\sigma = 6$
Elasticity of substitution for intermediate goods	$\epsilon = 6$
Standard deviation of individual productivities	$\sigma_z = 0.39$
Fixed cost of operation	$f = 0.05/n$
Set of possible connections	$\Omega$ see text

Table 3: Parameters of the calibrated economy

### 4.3 The shape of the network

We now consider the shape of the network of active connections in the efficient allocation. Figure 5 shows the distributions of the number of suppliers (in-degree) and the number of costumers (out-degree) in the benchmark economy (blue circles).<sup>24</sup> To properly highlight the shape of these distributions, the figure uses a log-log scale and plots the complementary cumulative distributions

<sup>20</sup>More generally, Zipf’s law also arises from proportional random growth (Gabaix, 2009) and in assignment models with complementarities (Geerolf, 2016).

<sup>21</sup>In addition to the connections drawn from Zipf’s law, I also add three potential connections to all firms. I do this for two reasons. First, it greatly helps the first-order conditions to converge, making the efficient allocation faster to solve and therefore allowing for a larger number of simulations. Second, empirical evidence from Bernard et al. (2015) finds that firms with few connections are more abundant than the power law would predict. Since these additional connections have a minimal impact on the highly-connected firms of the economy, those that are important for aggregate fluctuations, removing them has a minimal impact on the simulations.

<sup>22</sup>See Appendix C.1 for a description of the algorithm used to create  $\Omega$ .

<sup>23</sup>I discard and redraw simulations for which iterating on the first-order conditions does not converge to a point  $\theta$  such that  $\theta_j \in \{0, 1\}$  for all  $j$ . All moments in the next sections are computed on each network  $\Omega$  first and then averaged across networks.

<sup>24</sup>Another measure often used in the network literature is that of the centrality of a node. I computed the centrality vector in this economy using the same measure as Carvalho (2014). The distribution of centrality is close to a power law with shape parameter 1.39. Carvalho (2014) estimates this parameter in the U.S. data at the sector level and finds a value of 1.4. This centrality measure plays a role very similar to the out-degree variable and is therefore omitted from the analysis.

(CCDF) on the vertical axis. We see that the distributions are roughly linear in that space, suggesting that they are close to power laws.

In-degree and out-degree distributions have also been found to roughly follow power laws in the data. [Atalay et al. \(2011\)](#) and [Saito et al. \(2007\)](#) find that the empirical in-degree distributions for U.S. and Japanese firms are also well approximated by power laws with shape parameters 1.37 and 1.3.<sup>25</sup> The corresponding number in the benchmark economy is 1.43.<sup>26</sup> The efficient allocation is therefore able to replicate this important property of empirical networks of production.

To see how the in-degree and out-degree distributions are affected by features of the environment, Panels (a) and (b) also show an economy with lower elasticity of substitution between goods  $\epsilon$  (red squares). In this case, the value of an extra input to a firm is higher and the planner naturally creates a more densely connected network. From the figure, we see that all firms are essentially gaining customers and suppliers so that both the in-degree and out-degree distributions roughly shift to the right. The number of active firms also increases substantially.<sup>27</sup>

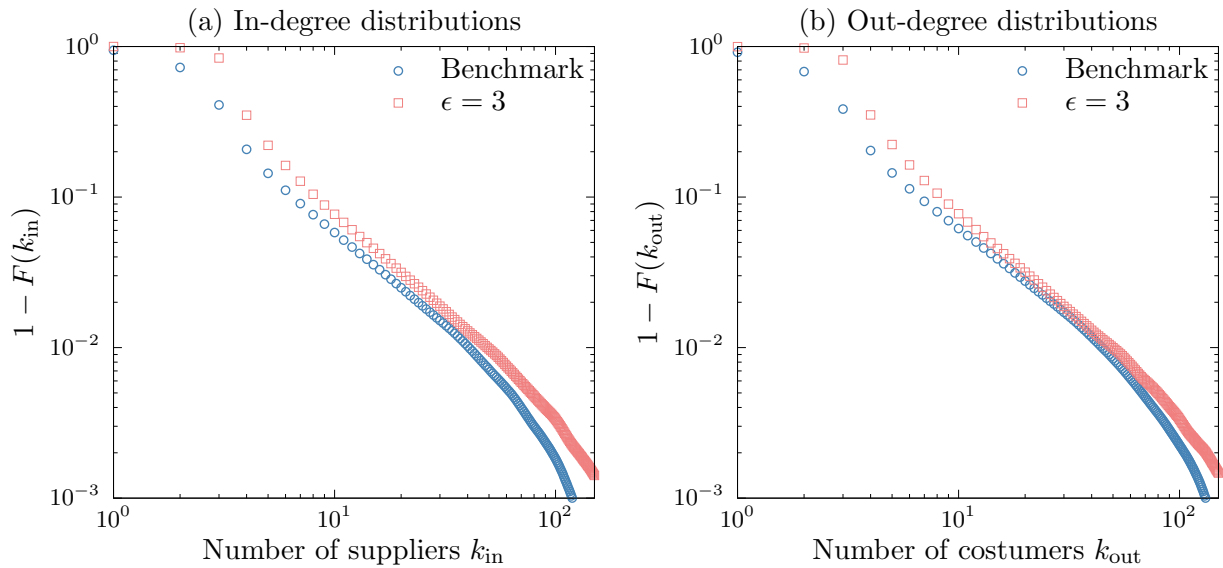


Figure 5: Distribution of the number of suppliers and the number of customers

<sup>25</sup>[Atalay et al. \(2011\)](#) reports that the mean of the number of suppliers in their sample is 3.67 which amounts to a shape parameter of 1.37. They also document a departure from the power law for firms with many suppliers. They provide an explanation that relies on firms exit and links reattachment. This departure is less visible in the Japanese data ([Bernard et al., 2015](#)).

<sup>26</sup>All shape parameters of power law distributions are estimated using the approach of [Gabaix and Ibragimov \(2011\)](#). To properly capture the shape of the right tail of the distributions, I exclude observations with less than three connections from the sample.

<sup>27</sup>When, instead, the share of intermediate inputs  $\alpha$  is larger, inputs matter more for production and the planner connects the high- $q$  firms with more customers. This mechanism is also at work in [Oberfield \(2013\)](#).

**Comparison with a random network.** To understand what features of a network are more desirable for efficiency, it is useful to compare the *optimal* network, the one that arises in the efficient allocation, to a completely *random* one.<sup>28</sup> Table 4 shows the shape parameters of the in-degree and out-degree distributions for the benchmark economy in this random network as well as in the optimal one. Since its in-degree and out-degree distributions possess fatter tails (smaller shape parameters), the optimal network features a larger fraction of highly connected suppliers and highly connected customers than the random network. These shifts in the distributions occur because high in-degree firms and high out-degree firms are particularly valuable. The former, because of their many suppliers, have high productivity  $q$  which the planner values. The latter provide valuable inputs to many customers thereby increasing these firms' productivity  $q$ . As a result, when designing the network optimally, the planner tends to operate these influential firms more than their poorly-connected counterparts.

	Optimal network	Random network
A. Pareto shape parameters		
In-degree	1.43	1.48
Out-degree	1.37	1.48
B. Clustering coefficient	0.027	0.018

Notes: The distances and the clustering coefficients are computed in the undirected graphs to avoid the many infinite distances that arise in the directed graphs. Infinite distances (arising from unconnected parts of the network) are excluded.

Table 4: Shape of the optimal and random networks

The optimal and the random networks also feature different degrees of clustering between firms. To quantify this difference, I use the standard global clustering coefficient commonly used in the network literature. This measure considers triplets of connected nodes, and compares the total number of such triplets that are fully connected with the number of triplets that are not. As Table 4 shows, this clustering coefficient is 47% larger in the optimal network than in the random one, indicating that designing the network optimally involves creating more tightly connected clustered firms, as the theory suggested.<sup>29</sup>

Another distinctive feature of the optimal network is the correlation between the in-degree and out-degree of a firm. In the random network, these those quantities are uncorrelated by construction. In the optimal network, however, the correlation is positive. The planner constructs the network in this way for two reasons. First, increasing the number of suppliers of a firm with high out-degree increases the productivity  $q$  of its many customers, which directly leads to higher output. Second, since firms with many suppliers are more productive, the planner seeks to benefit

<sup>28</sup>The random network is built by drawing a random set of active firms  $\theta$  for each network  $\Omega$ . I redraw  $\theta$  until the random network has as many active firms as the optimal one. Except for  $\theta$ , all the other quantities in the random network economy are chosen optimally.

<sup>29</sup>A consequence of this increased clustering is that firms are closer to one another: the average distance between two firms, in terms of number of connections, is 15% smaller in the optimal network.



as much as possible from this high  $q$  by increasing their number of customers. Together, these two forces work to create the positive correlation observed in the simulated data.

#### 4.4 Firm-level outcomes

We now consider the implications of the endogenous formation of the network for firm-level outcomes. Table 5 shows the ordinary least-square coefficients from regressing the labor productivity  $q$  and the employment  $l$  of firms on their number of suppliers and customers.

The regressions find that firms with more suppliers and more customers also employ more workers. Again, this finding is not surprising in view of equation 9. Firms with many suppliers are high- $q$  firms and the planner therefore allocates a lot of labor to them (ratio  $q/Q$  in the equation). High out-degree firms also receive more labor for them to supply to their many customers (term  $\Gamma$  in the equation). The allocation therefore features a similar correlation between the number of neighbors and output. Saito et al. (2007) and Bernard et al. (2015) find that firms with many suppliers and many customers are also larger in terms of sales in the Japanese data.

The role of a firm's in-degree and out-degree on its productivity  $q$  is less clear. In the first two columns of Table 5, the regressions show a positive relationship between in-degree and  $q$  but a negative relationship between out-degree and  $q$ . The first result is not surprising. Because of the production function, firms with many suppliers have higher  $q$ . The second result, however, emerges because of the positive selection of high out-degree firms. Since they are important providers of goods in the network, the planner tends to operate them *even if they have low productivity*. In contrast, low productivity firms with few customers do not operate. This selection process is responsible for the negative correlation observed in the simulated data.

The last two columns of the table confirm this conclusion. They show the same ordinary least-square coefficients as the first two columns but with the addition of firms fixed effects in the regressions. The coefficients therefore capture how *changes* in a firm's in- and out-degrees are related to its employment  $l$  and its labor productivity  $q$ . All coefficients are now positive. In particular, firms with many suppliers are now associated with a larger  $q$ . To understand why, consider a firm that is subject to a positive shock  $z$ , the source of variation in this regression. Because of the shock, this firm has a larger  $q$  than usual. To maximize the benefit of this  $q$  the planner tends to operate producers that are downstream from the firm, thereby increasing its out-degree, thereby creating this positive correlation.

The results of Table 5 highlight a two-way interaction between the shape of the network and firm-level outcomes. On the one hand, the shape of the network affects firm-level variables; for instance, firms with many suppliers have larger  $q$ 's. On the other hand, firm-level variables affect the shape of the network itself; for instance, the planner increases the number of customers of productive firms. Table 5 shows how this interaction creates correlations at the firm level.



Dependent variable	Employment $l$	Labor prod. $q$	Employment $l$	Labor prod. $q$
In-degree	0.36	0.08	0.75	0.15
Out-degree	0.44	-0.05	0.64	0.03
Firm fixed effects	no	no	yes	yes

Notes: Ordinary least-square regressions. Standard errors are negligible. All variables are in logs. Labor productivity is  $q$  and employment is  $l$ . Firm fixed effects are included to control for firm positions in the network of potential connections  $\Omega$ .

Table 5: Suppliers, customers and firm-level outcomes

The same interaction also has important implications for the distributions of firm-level outcomes. One way to understand the role it plays is to once again compare the optimal network and the random network economies. To do so, Table 6 shows moments of the cross-sectional distributions of labor productivity  $q$  and employment  $l$  in both economies. Consider skewness first. The table shows that the distributions in the optimal network economy are substantially skewed to the right. In contrast, the same distributions are essentially symmetrical in the random network economy. The endogenous adjustment of the network in response to shocks is responsible for this difference through two channels. First, the planner tends to shutdown firms with low productivity. This *selection* mechanism makes the left tails of the distributions thinner. Second, the planner reorganizes the network to *magnify* the impact of high- $z$  firms by also operating their suppliers. Doing so increases these firms' labor productivity  $q$  and, through the optimal allocation of labor, their employment. Doing so pushes these firms, already in the right tails of the distributions, even further to the right. Organizing the network efficiently therefore implies a thicker right tail in the firm-level distributions and the positive skewnesses.

These selection and magnification mechanisms also compete to influence the standard deviations of the distributions. In the calibrated economy, the selection dominates and the optimal network economy features more concentrated distributions. Figure 6 shows the effects of these forces on the distributions of labor productivity  $q$ .

	Labor prod. $q$	Employment $l$
A. Optimal network economy		
Standard deviation	0.29	1.24
Skewness	0.39	0.85
B. Random network economy		
Standard deviation	0.44	2.21
Skewness	-0.03	-0.05

Notes: All variables are in logs.

Table 6: Moments of firm-level outcomes in the optimal and random network economies

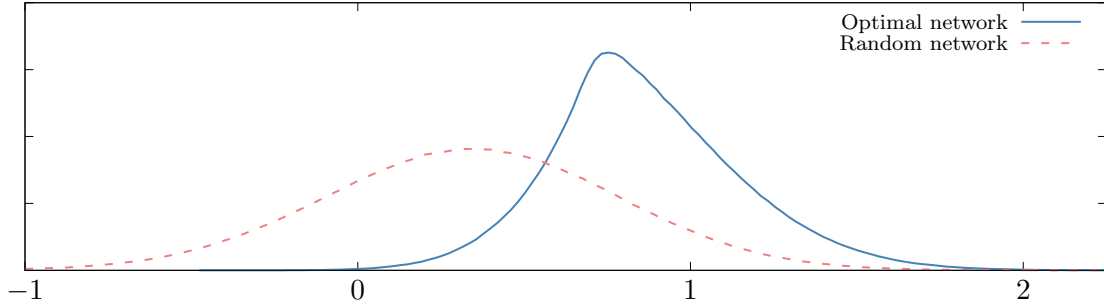


Figure 6: Distributions of the log of labor productivity  $q$

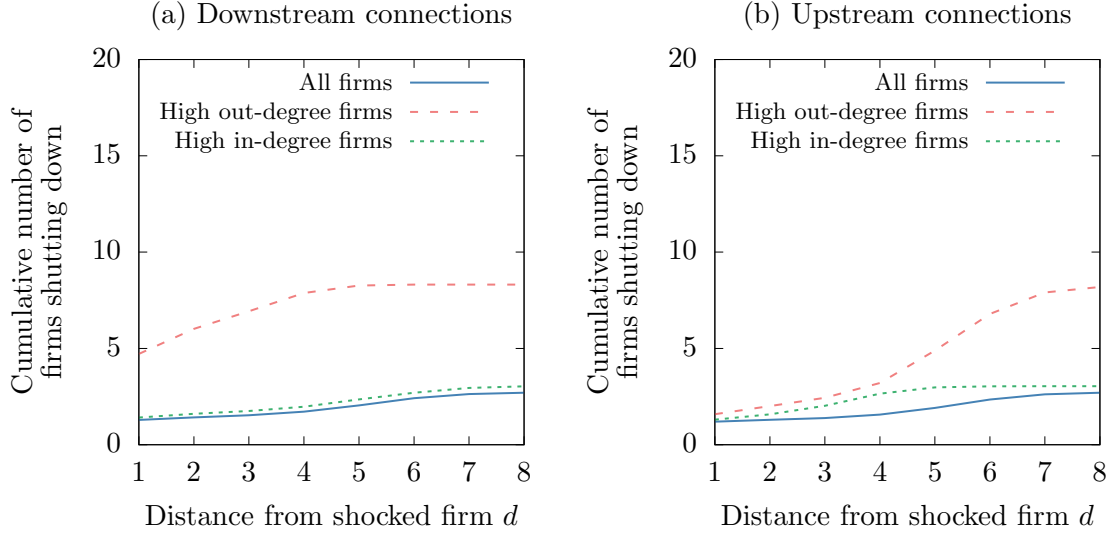
#### 4.5 Cascades of firm shutdowns

A severe productivity shock to an important firm in the network can trigger a cascade of firm shutdowns. Indeed, if the productivity of the targeted firm falls too much, the planner prefers to shut it down and to reallocate the labor from the fixed cost to more productive parts of the network. The neighbors of the targeted firms then lose a customer or a supplier, which lowers their contribution to aggregate output, and might lead to them shutting down as well. The same reasoning applies to the targeted firm's second neighbors and so on. As this cascade of shutdowns propagates through the network it can affect a large fraction of the firms in the economy.

To evaluate the importance of these cascades, I simulate a large number of economies in which a randomly selected producer is forced to shutdown by a sufficiently large negative  $z$  shock. Figure 7 shows the impact of the shuttered firm's exit on its neighbors. The plot in Panel (a) represents the cumulative number of *downstream* producers shutting down as we move further away from the targeted firm. Similarly, Panel (b) shows how *upstream* firms are affected. As we can see from the figure, a severe shock to an average firm (solid blue curves), is likely to lead to the shutdown of about three firms in the whole economy. While firms that are close to the shock are disproportionately affected in terms of their probability of shutting down (there are few firms there), the bulk of the shutdowns happen for firms that are far from the initial shock (where many firms are).

The situation is different if we consider only shocks to the largest suppliers in the network: firms in the 99<sup>th</sup> percentile of the out-degree distributions (dashed red curves). These shocks trigger the shutdown of many producers close to the distressed firm. Comparing Panels (a) and (b), we see that their impact propagates downstream, through the firms' customers, whose labor productivity  $q$  is directly affected by the shutdown of their supplier. The figure also shows shocks to the largest customers in the network: firms in the 99<sup>th</sup> percentile of the in-degree distribution (dashed green curves). While these shocks primarily trigger the shutdown of the suppliers of the targeted firm, their overall impact on the economy is not very different from a shock to an average firm.

Two key parameters, the elasticity of substitution between inputs  $\epsilon$  and the share of inputs into the production function  $\alpha$ , have a large impact on the magnitude of the cascades and how



Notes: The distance  $d$  measures the shortest path, in terms of numbers of connections, from the targeted firm. “High in-degree firms” refers to the 99th percentile of the in-degree distribution and similarly for out-degree firms.

Figure 7: Cascades of firm shutdowns in the benchmark calibration

they propagate. For instance, a higher  $\alpha$  increases the importance of the network in determining output. Shocks therefore lead to a larger reorganization of the network and, as a result, the cascades are more important. Lowering the elasticity of substitution  $\epsilon$  also affects the cascades in important ways. In particular, shocks to the largest customers can now trigger important cascades of shutdowns that propagates *upstream*, from supplier to supplier. This stark difference from the benchmark simulations can be explained by understanding why the planner operates small firms in this economy. Since  $\epsilon$  is small compared to  $\sigma$  the planner derives utility from them mostly as the providers of intermediate input to the high- $q$  firms and not so much for their direct contribution to final goods production. As a result, if one of the largest customers (a high- $q$  firm) shuts down, the planner is likely to shut its small suppliers down as well, thereby triggering an upstream cascade. In contrast, in the benchmark economy where  $\epsilon = \sigma$  the planner values the direct contribution to the final good from these small providers similarly to their contribution to the productivity of the high- $q$  firms. They are therefore more likely to remain if one of their large customer shuts down. Appendix A provides additional details and figures for cascades in these alternative economies.

In order to trigger a cascade, a firm must be hit by a shock strong enough to shut it down. Some firms are however much more resilient to shocks than others. To see this, Table 7 provides the probability that a one standard deviation  $z$  shock shuts down the firm — a measure of its resilience. Unsurprisingly, the largest customers and suppliers in the economy are much more likely to survive a bad shock than the average firm. Since these firms are important contributors to output, either directly or through their contributions to other firms, the planner keeps them active even when they are facing very low  $z$ ’s. Together, Figure 7 and Table 7 highlight a tradeoff between the resilience

of a firm and its impact on the network upon failure: well-connected firms resist well to shocks but trigger the shutdown of many of their neighbors when they stop production.<sup>30</sup>

Probability of firm shut down	
All firms	92%
High out-degree firms	20%
High in-degree firms	56%

Notes: “High in-degree firms” refers the 99th percentile of the in-degree distribution and similarly for out-degree firms. Firms are subject to a 1 standard deviation  $z$  shock.

Table 7: Resilience of targeted firms to shock

These cascades are efficient and are simply the manifestation of the optimal reorganization of the network in response to shocks. One way to estimate their importance in that regard is to consider how much more output would have fallen if the network was unable to reorganize itself after a shock. In order to do so, I compute the distribution of the ratio between the drop in output in an alternative rigid network to the one in the original flexible network. At the median point, the drop in output is 35% more severe in the rigid network, which suggests that allowing the network to reorganize itself in response to shocks is a powerful mechanism to prevent deep output losses.<sup>31</sup>

## 4.6 Aggregate fluctuations

We now consider how the optimal design of the network influences aggregate fluctuations. To do so, it is useful to first examine how the shape of the network varies with aggregate output. Table 8 provides the standard deviation of various measure of the shape of the network in the optimal allocation and their correlation with output. We see that, in booms, the tails of the in-degree and out-degree distributions become thicker, indicating that the economy features more highly-connected firms during expansions than during recessions. The clustering coefficient is also positively correlated with output. The network structure of production is therefore more tightly connected when output is high.

To understand why the planner adjusts the network in this way, consider the origin of fluctuations in this economy. Since the model does not feature aggregate shocks, expansions are mostly driven by positive shocks to influential firms in the network — those in the right tail of the in-degree and out-degree distributions — and by the ability of the planner to cluster production tightly around them. Recessions are therefore periods in which these influential firms suffer from bad shocks and are shut down, bringing with them the cluster that was built around them. This mechanism generates the correlations in the first column of Table 8.

The last two columns of Table 8 show that these correlations between the shape of the net-

<sup>30</sup>Appendix A shows that firms in the  $\alpha = 0.75$  and  $\epsilon = 3$  economies are much more resilient to shocks.

<sup>31</sup>In contrast, Baqaee (2016) proposes a model where, because of external effects, cascades can be inefficient.

work and output are also present in the two datasets on the U.S. production network constructed by Cohen and Frazzini (2008) and Atalay et al. (2011). As we can see from the table, the U.S. production network is more clustered and features more highly-connected firms during expansions than during recessions. These findings confirm that the predictions of the theory have support in the data.

	Correlation with output		
	Model	Data	
		CF (2008)	AHRS (2011)
A. Power law shape parameters			
In-degree	-0.10	-0.10	-0.21
Out-degree	-0.31	-0.24	-0.13
B. Clustering coefficient	0.47	0.70	0.15

Notes: All time series are in logs and linearly detrended. The clustering coefficients are computed on the undirected graphs. Data output is annual real gross domestic product over the same time period as the network sample.

Table 8: Aggregate fluctuations and the shape of the network

To quantify the impact of the optimal network on fluctuations, Table 9 shows the standard deviation of output  $Y$ , labor productivity  $Q$  and productive labor  $L - f \sum_j \theta_j$  in the optimal and random network economies. We see that allowing the shape of the network to adapt optimally to the  $z$  shocks leads to a substantial decline in aggregate fluctuations: the standard deviation of output is 1.5 percentage point larger in the economy with a random network.<sup>32</sup> The table also shows that this decline can be attributed almost entirely to lower fluctuations in  $Q$  and, therefore, to changes in the shape of the network.

The intuition behind this decline in fluctuations is as follows. In the random network allocation, the shape of the network is fixed at a given  $\theta^{\text{rand}}$  such that all fluctuations are coming from productivity  $z$ . The network  $\theta^{\text{rand}}$  might turn out to be close to what the optimal network would be for some  $z$ . In these cases, output will reach a relatively large value. For many other  $z$ 's, however,  $\theta^{\text{rand}}$  is far from the optimal network such that output is then low. As a result, the standard deviation of output is quite large in the economy with a random network. In contrast, in the optimal network allocation, the planner compares *all the  $2^n$  possible networks  $\theta$*  and selects the one that provides the maximum output for each particular vector  $z$ . This selection process provides substantial smoothing on fluctuations.<sup>33,34</sup>

<sup>32</sup>Unsurprisingly, the planner can also reach a higher level of output by adjusting the shape of the network in response to shocks. The average difference between the optimal network and the random network is about 31%.

<sup>33</sup>This intuition is made formal by extreme value theory. For instance, the variance of the maximum  $Y_m$  of a large number  $m$  of independent standard normal random variables decreases at a rate of  $\ln m$  for  $m$  large.

<sup>34</sup>The decline in fluctuations generated by the optimal organization of the network depends on various features of the economy. For instance, in an economy with lower elasticity of substitution between inputs  $\epsilon$ , the value of a connection is larger and the planner therefore creates a tightly connected network of firms that changes little in response to productivity shocks. The change in the standard deviation of output between the optimal and random

	Output $Y$	Labor Prod. $Q$	Prod. labor $L - f \sum_j \theta_j$
Optimal network	0.039	0.039	0.0014
Random network	0.054	0.054	0

Notes: All variables are in log. In the random network economy, the standard deviation of productive labor is 0 by construction.

Table 9: Standard deviation of aggregates in the optimal and random network economies

## 5 Conclusion

This paper proposes a theory of endogenous network formation and aggregate fluctuations. Because of a complementarity between firms’ operating decisions, production tends to be organized in tightly connected clusters of producers centered around the most productive firms. Long cascades of firm shutdowns can be triggered by a severe shock to an important producer. The theory also predicts how the shape of the network evolves with aggregate output. In particular, expansions feature more clustering among firms as well more highly connected producers. I find that these correlations are also present in the U.S. data. Finally, the theory predicts that the optimal reorganization of the network in response to shocks is responsible for a substantial decline in the size of aggregate fluctuations. These findings highlight the importance of considering the evolution of the production network to further understand the origin of aggregate fluctuations.

The paper also provides a novel reshaping approach to solve optimization problems with discrete adjustment margins and complementarities. Using this approach, it is possible to quickly find the optimal allocation in economies that were previously impossible to solve. The approach could potentially be useful in a broad range of economic environments, including models with menu costs, or fixed costs in investment or in hiring.

For the sake of clarity and tractability, the mechanisms of the model were exposited in a simple framework but several extensions may be worth investigating. In particular, it would be interesting to understand how requiring the payment of a fixed cost to *change* the operating status of a firm would affect the main mechanism of the model. In addition, a modified version of the model in which firms are financial institutions and links in the network are financial contracts might shed some light on the formation of financial networks and the propagation of financial distress.

## References

ACEMOGLU, D., V. M. CARVALHO, A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2012): “The Network Origins of Aggregate Fluctuations,” *Econometrica*, 80, 1977–2016.

---

network is therefore less important. Alternatively, an increase in the share of intermediate inputs into production  $\alpha$  increases the importance of the network in driving fluctuations and therefore yields a higher gap between the standard deviations of output in the two economies.

- ACEMOGLU, D., A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2015a): “Microeconomic Origins of Macroeconomic Tail Risks,” Working Paper 20865, National Bureau of Economic Research.
- (2015b): “Systemic Risk and Stability in Financial Networks,” *American Economic Review*, 105, 564–608.
- ATALAY, E. (2015): “How Important Are Sectoral Shocks,” Working paper.
- ATALAY, E., A. HORTAÇSU, J. ROBERTS, AND C. SYVERSON (2011): “Network structure of production,” *Proceedings of the National Academy of Sciences*, 108, 5199–5202.
- BAK, P., K. CHEN, M. WOODFORD, AND J. SCHEINKMAN (1993): “Aggregate Fluctuations from Independent Sectoral Shocks: Self-Organized Criticality in a Model of Production and Inventory Dynamics,” *Ricerche Economiche*, 47, 3–30.
- BAQAEE, D. R. (2016): “Cascading Failures in Production Networks,” working paper.
- BARTELSMAN, E., J. HALTIWANGER, AND S. SCARPETTA (2013): “Cross-Country Differences in Productivity: The Role of Allocation and Selection,” *American Economic Review*, 103, 305–34.
- BERNARD, A. B., A. MOXNES, AND Y. U. SAITO (2015): “Production Networks, Geography and Firm Performance,” Working Paper 21082, National Bureau of Economic Research.
- BIGIO, S. AND J. LA’O (2016): “Financial Frictions in Production Networks,” Working Paper 22212, National Bureau of Economic Research.
- BLUME, L., D. EASLEY, J. KLEINBERG, R. KLEINBERG, AND E. TARDOS (2011a): “Network formation in the presence of contagious risk,” in *Proceedings of the 12th ACM conference on Electronic commerce*, ACM, 1–10.
- BLUME, L., D. EASLEY, J. KLEINBERG, R. KLEINBERG, AND É. TARDOS (2011b): “Which networks are least susceptible to cascading failures?” in *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, IEEE, 393–402.
- BRESNAHAN, T. F. AND V. A. RAMEY (1994): “Output Fluctuations at the Plant Level,” *The Quarterly Journal of Economics*, 593–624.
- BRODA, C. AND D. E. WEINSTEIN (2006): “Globalization and the Gains From Variety,” *The Quarterly Journal of Economics*, 121, 541–585.
- CARVALHO, V. M., , M. NIREI, AND Y. SAITO (2014): “Supply Chain Disruptions: Evidence from the Great East Japan Earthquake,” Working Paper 14-E-035, The Research Institute of Economy, Trade and Industry.

- CARVALHO, V. M. (2014): “From Micro to Macro via Production Networks,” *Journal of Economic Perspectives*, 28, 23–48.
- CARVALHO, V. M. AND N. VOIGTLÄNDER (2014): “Input Diffusion and the Evolution of Production Networks,” Working Paper 20025, National Bureau of Economic Research.
- CATANZARO, M., M. BOGUÑÁ, AND R. PASTOR-SATORRAS (2005): “Generation of uncorrelated random scale-free networks,” *Phys. Rev. E*, 71, 027103.
- COHEN, L. AND A. FRAZZINI (2008): “Economic Links and Predictable Returns,” *The Journal of Finance*, 63, 1977–2011.
- DE SOLLA PRICE, D. J. (1965): “Networks of Scientific Papers,” *Science*, 149, 510–515.
- DUPOR, B. (1999): “Aggregation and irrelevance in multi-sector models,” *Journal of Monetary Economics*, 43, 391 – 409.
- ELLIOTT, M., B. GOLUB, AND M. O. JACKSON (2014): “Financial Networks and Contagion,” *American Economic Review*, 104, 3115–53.
- GABAIX, X. (2009): “Power Laws in Economics and Finance,” *Annual Review of Economics*, 1, 255–294.
- GABAIX, X. AND R. IBRAGIMOV (2011): “Rank -  $1/2$ : A Simple Way to Improve the OLS Estimation of Tail Exponents,” *Journal of Business & Economic Statistics*, 29, 24–39.
- GAREY, M. R. AND D. S. JOHNSON (1990): *Computers and Intractability; A Guide to the Theory of NP-Completeness*, New York, NY, USA: W. H. Freeman & Co.
- GEEROLF, F. (2016): “A Theory of Pareto Distributions,” working paper.
- HALL, G. J. (2000): “Non-convex costs and capital utilization: A study of production scheduling at automobile assembly plants,” *Journal of Monetary Economics*, 45, 681–716.
- HENDRICKS, K. B. AND V. R. SINGHAL (2005): “An Empirical Analysis of the Effect of Supply Chain Disruptions on Long-Run Stock Price Performance and Equity Risk of the Firm,” *Production and Operations Management*, 14, 35–52.
- HORVATH, M. (1998): “Cyclicality and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks,” *Review of Economic Dynamics*, 1, 781 – 808.
- JONES, C. I. (2011): “Intermediate Goods and Weak Links in the Theory of Economic Development,” *American Economic Journal: Macroeconomics*, 3, 1–28.



- JORGENSEN, D. W., F. GOLLOP, AND B. FRAUMENI (1987): *Productivity and U.S. Economic Growth*, Harvard University Press, reprinted- Universe, 1999.
- KARP, R. M. (1972): *Reducibility among Combinatorial Problems*, Boston, MA: Springer US, 85–103.
- KELLY, B., H. LUSTIG, AND S. V. NIEUWERBURGH (2013): “Firm Volatility in Granular Networks,” working paper.
- KENNAN, J. (2001): “Uniqueness of Positive Fixed Points for Increasing Concave Functions on  $R^n$ : An Elementary Result,” *Review of Economic Dynamics*, 4, 893 – 899.
- LIM, K. (2016): “Firm-to-firm Trade in Sticky Production Networks,” working paper.
- LONG, J. B. AND C. I. PLOSSER (1983): “Real Business Cycles,” *Journal of Political Economy*, 91, 39–69.
- OBERFIELD, E. (2013): “Business Networks, Production Chains, and Productivity: A Theory of Input-Output Architecture,” *Working Paper*.
- SAITO, Y. U., T. WATANABE, AND M. IWAMURA (2007): “Do larger firms have more interfirm relationships?” *Physica A: Statistical Mechanics and its Applications*, 383, 158 – 163.
- SIMON, H. A. (1955): “On a Class of Skew Distribution Functions,” *Biometrika*, 42, 425–440.
- WAGNER, S. M. AND C. BODE (2008): “AN EMPIRICAL EXAMINATION OF SUPPLY CHAIN PERFORMANCE ALONG SEVERAL DIMENSIONS OF RISK,” *Journal of Business Logistics*, 29, 307–325.
- WU, D. A. (2016): “Shock Spillover and Financial Response in Supply Chain Networks: Evidence from Firm-Level Data,” working paper.
- ZURICH INSURANCE GROUP (2015): “Potential business impact of loss of the main supplier for small and medium enterprises (SMEs) in 2015,” .

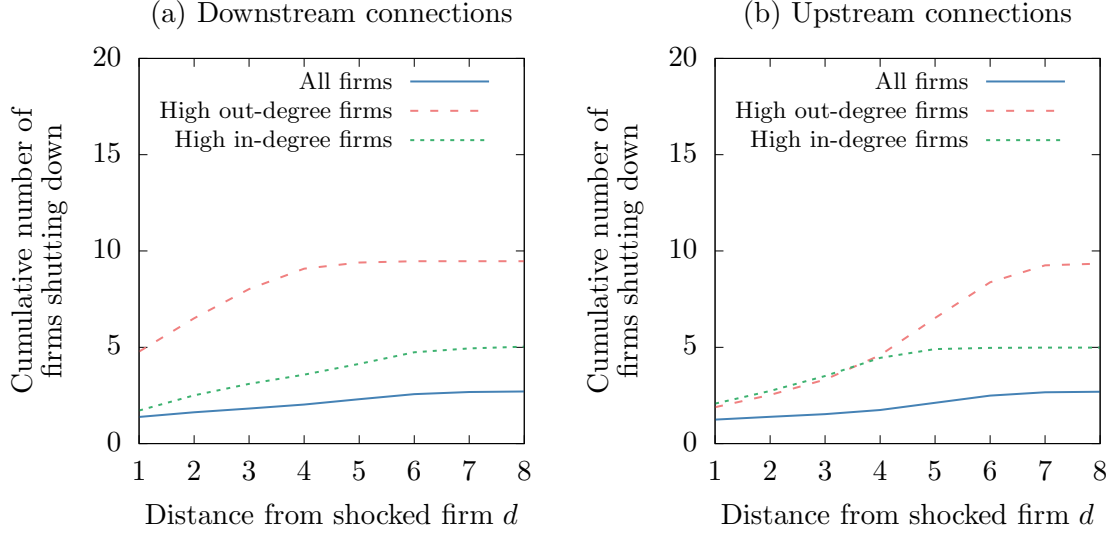
## A Cascades under different parameters

This appendix provides additional figures and tables to understand how cascades propagate in economies with different parameters. See 4.5 for a discussion.

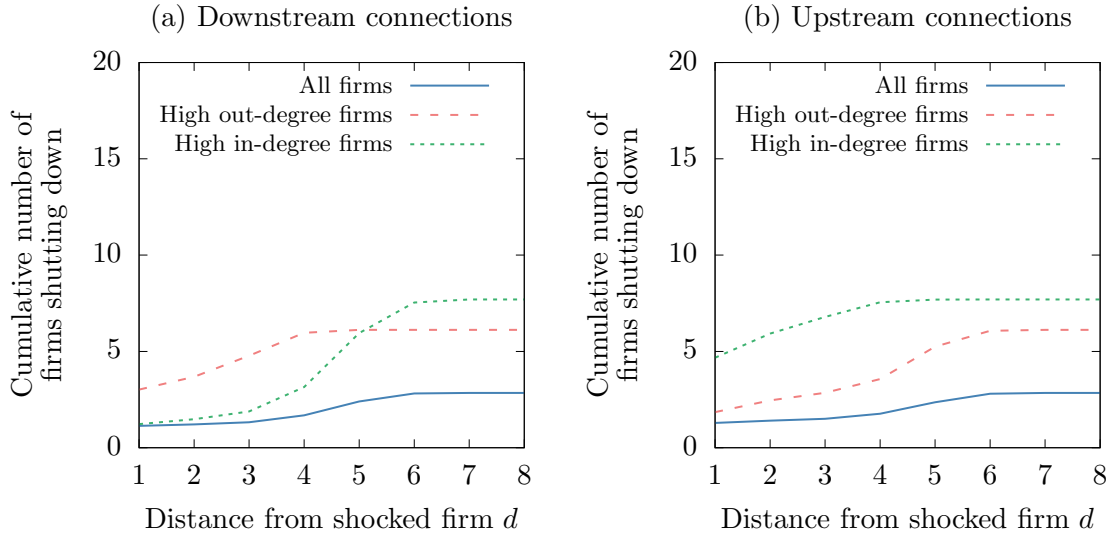
	Probability of firm shutdown		
	Benchmark	$\alpha = 0.75$	$\epsilon = 3$
All firms	92%	82%	32%
High out-degree firms	20%	8%	0%
High in-degree firms	56%	19%	15%

Notes: “High in-degree firms” refers to the 99th percentile of the in-degree distribution and similarly for out-degree firms. Firms are subject to a 1 standard deviation  $z$  shock.

Table 10: Resilience of targeted firms to shock with alternative parameter values



(a)  $\alpha = 0.75$



(b)  $\alpha = 0.75$

Notes: The distance  $d$  measures the shortest path, in terms of numbers of connections, from the targeted firm. “High in-degree firms” refers to the 99th percentile of the in-degree distribution and similarly for out-degree firms.

Figure 8: Cascades of firm shutdowns with alternative parameter values

## B Proofs

This section contains the proofs of the lemmas and propositions stated in the main text of the paper.

### B.1 Determination and Uniqueness of $q$

An intermediate step in the proof of the uniqueness of  $q$  uses the following definition of R-concavity.

**Definition.** (Kennan, 2001) A function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is radially quasiconcave (“R-concave”) if  $g(x) = 0$  and  $x > 0$  and  $0 \leq \lambda \leq 1$  implies  $g(\lambda x) \geq 0$ . If (in addition)  $0 < \lambda < 1$  implies  $g(\lambda x) > 0$ , then  $g$  is strictly R-concave.

In addition, we also need the concept of a quasi-increasing function.

**Definition.** (Kennan, 2001) A function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is quasi-increasing if  $y_i = x_i$  and  $y_j \geq x_j$  for all  $j$  implies  $g_i(y) \geq g_i(x)$ .

The following Lemma is used as an intermediate step to prove the uniqueness of the vector  $q$ .

**Lemma 3.** Consider the mapping  $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$  defined as

$$g_j(p) = (z_j A)^\epsilon \left( \sum_{i=1}^n \Omega_{ij} p_i^{\frac{\epsilon-1}{\epsilon}} \right)^{\alpha \frac{\epsilon}{\epsilon-1}} - p_j.$$

Then  $f$  is strictly R-concave.

*Proof.* Suppose that there exists a  $p^* > 0$  such that  $g(p^*) = 0$ . Then, for  $0 \leq \lambda \leq 1$ ,

$$\begin{aligned} g_j(\lambda p^*) &= (z_j A)^\epsilon \lambda^\alpha \left( \sum_{i=1}^n \Omega_{ij} (p_i^*)^{\frac{\epsilon-1}{\epsilon}} \right)^{\alpha \frac{\epsilon}{\epsilon-1}} - \lambda p_j^* \\ &\geq (z_j A)^\epsilon \lambda \left( \sum_{i=1}^n \Omega_{ij} (p_i^*)^{\frac{\epsilon-1}{\epsilon}} \right)^{\alpha \frac{\epsilon}{\epsilon-1}} - \lambda p_j^* \\ &\geq \lambda g_j(p^*) \\ &\geq 0 \end{aligned}$$

with strict inequality for  $0 < \lambda < 1$  since  $0 < \alpha < 1$  and  $\sum_{i=1}^n \Omega_{ij} \geq 1$ , which completes the proof.  $\square$

**Proposition 1.** The Lagrange multipliers associated with the efficient allocation satisfy

$$q_j = z_j \theta_j A \left( \sum_{i=1}^n \Omega_{ij} q_i^{\epsilon-1} \right)^{\frac{\alpha}{\epsilon-1}}, \quad \forall j \in \{1, \dots, n\} \quad (7)$$

where  $q_j = w/\lambda_j$ . Furthermore, there is a unique vector  $q$  that satisfies (7) such that  $q_j > 0$  if firm  $j$  is active and part of a strongly connected set of active firms.

*Proof.* The first-order conditions with respect to  $l_j$  and  $x_{ij}$  are

$$\begin{aligned} \lambda_j (1 - \alpha) y_j &= w l_j \\ \lambda_j \alpha \theta_j \frac{A z_j}{\alpha^\alpha (1 - \alpha)^{1-\alpha}} \left( \sum_{i=1}^n (\Omega_{ij} x_{ij})^{\frac{\epsilon-1}{\epsilon}} \right)^{\alpha \frac{\epsilon}{\epsilon-1} - 1} l_j^{1-\alpha} \Omega_{ij} x_{ij}^{\frac{\epsilon-1}{\epsilon}} &= \lambda_i x_{ij}. \end{aligned} \quad (13)$$

Combining these first-order conditions with the production function yields

$$x_{ij} \lambda_i^\epsilon = \lambda_j^\epsilon \alpha \left( z_j \theta_j A \left( \frac{\lambda_j}{w} \right)^{1-\alpha} \right)^{\frac{\epsilon-1}{\alpha}} y_j. \quad (14)$$

Plugging this equation and the first-order condition for  $l_j$  back into the production function yields (7).

To prove the uniqueness of  $q$ , I rely on a result by Kennan (2001). Consider the change of variable  $p_j = q_j^\epsilon$  and, since  $q_j = 0$  if  $\theta_j = 0$ , let us only consider the  $j$ 's such that  $\theta_j > 0$ . The recursive equation becomes

$$p_j = (z_j A)^\epsilon \left( \sum_{i=1}^n \Omega_{ij} p_i^{\frac{\epsilon-1}{\epsilon}} \right)^{\alpha \frac{\epsilon}{\epsilon-1}}. \quad (15)$$

Denote the right-hand side of that equation by  $f_j(p)$  and define the function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$  as  $g(p) = f(p) - p$ . Then, from Lemma 3, we now that  $g$  is strictly R-concave. Note also that  $g$  is quasi-increasing.

Consider the mapping  $h : \mathbb{R} \rightarrow \mathbb{R}^n$  defined as  $h(s) = f(\mathbb{1}s)$  where  $\mathbb{1}$  is a vector full of 1. Then  $h(s)$  is strictly concave, strictly increasing and differentiable with  $h(0) = 0$ ,  $\lim_{s \rightarrow 0} h'(s) = \infty$  and  $\lim_{s \rightarrow \infty} h'(s) = 0$ . As a result, there exist  $b > a > 0$  such that  $h(a) > a$  and  $h(b) < b$ . Then Theorem 3.1 and Theorem 3.2 of Kennan (2001) apply and 15 has a unique positive fixed point  $p^*$ . There is therefore a unique positive  $q^*$  that satisfies (7). It is such that  $q_j^* = \left(p_j^*\right)^{\frac{1}{\epsilon}}$  for  $\theta_j > 0$  and  $j$  is part of a strongly connected set,<sup>35</sup> and  $q_j^* = 0$  otherwise. Note that the proof goes through if we include the reshaping constant in (7).  $\square$

---

<sup>35</sup>This assumption is necessary to be sure that the firm has access to at least one input.

## B.2 Efficient Allocation with Exogenous Firm Status

**Lemma 1.** Aggregate output in the optimal allocation is given by

$$Y = Q \left( L - f \sum_{j=1}^n \theta_j \right). \quad (8)$$

where  $Q \equiv \left( \sum_{j=1}^n q_j^{\sigma-1} \right)^{\frac{1}{\sigma-1}}$ .

*Proof.* The first-order condition on  $y_j^0$  yields

$$y_j^0 = \left( \frac{q_j}{w} \right)^\sigma Y.$$

Raising both side of this equation to the power  $\frac{\sigma-1}{\sigma}$  and summing across firm we find that  $w = Q$ , so that the aggregate labor productivity is equal to the shadow value of labor. Using the first-order conditions (13) and (14) into the resource constraints (2), we find

$$\begin{aligned} y_j^0 + \sum_{k=1}^n x_{jk} - y_j &\leq 0 \\ y_j^0 + \sum_{k=1}^n \Omega_{jk} \left( \frac{q_j}{q_k} \right)^\epsilon \alpha \left( z_k \theta_k A \left( \frac{1}{q_k} \right)^{1-\alpha} \right)^{\frac{\epsilon-1}{\alpha}} \frac{q_k l_k}{1-\alpha} - \frac{q_j l_j}{1-\alpha} &\leq 0 \\ y_j^0 + \sum_{k=1}^n \Omega_{jk} q_j^\epsilon \frac{\alpha}{1-\alpha} \left( \frac{z_k A \theta_k}{q_k} \right)^{\frac{\epsilon-1}{\alpha}} l_k - \frac{q_j}{1-\alpha} l_j &\leq 0 \end{aligned}$$

which, taking the other variables as fixed, is a linear equation in  $l$ . Combining with the first-order condition on  $y_j^0$  we find

$$\left( \frac{q_j}{Q} \right)^{\sigma-1} \frac{Y}{Q} = \frac{1}{1-\alpha} \left[ l_j - \sum_{k=1}^n \Omega_{jk} q_j^{\epsilon-1} \alpha \left( \frac{z_k A \theta_k}{q_k} \right)^{\frac{\epsilon-1}{\alpha}} l_k \right]. \quad (16)$$

Define the  $n \times n$  matrix  $\Gamma$  with elements

$$\Gamma_{jk} = \Omega_{jk} q_j^{\epsilon-1} \left( \frac{z_k A \theta_k}{q_k} \right)^{\frac{\epsilon-1}{\alpha}}.$$

Using (7) we can write  $\Gamma_{jk}$  as

$$\begin{aligned}\Gamma_{jk} &= \Omega_{jk}(q_j)^{\epsilon-1} \left( \frac{z_k A \theta_k}{\theta_k A z_k \left( \sum_{i=1}^n \Omega_{ik} q_i^{\epsilon-1} \right)^{\frac{\alpha}{\epsilon-1}}} \right)^{\frac{\epsilon-1}{\alpha}} \\ &= \frac{\Omega_{jk} q_j^{\epsilon-1}}{\sum_{i=1}^n \Omega_{ik} q_i^{\epsilon-1}}.\end{aligned}$$

Summing (16) across firms, we get

$$\begin{aligned}\frac{Y}{Q} &= \frac{1}{1-\alpha} \left[ \left( L - f \sum_{j=1}^n \theta_j \right) - \alpha \sum_{k=1}^n \frac{\sum_{j=1}^n \Omega_{jk} q_j^{\epsilon-1}}{\sum_{i=1}^n \Omega_{ik} q_i^{\epsilon-1}} l_k \right] \\ &= \left( L - f \sum_{j=1}^n \theta_j \right).\end{aligned}$$

where I used the resource constraint on labor, which completes the proof.

We can derive all the relevant quantities in the allocation by using (16). Beginning by rewriting the equation in matrix form

$$\left( \frac{q}{Q} \right)^{\circ(\sigma-1)} \frac{Y}{Q} = \frac{1}{1-\alpha} [I_n - \alpha \Gamma] l.$$

Inverting this equation<sup>36</sup>, we find the labor allocation

$$l = (1-\alpha) \frac{Y}{Q} [I_n - \alpha \Gamma]^{-1} \left( \frac{q}{Q} \right)^{\circ(\sigma-1)}.$$

Returning to the first-order condition on labor we can find the output of each firm

$$\begin{aligned}y_j &= \frac{1}{1-\alpha} l_j q_j \\ &= \frac{1}{1-\alpha} (1-\alpha) Y [I_n - \alpha \Gamma]^{-1} \left( \frac{q}{Q} \right)^{\circ\sigma}.\end{aligned}$$

and (14) we find the amount of goods transiting between any two firms

$$x_{ij} = \left( \frac{q_i}{q_j} \right)^{\epsilon} \alpha \left( z_j \theta_j A \left( \frac{1}{q_j} \right)^{1-\alpha} \right)^{\frac{\epsilon-1}{\alpha}} y_j.$$

---

<sup>36</sup>If no self-connection are allowed,  $\Omega_{jj} = 0$  for all  $j$ , then  $[I_n - \alpha \Gamma]^T$  is strictly diagonally dominant (since the columns of  $\Gamma$  sum to 1 and  $0 < \alpha < 1$ ) and it is therefore invertible. Since the transpose of an invertible matrix is invertible,  $[I_n - \alpha \Gamma]$  is also invertible. In the numerical exercises,  $I_n - \alpha \Gamma$  is always invertible.

□

### B.3 Reshaping approach

#### Preliminary results

The second concavity proof uses the following lemmas.

**Lemma 4.** *Let  $F = A + \varepsilon B - \varepsilon C$  where  $\varepsilon > 0$ ,  $A$  is the  $n \times n$  matrix with every elements equal to 1,  $B$  is positive definite and  $C$  is an  $n \times n$  matrix with coefficients  $c_{ij} \geq 0$ . Let  $P$  denote an orthogonal projection onto the subspace  $S : \sum_{i=1}^n x_i = 0$ . If  $C$  is negative semi-definite on  $S$  then  $F$  is positive semi-definite for  $\varepsilon > 0$  small enough.*

*Proof.* By the Min-Max theorem the positive definiteness of  $B$  implies that  $z'Bz \geq c\|z\|^2$  for all  $z$  and  $c > 0$ . Take a vector  $z = x + y$  where  $x \in S$  and  $y \perp S$ . Then, for  $\varepsilon$  small enough,

$$\begin{aligned} z'(A + \varepsilon B - \varepsilon C)z &= z'Az + \varepsilon z'Bz - \varepsilon z'Cz = n\|y\|^2 + \varepsilon c\|z\|^2 - \varepsilon x'Cx - \varepsilon y'Cy - 2\varepsilon y'Cx \\ &\geq (n - 1/2)\|y\|^2 + c\varepsilon(\|x\|^2 + \|y\|^2) - 2\varepsilon\|C\|\|x\|\|y\| \\ &\geq 0. \end{aligned}$$

which is the result. □

**Lemma 5.** *Let  $X \subset \mathbb{R}^n$  be a nonempty convex set such that  $\text{int}(X)$  is also nonempty and let  $f : X \rightarrow \mathbb{R}$  be a continuous function. If  $f$  is convex on  $\text{int}(X)$  then  $f$  is convex on  $X$ .*

*Proof.* Since  $\text{int}(X) \neq \emptyset$ , then  $\text{cl}(\text{int}(X)) = \text{cl}(X)$ . Therefore, for all  $x \in X$  there exists a sequence  $x_n \rightarrow x$  with  $x_n \in \text{int}(X)$ . Take two points  $x, y \in X$  and  $x_n \rightarrow x$ ,  $y_n \rightarrow y$  with  $x_n \in \text{int}(X)$  and  $y_n \in \text{int}(X)$ . For any  $\lambda \in [0, 1]$ , by the convexity of  $f$  on  $\text{int}(X)$  we have

$$f(\lambda x_n + (1 - \lambda)y_n) \leq \lambda f(x_n) + (1 - \lambda)f(y_n).$$

Therefore, by continuity of  $f$  on  $X$ ,

$$\begin{aligned} f(\lambda x + (1 - \lambda)y) &= \lim_{n \rightarrow \infty} f(\lambda x_n + (1 - \lambda)y_n) \\ &\leq \lim_{n \rightarrow \infty} \lambda f(x_n) + (1 - \lambda)f(y_n) = \lambda f(x) + (1 - \lambda)f(y) \end{aligned}$$

so that  $f$  is convex on  $X$ . □

#### Proofs of concavity

**Proposition 2.** If the network of potential connections is such that  $\Omega_{ij} = c_i d_j$  for some vectors  $c, d \in \mathbb{R}^n$ , then the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a



solution to  $\mathcal{P}_{RR}$ .

*Proof.* The proof first finds conditions under which  $V_{RR}$  is log-concave. Raising both sides of (20) to the power  $\epsilon - 1$ , multiplying them by  $c_j \theta_j^b$  and summing across firms we find that

$$\sum_{j=1}^n c_j \theta_j^b (q_j(\theta))^{\epsilon-1} = \left( \sum_{j=1}^n c_j d_j^\alpha (Az_j)^{\epsilon-1} \theta_j^{a(\epsilon-1)+b} \right)^{\frac{1}{1-\alpha}}$$

so that

$$q_j(\theta) = z_j \theta_j^a A d_j^\alpha \left( \sum_{i=1}^n c_i d_i^\alpha (Az_i)^{\epsilon-1} \theta_i^{a(\epsilon-1)+b} \right)^{\frac{\alpha}{1-\alpha} \frac{1}{\epsilon-1}}.$$

Using the expression for  $V_{RR}(\theta)$ , we find that

$$\begin{aligned} \log(V_{RR}(\theta)) &= \frac{1}{\sigma-1} \log \left( \left( \sum_{j=1}^n (z_j \theta_j^a A d_j^\alpha)^{\sigma-1} \right) \left( \sum_{i=1}^n c_i d_i^\alpha (Az_i)^{\epsilon-1} \theta_i^{a(\epsilon-1)+b} \right)^{\frac{\alpha}{1-\alpha} \frac{\sigma-1}{\epsilon-1}} \right) \\ &\quad + \log \left( L - f \sum_{j=1}^n \theta_j \right) \\ &= \frac{1}{\sigma-1} \log \left( \sum_{j=1}^n (z_j \theta_j^a A d_j^\alpha)^{\sigma-1} \right) + \frac{1}{\epsilon-1} \frac{\alpha}{1-\alpha} \log \left( \sum_{i=1}^n c_i d_i^\alpha (Az_i)^{\epsilon-1} \theta_i^{a(\epsilon-1)+b} \right) \\ &\quad + \log \left( L - f \sum_{j=1}^n \theta_j \right). \end{aligned}$$

If  $0 < a \leq (\sigma-1)^{-1}$ ,  $-a(\epsilon-1) \leq b \leq 1-a(\epsilon-1)$  (and in particular if  $a$  and  $b$  satisfy  $(\star)$ ) the exponents on  $\theta$  are all between 0 and 1 so that the first two summations are concave functions of  $\theta$ . The last summation is linear in  $\theta$  so it is concave. Since the log of a concave function is concave, we have that  $\log(V_{RR})$  is also a concave function of  $\theta$ . Since, in addition, the constraint set  $\theta \in [0, 1]^n$  is convex and the Slater's qualification condition is obviously satisfied, the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize an optimal solution to the maximization of  $\log(V_{RR}(\theta))$  on the set  $\theta \in [0, 1]^n$ . Since log is a monotone transformation, a solution to this problem also solves  $\mathcal{P}_{RR}$ .  $\square$

**Proposition 3.** Let  $\sigma = \epsilon$  and suppose that the fixed cost  $f > 0$  and the dispersion in productivity  $\bar{z} - \underline{z} > 0$  are not too big. If the network  $\Omega$  is sufficiently connected with  $\Omega$  invertible, then the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution to  $\mathcal{P}_{RR}$ .

*Proof.* Define  $p_j = q_j^{\sigma-1}$  and

$$g^j = \frac{p_j}{z_j^{\sigma-1} (\sum_i \Omega_{ij} p_i)^\alpha}.$$

Using this notation, the optimization problem  $\mathcal{P}_{RR}$  can be written as

$$\min_{p \in P} -\frac{1}{\sigma-1} \log \left( \sum_j p_j \right) - \log \left( L - f \sum_{j=1}^n g^j(p) \right)$$

where the set  $P$  is defined as the set of  $p$  such that

$$\begin{aligned} -p_j &\leq 0 \\ p_j &\leq z_j^{\sigma-1} \left( \sum_i \Omega_{ij} p_i \right)^\alpha \end{aligned}$$

for each  $j \in \{1, \dots, n\}$ . Notice that  $P$  is convex and that Slater's condition is satisfied.

We first show that the objective function, which we denote by  $\Lambda(p)$ , is convex on the interior of  $P$ . The hessian of  $\Lambda$  is

$$\frac{\partial^2 \Lambda}{\partial p_k \partial p_l} = \frac{1}{\epsilon-1} \frac{1}{\left( \sum_j p_j \right)^2} + f \frac{\sum_j g_{kl}^j(p)}{L - f \sum_j g^j(p)} + f^2 \frac{\left( \sum_j g_k^j(p) \right) \left( \sum_j g_l^j(p) \right)}{\left( L - f \sum_j g^j(p) \right)^2}.$$

The last term can clearly be expressed as a Gramian matrix and its contribution to the hessian is therefore positive semidefinite. Using the derivatives of  $g$  explicitly, the first two first terms become

$$\frac{1}{\epsilon-1} \frac{1}{\left( \sum_j p_j \right)^2} + f \frac{1}{L - f \sum_j g^j(p)} \left( \sum_j \alpha(\alpha+1) \frac{p_j \Omega_{kj} \Omega_{lj}}{z_j^{\sigma-1} \left( \sum_i \Omega_{ij} p_i \right)^{\alpha+2}} - \alpha \frac{\Omega_{lk}}{z_k^{\sigma-1} \left( \sum_i \Omega_{ik} p_i \right)^{\alpha+1}} - \alpha \frac{\Omega_{kl}}{z_l^{\sigma-1} \left( \sum_i \Omega_{il} p_i \right)^{\alpha+1}} \right) \quad (17)$$

The first term in the parenthesis can be written as a Gramian matrix of the scalar product of pairs of vectors  $v_k$  and  $v_l$  with

$$v_m = \left[ \sqrt{\frac{\alpha(\alpha+1)p_1}{z_1^{\sigma-1} \left( \sum_i \Omega_{i1} p_i \right)^{\alpha+2}}} \Omega_{m1} \quad \dots \quad \sqrt{\frac{\alpha(\alpha+1)p_j}{z_j^{\sigma-1} \left( \sum_i \Omega_{ij} p_i \right)^{\alpha+2}}} \Omega_{mj} \quad \dots \quad \sqrt{\frac{\alpha(\alpha+1)p_n}{z_n^{\sigma-1} \left( \sum_i \Omega_{in} p_i \right)^{\alpha+2}}} \Omega_{mn} \right]'.$$

Consider the matrix  $M$  that has the vectors  $v_1, \dots, v_n$  as columns. We can write

$$M = \text{diag} \left( \left[ \sqrt{\frac{\alpha(\alpha+1)p_1}{z_1^{\sigma-1} \left( \sum_i \Omega_{i1} p_i \right)^{\alpha+2}}} \quad \dots \quad \sqrt{\frac{\alpha(\alpha+1)p_j}{z_j^{\sigma-1} \left( \sum_i \Omega_{ij} p_i \right)^{\alpha+2}}} \quad \dots \quad \sqrt{\frac{\alpha(\alpha+1)p_n}{z_n^{\sigma-1} \left( \sum_i \Omega_{in} p_i \right)^{\alpha+2}}} \right] \right) \Omega.$$

Since  $\Omega$  is invertible, the determinant of  $M$  is non-zero.<sup>37</sup> The vectors  $v_m$  are therefore linearly independent and the contribution to the hessian of the first term in the parenthesis of (17) is positive definite.

---

<sup>37</sup>Note that since we are proving convexity on the interior of  $P$  only we have  $p_j > 0$  for all  $j$ .

Summarizing the results so far, the hessian  $H$  can be written as

$$H = A + f \frac{1}{L - f \sum_j g^j(p)} (B - C) + f^2 D$$

where all the elements of matrix  $A$  are the same,  $B$  is positive definite,  $D$  is positive semidefinite and  $C$  is the symmetric matrix with elements

$$c_{kl} = \alpha \frac{\Omega_{lk}}{z_k^{\sigma-1} (\sum_i \Omega_{ik} p_i)^{\alpha+1}} + \alpha \frac{\Omega_{kl}}{z_l^{\sigma-1} (\sum_i \Omega_{il} p_i)^{\alpha+1}}.$$

By Lemma 4,  $H$  is positive semidefinite, for  $f > 0$  small enough, if  $C$  is negative semidefinite on the subspace  $S : \sum_{i=1}^n x_i = 0$ . Denote the largest element of  $C$  by  $\bar{c}$ . Then

$$C = \bar{c}O - \tilde{C}$$

where  $O$  is a matrix such that  $o_{ij} = 1$  for all  $i, j$  and where  $\tilde{C}$  is a matrix with elements  $\tilde{c}_{ij} = \bar{c} - c_{ij} \geq 0$ . Since  $x \in S$   $x'Ox = 0$  and thus  $x'Cx = -x'\tilde{C}x$ . The matrix  $C$  is therefore negative semidefinite on  $S$  if  $\tilde{C}$  is positive semidefinite on  $S$ . Since  $\Omega_{ii} = 0$  for all  $i$ ,  $c_{ii} = 0$  for all  $i$ . Then,  $\tilde{C}$  is diagonally dominant if

$$\bar{c} \geq \sum_{l \neq k} \left( \bar{c} - \alpha \frac{\Omega_{lk}}{z_k^{\sigma-1} (\sum_i \Omega_{ik} p_i)^{\alpha+1}} - \alpha \frac{\Omega_{kl}}{z_l^{\sigma-1} (\sum_i \Omega_{il} p_i)^{\alpha+1}} \right). \quad (18)$$

for all  $k$ . This inequality is strictly satisfied when the network is sufficiently connected and when the dispersion in  $z$  is sufficiently small. Indeed, in the limit ( $\bar{z} - \underline{z} \rightarrow 0$  and  $\Omega_{ij} = 1$ )

$$\bar{c} - \alpha \frac{\Omega_{lk}}{z_k^{\sigma-1} (\sum_i \Omega_{ik} p_i)^{\alpha+1}} - \alpha \frac{\Omega_{kl}}{z_l^{\sigma-1} (\sum_i \Omega_{il} p_i)^{\alpha+1}} \rightarrow 0$$

for all  $l \neq k$  so that (18) is strictly satisfied in this case for each  $k$ .

There is therefore a positive measure of parameters  $\bar{z}$ ,  $\underline{z}$  and  $\Omega$ , close to the limit, such that  $\tilde{C}$  is positive semidefinite on  $S$  such that  $C$  is negative semidefinite on  $S$ . By Lemma 4, the hessian matrix  $H$  is therefore positive semidefinite for  $f$  sufficiently small and  $\Lambda$  is therefore convex on  $\text{int}(P)$ . Since  $\Lambda$  is continuous, it is also convex by Lemma 5. The Karush-Kuhn-Tucker conditions are therefore necessary and sufficient to characterize a solution to  $\mathcal{P}_{RR}$ .  $\square$

## Proof of equivalence of the solutions

Denote by  $V_{RR} : [0, 1]^n \rightarrow \mathbb{R}^+$  the objective function of  $\mathcal{P}_{RR}$  defined as

$$V_{RR}(\theta) = \left( \sum_{j=1}^n (q_j(\theta))^{\sigma-1} \right)^{\sigma-1} \left( L - f \sum_{j=1}^n \theta_j \right) \quad (19)$$

where  $q_j$  is implicitly defined as

$$q_j(\theta) = z_j \theta_j^a A \left( \sum_{i=1}^n \theta_i^b \Omega_{ij} (q_i(\theta))^{\epsilon-1} \right)^{\frac{\alpha}{\epsilon-1}}. \quad (20)$$

Similarly, denote by  $V_{SP}$  the objective function of  $\mathcal{P}_{SP}$ .

**Proposition 4.** If  $\theta^*$  solves  $\mathcal{P}_{RR}$  and that  $\theta_j^* \in \{0, 1\}$  for all  $j$ , then  $\theta^*$  also solves  $\mathcal{P}_{SP}$ .

*Proof.* From the assumptions of the proposition, assume that  $\theta^* \in \{0, 1\}^n$  solves  $\mathcal{P}_{RR}$ . By construction, the objective function  $V_{RR}$  of  $\mathcal{P}_{RR}$  and the objective function  $V_{SP}$  of  $\mathcal{P}_{SP}$  coincide over  $\theta \in \{0, 1\}^n$ . Therefore  $V_{RR}(\theta^*) = V_{SP}(\theta^*)$ . Since the feasible set of  $\mathcal{P}_{RR}$ ,  $\theta \in [0, 1]^n$ , contains the feasible set of  $\mathcal{P}_{SP}$ ,  $\theta \in \{0, 1\}^n$ , it must be that  $V_{SP}(\theta^*) \geq V_{SP}(\theta)$  for  $\theta \in \{0, 1\}^n$ , otherwise  $\theta^*$  would not be a solution to  $\mathcal{P}_{RR}$ . Since  $\theta^*$  maximizes  $V_{SP}$  on the set  $\theta \in \{0, 1\}^n$  it must be that  $\theta^*$  solves  $\mathcal{P}_{SP}$ , which completes the proof.  $\square$

**Lemma 2.** The first-order condition (11) can be written as

$$z_k^{\sigma-1} \theta_k^{a(\sigma-1)-1} F_1(Q, B_k) \left( L - f \sum_{j=1}^n \theta_j \right) - fQ + z_k^{\epsilon-1} \theta_k^{b-1+a(\epsilon-1)} \sum_{j=1}^n \beta_j \Omega_{kj} F_2(B_k, B_j) \frac{\partial q_j}{\partial B_j} = \bar{\mu}_k - \underline{\mu}_k \quad (12)$$

where  $F_1$  and  $F_2$  are functions that only depend on the aggregates  $B$ 's and  $Q$ .

*Proof.* Consider the term

$$\begin{aligned} \frac{\partial q_k}{\partial \theta_k} \frac{\partial Q}{\partial q_k} &= z_k a \theta_k^{a-1} A B_k^\alpha \times \left( \sum_{j=1}^n q_j^{\sigma-1} \right)^{\frac{1}{\sigma-1}-1} q_k^{\sigma-2} \\ &= z_k a \theta_k^{a-1} A B_k^\alpha \times Q^{2-\sigma} (z_k \theta_k^a A B_k^\alpha)^{\sigma-2} \\ &= z_k^{\sigma-1} \theta_k^{a(\sigma-1)-1} F_1(Q, B_k). \end{aligned}$$

Similarly, for the term in (11) that is inside the summation

$$\begin{aligned}
\frac{\partial B_j}{\partial \theta_k} + \frac{\partial q_k}{\partial \theta_k} \frac{\partial B_j}{\partial q_k} &= \frac{1}{\epsilon - 1} B_j^{2-\epsilon} b \theta_k^{b-1} \Omega_{kj} q_k^{\epsilon-1} + a z_k \theta_k^{a-1} A B_k^\alpha \times B_j^{2-\epsilon} \theta_k^b \Omega_{kj} q_k^{\epsilon-2} \\
&= B_j^{2-\epsilon} \theta_k^{b-1} \Omega_{kj} q_k^{\epsilon-1} \left( a + b \frac{1}{\epsilon - 1} \right) \\
&= z_k^{\epsilon-1} \theta_k^{b-1+a(\epsilon-1)} \Omega_{kj} B_j \left( \frac{A B_k^\alpha}{B_j} \right)^{\epsilon-1} \left( a + b \frac{1}{\epsilon - 1} \right) \\
&= z_k^{\epsilon-1} \theta_k^{b-1+a(\epsilon-1)} \Omega_{kj} F_2(B_k, B_j).
\end{aligned}$$

Substituting these two expressions in (11) yields the result.  $\square$

## C Algorithms

### C.1 Network Creation

#### Random network

For a given number of firms  $n$  and a threshold  $t$  a connection  $\Omega_{ij} = 1$  is created if  $U < t$  where  $U \sim \text{iid Uniform}([0, 1])$ . In practice,  $t$  is set to  $t = c/n$  such that on average a firm is connected to  $c$  other firms. If a firm has no connection, the network is discarded and a new one is drawn.

#### Scale Free Network

I use the algorithm of [Catanzaro et al. \(2005\)](#). The idea is to create a number of stubs, or half-connections, drawn from power law (PL) distributions on each node and then link the stubs together.

For each firm  $j \in \{1, \dots, n\}$  draw a number of in-connections  $c_j^{in} \sim \text{iid PL}(\beta)$  where  $\beta$  is the shape of the distribution. Similarly, draw a number of out-connections  $c_j^{out} \sim \text{iid PL}(\beta)$ . If  $\sum_j c_j^{in} \neq \sum_j c_j^{out}$ , repeat the draws. Stack the in-connections in a list  $list_{in}$  and the out-connections in another list  $list_{out}$ . For each element  $i \in list_{in}$ , randomly assign an element from  $j \in list_{out}$ . If  $\Omega_{ij} = 0$ , set  $\Omega_{ij} = 1$  and remove  $i$  and  $j$  from their respective list. If not, draw another  $j$  from  $list_{out}$ . Repeat the procedure until both lists are empty. If the lists are not empty and no new connections can be created, repeat the draws.

### C.2 Solving for the Efficient Allocation

#### Exhaustive search

This algorithm performs an exhaustive search of the  $2^n$  possible vectors  $\theta \in \{0, 1\}^n$ . For each  $\theta$ , iterate on equation 7 to find  $q$ . Then compute  $Q$  and the value of the objective function  $Y$  using

8. The algorithm then picks the  $\theta^*$  with the largest  $Y$ . This algorithm is guaranteed to find the global maximum of  $\mathcal{P}_{SP}$  but it is impossible to use for large  $n$ . In practice, it takes about 1 minute for  $n = 17$  to check the 131072 possible vectors  $\theta$  on a 4 GHz Intel Core i7 processor. This time roughly doubles for each additional firms.

### Iterating on the first-order conditions

A convenient way to solve for the optimal allocation using the deformation approach is to iterate on a modified version of the first-order conditions of the log of  $\mathcal{P}_{RR}$ . The algorithm is as follows:

1. Initialize the 0-th iteration with  $\Delta\mu_k^0 = \underline{\mu}_j^0 - \bar{\mu}_j^0 = -1$  for all  $k$ .
2. For step  $p$ :
  - (a) Using the complementary slackness condition set  $\theta_k^p = 1$  if  $\Delta\mu_k^p \leq 0$  and  $\theta_k^p = 0$  if  $\Delta\mu_k^p > 0$
  - (b) Iterate on (10) to find  $q^p$
  - (c) Find  $\frac{\lambda_k^p q_k^p}{\theta_k^p}$  by solving the following linear equation derived from the first-order conditions:

$$\frac{\left(z_k A B_k^{\frac{\alpha}{\epsilon-1}}\right)^{\sigma-1}}{\sum_{j=1}^n q_j^{\sigma-1}} + \alpha (z_k A)^{\epsilon-1} B_k^\alpha \sum_{j=1}^n \Omega_{kj} \frac{\theta_j}{B_j} \frac{q_j \lambda_j}{\theta_j} = \frac{\lambda_k q_k}{\theta_k}$$

where  $B_j = \left(\sum_{i=1}^n \theta_i^b \Omega_{ij} q_i^{\epsilon-1}\right)^{\frac{1}{\epsilon-1}}$  is the composite of intermediate inputs.

- (d) Compute  $\Delta\mu_k^{p+1}$  using the following equation derived from the first-order conditions

$$\frac{f}{L - f \sum_{j=1}^n \theta_j} = \frac{1}{\sigma-1} \frac{q_k \lambda_k}{\theta_k} + \frac{\alpha}{\epsilon-1} \left(1 - \frac{\epsilon-1}{\sigma-1}\right) (z_k A)^{\epsilon-1} B_k^\alpha \sum_{j=1}^n \Omega_{kj} \frac{\theta_j}{B_j} \frac{q_j \lambda_j}{\theta_j} + \Delta\mu_k.$$

- (e) Repeats steps (a) to (d) until convergence.