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

Mathieu Taschereau-Dumouchel†

Cornell University

June 2022

Abstract

This paper studies an economy in which firms are connected through input-output linkages and must pay a fixed cost to produce. When economic conditions are poor, some firms might decide not to operate, thereby severing the links with their neighbors and changing the structure of the production network. In the model, producers benefit from having access to additional suppliers, and so nearby firms tend to operate, or not, together. As a result, the production network features clusters of operating firms, and the exit of a producer can create a cascade of firm shutdowns. While well-connected firms are better able to withstand shocks, they trigger larger cascades upon exit. The theory also predicts how the structure of the production network changes over the business cycle. As in the data, recessions are associated with more dispersed networks that feature fewer highly connected firms. In the calibrated economy, the endogenous reorganization of the network substantially dampens the impact of idiosyncratic shocks on aggregate fluctuations.

JEL Classifications: E23, D85, C67

*We thank Enghin Atalay, Julieta Caunedo, Jeremy Greenwood, Veronica Guerrieri, Christopher Huckfeldt, Philipp Kircher, Ben Lester, Kristoffer Nimark, Ezra Oberfield, Nicolas Petrosky-Nadeau, Carlos Ramírez, Stephen Redding, Gill Segal, Ali Shourideh, Gianluca Violante, three anonymous referees, and seminar participants at Wharton, the Philadelphia Fed, Drexel, FGV São Paulo, the Cornell/Penn State workshop, UT Austin, SED Edinburgh, SAET Faro, the ITAM-PIER Conference, the Tepper-LAEF Conference, the CIRANO – Sam M. Walton College of Business Workshop on Networks, LSE, Cornell, the CMSG annual meeting, Rutgers, the New York Fed, the Midwest Macro Meetings, Minnesota Macro, USC Marshall, Queen's, the NYU Search Conference, Princeton, the 2019 AEA, the Atlanta Fed, Binghamton and HKUST for helpful comments. We gratefully acknowledge support from Cornell's Institute for the Social Sciences.

†Email: mt763@cornell.edu; Address: 480 Uris Hall, Cornell University, Ithaca NY 14853

1 Introduction

Production in modern economies involves a complex network of specialized firms, each using inputs from suppliers and providing their own output to downstream producers. Recent research has shown that the structure of this production network is an important determinant of economic outcomes. At the micro level, it influences the size of firms and how they withstand shocks (Barrot and Sauvagnat, 2016), while at the macro level it affects how idiosyncratic shocks contribute to aggregate fluctuations (Acemoglu et al., 2012). Yet relatively little is known about how economic forces shape the network structure itself. This paper proposes a theory to study how the production network is formed, how it responds to shocks, and how the response of the network affects the aggregation of idiosyncratic shocks into macroeconomic fluctuations.

One distinguishing feature of the theory is that it focuses on the firms' extensive margin of production as the key driver behind the formation of the network. Consider, for instance, a firm that goes out of business after facing a severe shock. Since it no longer supplies to customers or purchases from suppliers, the links with its previous neighbors are cut. Similarly, when a new firm begins production, new connections with customers and suppliers are created. This process plays an important role in shaping the production network in the data.[1] It is also responsible for creating cascades of firm shutdowns: a chain reaction through which a shock to a vulnerable firm can lead to the exit of many of its (perhaps removed) neighbors. Policymakers were worried about such cascades during the financial crisis and the model can shed light on their origin and propagation.

In the model, a finite number of firms produce differentiated goods using labor and inputs from other producers. Production requires the payment of a fixed cost, so that firms operate or not as a function of economic conditions. When a firm does operate, it makes an additional input available to all of its customers, and it purchases intermediate goods from its suppliers, thereby creating new input-output relationships. Together, the operating decisions of the firms therefore determine the structure of the production network. The main goal of the paper is to study the efficient allocation in this environment. That allocation provides a natural benchmark as it captures some of the main forces at work in the environment. We also describe how it can be decentralized as an equilibrium.

In the model, firms combine intermediate inputs from their suppliers using a standard CES production technology. As a result, having access to an additional input lowers the marginal cost of production and makes the firm effectively more productive. Because of these gains from input variety, firms with multiple suppliers are more likely to operate—their high productivity more than compensate for the fixed cost of operation. Similarly, firms with many customers provide a valuable input to multiple producers and are also more likely to operate. These forces create complementarities between the operating decisions of nearby firms: neighbors tend to operate, or

[1] According to Factset Revere, a large dataset of firm-level input-output linkages in the U.S., about 40% of link destructions occur when either the supplier or the customer (or both) stops producing. See Section 6.1 for details.

not, together.

These complementarities have important implications for the structure of the production network. First, they lead to the creation of clusters of firms that are tightly connected with one another. By organizing production in this way, firms increase their number of customers and suppliers so as to take full advantage of the gains from input variety. Second, cascades of firm failures can arise in the efficient allocation. If a firm faces a severe shock and stops production, its customers, having lost a valuable input, and its suppliers, now producing a less useful product, are also more likely to shut down. The same logic applies to the firm's second neighbors, which are more likely to shut down as well, and so on. As a result, the initial shock can trigger a cascade of firm shutdowns that propagates upstream and downstream through the production network.

One other, and perhaps unusual, consequence of the operating complementarities between neighbors is that they can lead to a large reorganization of the network after a small change in the environment. For instance, a small decline in the productivity of a centrally located firm can lead to the shutdown of its whole neighborhood, as economic activity moves to a more productive part of the network. Through that mechanism, large changes in the distribution of firm-level outcomes can occur in response to arbitrarily small shocks.

Two features of the environment make the problem of the social planner particularly challenging to solve. First, since the decision to operate a firm is binary, the planner's optimization problem has a non-convex feasible set. Second, the increasing returns to scale generated by the fixed costs break the concavity of the objective function. As a result, the objective function generally features multiple local maxima and standard algorithms are unable to identify the global maximum. We provide a novel solution approach that involves reshaping the original optimization problem such that 1) this reshaped problem can be solved easily, and 2) its solution coincides with that of the original problem. We establish sufficient conditions under which this approach is guaranteed to find the efficient allocation. But even when those conditions are not met, numerical simulations show that it provides a rapid and robust way of tackling a class of challenging network formation problems (Carvalho and Tahbaz-Salehi, 2018).

While this paper focuses on the efficient allocation, distortions such as market power might play a role in shaping the production network in reality. In particular, superstar firms such as Walmart and IKEA are believed to have a large amount of pricing power when dealing with their suppliers (Bloom and Perry, 2001). We incorporate this feature in the model and show that it leads to inefficient entry decisions by firms. The forces at work in that equilibrium are similar to those of the efficient allocation, but there are differences in how the complementarities between firms operate and, as a result, how cascades propagate. We show that the proposed solution method can also be used to solve for that inefficient equilibrium in a straightforward way.

We provide a basic calibration of the model using firm-level data for the United States economy. To better understand how economic forces shape the production network, we compare the efficient network, designed optimally by the planner, to a neutral benchmark whose structure is randomly determined. We find that the efficient network features indegree (number of suppliers) and outdegree (number of customers) distributions with thicker right tails, as well as a higher amount of clustering between firms. These differences show that the planner takes advantage of the operating complementarities by creating tightly connected clusters of economic activity centered around well-connected firms.

We use the calibrated economy to investigate how cascades of firm shutdowns arise and propagate through the network. As in the data, highly connected firms are more resilient to shocks but, upon shutting down, they create larger cascades that lead to the exit of several of their neighbors. We also investigate how cascades affect macroeconomic aggregates. While the average cascade has a negligible effect, a cascade that originates from a highly connected firm can have a substantial negative impact on output. Finally, we describe through sensitivity analysis how the parameters of the model influence how cascades propagate and their aggregate impact. In particular, cascades become more damaging when intermediate inputs are less substitutable.

The model also allows us to understand the relationship between the structure of the production network and the business cycle. Since the number of firms in the economy is finite, fluctuations in aggregate output emerge from the idiosyncratic shocks faced by the firms. As these shocks also affect the structure of the production network, the model features comovements between the macroeconomy and the structure of the network.

One contribution of this paper is to highlight novel business cycle correlations between aggregate output and the structure of the production network. We find that, in the data and in the model, recessions are periods in which, 1) the tails of the degree distributions are thinner, and 2) there is less clustering between firms. These correlations are naturally explained through the lens of the model. Expansions are periods in which it is easy to leverage the complementarities at work in the economy by creating productive clusters of firms around highly connected producers. In contrast, recessions are periods in which creating these clusters would be too costly, perhaps because a few influential firms are facing bad shocks, and in which production therefore involves a more diffused, and less productive, network.

We then consider how the endogenous formation of the network interacts with firm-level shocks to influence aggregate fluctuations. To do so, we compare the benchmark economy, in which the production network reorganizes itself in response to shocks, to an alternative economy in which the structure of the network is kept fixed. We find that aggregate output is on average 11% lower and 20% more volatile when the network is fixed. This last finding highlights the importance of considering how the production network adapts to shocks to better understand the microeconomic origin of aggregate fluctuations.

Finally, we compare the efficient allocation to the inefficient equilibrium to evaluate the quantitative impact of pricing distortions. While they are mostly similar, there are a few notable differences between the two allocations. For instance, the equilibrium production network is less correlated with GDP, suggesting that it is more rigid and less able to adapt to changing economic conditions. Cascades also tend to propagate more downstream than in the efficient allocation, as predicted by the theory.

Relation to the Literature

The theoretical model is motivated by an empirical literature documenting that losing a supplier is disruptive to a firm's operations. Carvalho et al. (2014) document that firms that stopped production because of the Great East Japan Earthquake of 2011 had a significant negative impact on their customers and suppliers. Hendricks and Singhal (2005) find that firms facing supply chains disturbances suffer from large and long-lasting negative abnormal stock returns. Wagner and Bode (2008) survey business executives in Germany who report that issues with supply chains, including the loss of a supplier, were responsible for significant disturbances to 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 shut down.

This paper also relates to a literature that studies how shocks to interconnected sectors contribute to aggregate fluctuations in exogenous networks (Long and Plosser, 1983; Horvath, 1998; Dupor, 1999). In an influential paper, 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.[2] Acemoglu et al. (2015) further show that inter-sectoral linkages can generate larger tail-risks in aggregate output. This literature emphasizes the importance of the (fixed) structure of the network in transmitting idiosyncratic shocks. In contrast, the current paper studies how endogenizing the network affects aggregate fluctuations.[3]

One of the first papers to study the macroeconomic impact of cascades is Baqaee (2018), which considers a model with an exogenous sectoral network in which the mass of firms 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 to that paper, the present work considers a discrete adjustment margin that leads to the creation and destruction of nodes and edges in the production network, something that we observe in the firm-level data. Another paper that emphasizes the role of discreteness is Elliott et al. (2020), which considers a supply network in which each link is

[2] In a related paper, Gabaix (2011) shows that when the tail of the firm size distribution is sufficiently thick, firm-level shocks can have large effects on aggregates.

[3] See Carvalho (2014) and Carvalho and Tahbaz-Salehi (2018) for an overview of the literature on production networks. Recent contributions include di Giovanni et al. (2014), Atalay (2017), Baqaee and Farhi (2017a), Baqaee and Farhi (2017b), Bigio and La'O (2016), Caliendo et al. (2017b), Caliendo et al. (2017a), Grassi (2017), Ozdagli and Weber (2017), Liu (2019) and Chahrour et al. (2019).

at risk of failure. This discrete margin can make the network fragile, in the sense that aggregate output becomes very sensitive to small shocks. Acemoglu and Tahbaz-Salehi (2020) also look at the impact of supply chain disruptions in a model with bargaining and endogenous markups.

This paper contributes to a recent literature in which production networks are built endogenously by the decisions of economic agents.[4,5] One of the first in that literature is Oberfield (2018) who builds a model in which producers optimally choose one input from a randomly evolving set of suppliers, thereby creating a production network. He finds that star suppliers can emerge endogenously in equilibrium. Lim (2018) studies sourcing decisions in a model with sticky relationships. Unlike the present work, these papers feature a continuum of firms so that aggregate fluctuations do not arise from individual idiosyncratic shocks, a margin whose importance has been emphasized by the granularity literature (Gabaix, 2011) but that has proven challenging to incorporate in network formation models (Carvalho and Tahbaz-Salehi, 2018).

Acemoglu and Azar (2018) consider a network of competitive industries in which firms select a production technique that involves different sets of suppliers. They show that the endogenous evolution of the network can generate long-run growth. Tintelnot et al. (2018) build a model of endogenous network formation and international trade. In contrast to the current paper, they only consider acyclic networks. Boehm and Oberfield (2018) estimate a model of network formation using Indian micro data to study misallocation in the inputs market. Kopytov et al. (2022) look at the impact of uncertainty on the structure of the production network.

One distinguishing feature of this work is that it proposes a macroeconomic model in which the input-output network is built endogenously through the entry and exit decisions of the firms—a margin that accounts for a large fraction of link destructions in U.S. data and that allows for cascades of firm shutdowns to arise.

This paper proposes a new solution technique for some nonconvex optimization problems with binary variables. Several heuristics have been developed to handle these problems (Li and Sun, 2006). Closest to the present work are smoothing algorithms that attempt to get rid of the local maxima that emerge in the relaxed problem (Murray and Ng, 2010). In practice, finding an appropriate smoother is usually done through trial and error and there is no guarantee that the algorithm converges to a global maximum. In contrast, the current work explicitly describes how to reshape the problem and proposes a rapid and robust solution method.

The next section introduces the model. Section 3 describes the solution method. Section 4 discusses equilibrium allocations. Section 5 explores the forces at work in the economy. Section 6 provides a calibration to U.S. data.

[4] A related literature studies networks that are mostly formed through an exogenous random process. See for instance Atalay et al. (2011), Carvalho and Voigtländer (2015), and König et al. (2018).

[5] Some papers with CES production also have endogenous networks in the sense that the input-output matrix is endogenous. In these models, and in contrast to those mentioned here, the input-output matrix move smoothly with fundamentals and never reach zero, so that connections are never created or destroyed.

2 A model of production networks with endogenous entry

There is a set $\mathcal{N} = \{1, \dots, n\}$ of firms, each of which produces a differentiated good that can be used as intermediate input by other firms or consumed by a representative household. The preferences of the household are given by the utility function

$$C = \left(\sum_{j=1}^{n} \beta_j^{\frac{1}{\sigma}} c_j^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

where c_j is consumption of good j, $\sigma > 1$ is the elasticity of substitution between goods and $\beta_j \geq 0$ determines the household's taste for good j. Throughout, we sometimes refer to C as aggregate output or GDP. The household also supplies $L > 0$ units of labor inelastically.

To produce, a firm j must employ $f_j L \geq 0$ units of labor as a fixed cost, in which case we say that j is *operating*. This fixed cost captures overhead labor, such as managers and other non-production workers, that is necessary for production.[6] The vector $\theta \in \{0, 1\}^n$ keeps track of the operating decisions of the firms, such that $\theta_j = 1$ if j operates and $\theta_j = 0$ otherwise.

When it operates, firm j has access to a technology that converts l_j units of labor and a vector of intermediate inputs $x_j = (x_{j1}, \dots, x_{jn})$ into y_j units of good j according to the production function

$$y_j = \frac{A}{\alpha_j^{\alpha_j} (1 - \alpha_j)^{1-\alpha_j}} z_j \theta_j \left(\sum_{i=1}^{n} \Omega_{ij}^{\frac{1}{\varepsilon_j}} x_{ij}^{\frac{\varepsilon_j - 1}{\varepsilon_j}} \right)^{\frac{\varepsilon_j}{\varepsilon_j - 1} \alpha_j} l_j^{1-\alpha_j}, \tag{2}$$

where $\Omega_{ij} \geq 0$ denotes the factor intensity of input i, $\varepsilon_j > 1$ is the elasticity of substitution between inputs, $0 < 1 - \alpha_j < 1$ is the labor intensity, and $A > 0$ and $z_j > 0$ are aggregate and firm-specific total factor productivities. Without loss of generality, we assume that each firm has access to at least one input, so that $\sum_i \Omega_{ij} > 0$ for all j, otherwise j cannot produce and we can redefine the economy without it.[7]

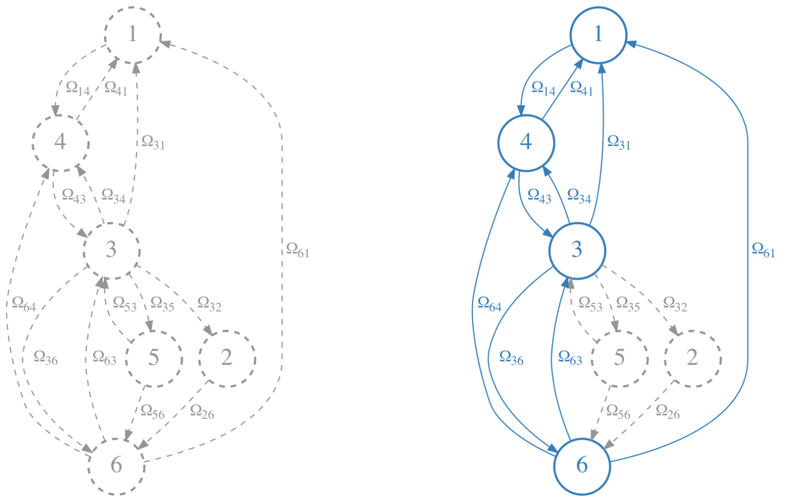
We see from (2) that a firm j can only use inputs from a supplier i if $\Omega_{ij} > 0$. As such, the matrix Ω describes a network of *potential connections* between firms. A potential connection (i, j) is *active*—with goods being traded—if firms i and j both operate, otherwise it is *inactive*. The production network is therefore jointly determined by Ω and θ, and economic conditions endogenously determine the structure of the network through their impact on operating decisions.

Panel (a) in Figure 1 provides an example of potential connections in an economy with six firms. Each arrow represents a connection $\Omega_{ij} > 0$, with the direction of the arrow showing the

[6] Empirical studies have found these fixed costs to be important to explain firm-level outcomes. For instance, Bresnahan and Ramey (1994) show that the extensive margin is responsible for about 80% of plant-level output fluctuations in the automobile industry.

[7] The restrictions $\sigma > 1$ and $\varepsilon > 1$ are necessary to avoid a complete shutdown of the economy, or of a customer, if a single producer does not operate. The term $\alpha_j^{\alpha_j} (1 - \alpha_j)^{1-\alpha_j}$ in (2) is a normalization to simplify some expressions.

potential movement of goods between the firms. The set of active connections, in blue in panel (b), is determined endogenously by the set of operating firms, also shown in blue.



(a) Potential connections

(b) Active firms and connections

Figure 1: The firms' operating decisions determine the production network

While we focus on the role of the firms' extensive margin of operation for the formation of the network, the model is general enough to accommodate the formation of individual links. Specifically, we can interpret a link between any two firms i and k, as a *pseudo* firm j that 1) only has potential connections in Ω with i as a supplier and k as a customer, and 2) produces a good that is not included in the production of the final good ($\beta_j = 0$). $\theta_j \in \{0, 1\}$ then indicates whether the link between i and k is active or not.[8]

It is useful to describe the set of firms that *can* produce under a given operating vector θ. For a firm to produce, it must receive some intermediate input from at least one supplier, and this supplier must also receive some input from a supplier, and so on. Since n is finite, this sequence of suppliers must contain a cycle for production to take place. As a result, a firm without access to such an operating cycle simply cannot produce.[9] We will use this fact later on to characterize an allocation under a given vector θ.

3 The efficient allocation and how to find it

To better understand some of the key forces at work in this environment, we first consider the problem of a social planner. In the next section, we discuss how the resulting efficient allocation

[8]In this context, j can be interpreted as a shipping technology whose cost f_j is the fixed cost of operating the link and its productivity z_j affects the variable cost of shipping goods through that link.

[9]Formally, an *operating cycle* is a sequence of operating firms $\{s_1, \ldots, s_k\}$, for some $k \geq 1$, such that 1) $\Omega_{s_i, s_{i+1}} > 0$ for all $i \in \{1, \ldots, k-1\}$, and 2) $\Omega_{s_k, s_1} > 0$. A firm s_j *has access to an operating cycle* if there exists a sequence of operating firms $\{s_1, \ldots, s_j\}$ such that 1) s_1 is part of an operating cycle, and 2) $\Omega_{s_i, s_{i+1}} > 0$ for all $i \in \{1, \ldots, j-1\}$.

can be decentralized as an equilibrium. We also consider an equilibrium in which market power in firm-to-firm transactions can lead to inefficiencies.

Consider the problem $\mathcal{P}$ of a social planner that maximizes the utility of the household

$$\mathcal{P}: \max_{\substack{c\geq 0,\, x\geq 0,\, l\geq 0\\ \theta\in\{0,1\}^n}} \left(\sum_{j=1}^n \beta_j^{\frac{1}{\sigma}} c_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \tag{3}$$

subject to a resource constraint for each intermediate good j,

$$c_j + \sum_{k=1}^n x_{jk} \leq y_j, \tag{4}$$

where y_j is given by (2), and a resource constraint for labor,

$$\sum_{j=1}^n l_j + \sum_{j=1}^n \theta_j f_j L \leq L. \tag{5}$$

An allocation is *efficient* if it solves $\mathcal{P}$. To solve $\mathcal{P}$, it is useful to first find the best allocation when the production network is fixed, that is, when the vector θ is taken as given. We can then take a step back to find the efficient vector θ.

3.1 Planner's problem with exogenous firm status θ

For a fixed θ, $\mathcal{P}$ is a standard convex maximization problem and the usual Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize its solution. Denote by λ_j the Lagrange multiplier on the resource constraint (4) for good j and by w the multiplier on the labor resource constraint (5). The first-order conditions imply that $(1-\alpha_j)\, y_j \lambda_j = w l_j$ so that, as in Oberfield (2018), we can define $q_j = w/\lambda_j$ as a measure of firm j's labor productivity.

From the planner's first-order conditions, we can characterize the vector $q = (q_1, \ldots, q_n)$ as a function of θ.

Proposition 1. *In the efficient allocation, the labor productivity vector q satisfies*

$$q_j = z_j \theta_j A \left(\sum_{i=1}^n \Omega_{ij} q_i^{\varepsilon_j - 1}\right)^{\frac{\alpha_j}{\varepsilon_j - 1}}, \tag{6}$$

for all j. Furthermore, there is a unique q that solves (6) such that $q_j > 0$ if j operates and has access to an operating cycle, and $q_j = 0$ otherwise.

The proof of Proposition 1 shows that q can be found by iterating on the mapping (6).

Several features of (6) are worth emphasizing. First, its recursive structure implies that a change in the labor productivity q_j of a firm j propagates downstream through supply chains. For instance, if j faces a negative TFP shock, the amount of labor needed to produce one unit of good j increases, which leads to a higher unit labor cost for j's customers, and for its customers' customers and so on. Second, (6) implies that a firm that has access to a greater set of active suppliers (more positive terms in the summation), is more productive (higher q_j). Intuitively, with a more diverse set of inputs a firm might be able to use better production techniques that would otherwise be unavailable. The elasticity ε_j governs how substitutable these inputs are with each other and is the key parameter determining the strength of this mechanism. When ε_j is small, intermediate inputs are poor substitutes and the benefit of having an additional supplier is large. In contrast, when ε_j is large, firm j's labor productivity is almost entirely driven by its most productive supplier. As we will see in Section 5, these mechanisms have important implications for the structure of the network and for the propagation of shocks in this economy.

With q in hand, it is straightforward to derive all other quantities in the efficient allocation (see Appendix A for the equations). In particular, the following lemma shows that GDP C can be computed as the product of aggregate productivity

$$Q = \left(\sum_{j=1}^{n} \beta_j q_j^{\sigma-1} \right)^{\frac{1}{\sigma-1}}. \tag{7}$$

and the amount of labor available after fixed costs have been paid.

Lemma 1. *In the efficient allocation, aggregate output is*

$$C = Q \left(1 - \sum_{j=1}^{n} \theta_j f_j \right) L. \tag{8}$$

We see from (7) that aggregate productivity Q is a CES aggregator of the underlying firm-level labor productivities q_j, with an elasticity of substitution controlled by σ. When σ is small, the differentiated goods are poor substitutes and each additional good is more highly valued by the planner. When, instead, σ is large the household derives utility primarily from the most productive firm. As a result, σ affects the planner's incentives to operate more firms and will also play an important role in shaping the structure of the network.

3.2 The planner's problem with θ as a choice variable

Lemma 1 completes the solution of the planner's problem with a fixed θ and we can now take a step back to consider the full problem $\mathcal{P}$, in which the network itself is a choice variable. By combining Lemma 1 and Proposition 1, we can rewrite $\mathcal{P}$ as the problem of finding the vector θ^*

that maximizes consumption (8), and where q solves (6). The left-hand side of Figure 2 lays out that version of $\mathcal{P}$.

Equations (6) and (8) highlight the trade-off faced by the planner when deciding whether to operate a firm j. Because of the recursive structure of (6), operating j improves the labor productivity q not only of j itself, but also of all its downstream customers, which benefits aggregate productivity Q. On the other hand, operating j also takes $f_j L$ units of labor away from other uses.

The full problem $\mathcal{P}$ with endogenous θ is hard to solve for two reasons.[10] First, θ is limited to the *corners* $\{0,1\}^n$ of the n-dimensional unit hypercube—a non-convex set. But even if θ could move freely over $[0,1]^n$, the fixed costs create firm-level increasing returns to scale that break the (quasi) concavity of the objective function. As a result, there are usually multiple local maxima, and the standard Karush-Kuhn-Tucker conditions are not sufficient to find the global maximum.

There is however a brute-force way of solving $\mathcal{P}$. Since there are only a finite number of vectors θ in the feasible set $\{0,1\}^n$, one can try them all. For each θ, q can be found by iterating on (6) and the objective function can then be computed using (8). While this *exhaustive search* approach is guaranteed to find the correct solution, it is in practice limited to economies with only a few firms. Since there are 2^n possible θ in $\{0,1\}^n$, the number of vectors to try explodes as n grows.[11]

Reshaping the planner's problem

To handle economies with large n, this paper proposes a novel solution method that is less computationally intensive. The key idea is to find an alternative optimization problem that is easy to solve and whose solution coincides with that of $\mathcal{P}$. This alternative problem, denoted by $\mathcal{R}$, is obtained by *relaxing* and *reshaping* $\mathcal{P}$. $\mathcal{R}$ is defined on the right-hand side of Figure 2.

$\mathcal{R}$ differs from $\mathcal{P}$ in two important ways—emphasized in blue in Figure 2. First, the binary constraint $\theta \in \{0,1\}^n$ is relaxed, and θ can now take values *inside* the unit hypercube $[0,1]^n$. While this relaxation has the advantage of convexifying $\mathcal{P}$'s feasible set, it also augments the planner's problem with points that have no real economic meaning. For instance, $\theta_j = 0.5$ does not correspond to any physical reality in the economic environment. One of the key insight of the paper is that, since they have no interest on their own, we can change the value of the objective function over these new points—and only over these points—to help us solve $\mathcal{P}$.

This is done in (9), which is a transformed version of (6) that includes the *shape parameters* $a_j > 0$ and b_{ij}.[12] These parameters modify the shape of the optimization problem everywhere

[10] It belongs to the class of Mixed Integer Nonlinear Problems (MINLP). Their combinatorial nature makes them notoriously challenging to solve and they are, in general, NP-Hard (Garey and Johnson, 1990).

[11] In some models, it is possible to order firms in some way and to progressively shut down the "worst" ones until the desired allocation is found. Here, however, firms differ along several dimensions of heterogeneity so that this approach does not apply directly.

[12] For 9 to be well-defined as $\theta_j \to 0$, we impose that $b_{ij} \geq -a_i (\varepsilon_j - 1)$ for all i, j.

$\mathcal{P}$: Original planner's problem

$$\max_{\theta \in \{0,1\}^n} Q\left(1 - \sum_{j=1}^{n} \theta_j f_j\right) L \qquad (8)$$

where q solves, for each $j \in \mathcal{N}$,

$$q_j = z_j \theta_j A \left(\sum_{i=1}^{n} \Omega_{ij} q_i^{\varepsilon_j - 1}\right)^{\frac{\alpha_j}{\varepsilon_j - 1}} \qquad (6)$$

$\mathcal{R}$: Relaxed and reshaped problem

$$\max_{\theta \in \boldsymbol{[0,1]^n}} Q\left(1 - \sum_{j=1}^{n} \theta_j f_j\right) L \qquad (8)$$

where q solves, for each $j \in \mathcal{N}$,

$$q_j = z_j \theta_j^{\boldsymbol{a_j}} A \left(\sum_{i=1}^{n} \Omega_{ij} \left(\boldsymbol{\theta_i^{b_{ij}}} q_i\right)^{\varepsilon_j - 1}\right)^{\frac{\alpha_j}{\varepsilon_j - 1}} \qquad (9)$$

Figure 2: Differences between the original and the reshaped problems

except over $\mathcal{P}$'s original feasible set $\{0,1\}^n$. Indeed, for $\theta \in \{0,1\}^n$, $\theta_j^{a_j} = \theta_j$ for all j. Similarly, for b_{ij}, if $\theta_i = 0$ then $q_i = 0$ anyway, and if $\theta_i = 1$ then $\theta_i^{b_{ij}} = 1$. In both cases, the term in the summation is unchanged. This reshaping procedure therefore preserves the ranking, in terms of utility, of the corners $\{0,1\}^n$—the only points with actual economic meaning—while changing the shape of the optimization problem elsewhere.

The shape parameters a_j and b_{ij}

For this solution method to be useful it must be that the reshaped problem $\mathcal{R}$ is easy to solve and that the solution to $\mathcal{R}$ also solves the original planner's problem $\mathcal{P}$. Crucially, we are free to pick the parameters a_j and b_{ij} to achieve these objectives. In particular, we can pick these parameters to increase the concavity of $\mathcal{R}$ with the goal of removing the undesirable local maxima that prevent an easy resolution of the relaxed problem. On the other hand, too much concavity can create a new global maximum somewhere in the middle of $[0,1]^n$, in which case the solutions of $\mathcal{P}$ and $\mathcal{R}$ would clearly differ.

To figure out how to properly set a_j and b_{ij}, it is useful to consider $\mathcal{R}$'s first-order condition with respect to θ_j.

Lemma 2. *The first-order conditions of the reshaped planner's problem $\mathcal{R}$ can be written as*

$$(1 + a_j)\, \lambda_j c_j + \sum_{k=1}^{n} (1 + a_j + b_{jk})\, \lambda_j x_{jk} - \sum_{i=1}^{n} \lambda_i x_{ij} - w l_j - w \theta_j f_j L = \theta_j \Delta\mu_j, \qquad (10)$$

where $\Delta\mu_j = \overline{\mu}_j - \underline{\mu}_j$ is the difference between the Lagrange multipliers on the constraints $\theta_j \leq 1$ and $\theta_j \geq 0$, respectively.

Lemma 2 provides the accounting of the resources that go into the decision to operate firm j. The first two terms in (10) capture the value of the goods produced by firm j that go to the household and other producers (recall that λ_j is the Lagrange multiplier associated with the resource constraint (4) for good j). The last three terms on the left-hand side correspond to the intermediate inputs and the amount of labor that are needed to operate j. It is clear from (10) that a_j and b_{ij} are essentially changing the value of good j in the first-order conditions. Why is such a change needed? Consider the first term in (10). If we set $a_j = 0$, the marginal value of an extra unit of consumption c_j is equal to the marginal utility of consumption $\lambda_j = \frac{\partial C}{\partial c_j}$. But this reasoning at the margin is misleading since the choice of the planner is binary: the firm is either operating or not. We can compute instead the full gain in utility provided by operating j, which is

$$\Delta C = \int_0^{c_j} \frac{\partial C}{\partial c_j} d\tilde{c}_j = \int_0^{c_j} \tilde{c}_j^{-\frac{1}{\sigma}} C^{\frac{1}{\sigma}} d\tilde{c}_j = \frac{\sigma}{\sigma - 1} c_j \frac{\partial C}{\partial c_j}.$$

This suggests that the benefit of operating firm j in (10) should be proportional to $\frac{\sigma}{\sigma-1}$ to capture the full impact of operating j on the household's utility. A similar reasoning implies that the benefit of operating firm j to a customer k (second term in 10) should be proportional to $\frac{\varepsilon_k}{\varepsilon_k - 1}$. Based on that discussion, we therefore set

$$a_j = \frac{1}{\sigma - 1} \quad \text{and} \quad b_{ij} = \frac{1}{\varepsilon_j - 1} - \frac{1}{\sigma - 1} \tag{$\star$}$$

from now on and verify later that these specific values push the solution to $\mathcal{R}$ toward the efficient allocation. In the next section, we will discuss how different values of a_j and b_{ij} can instead push that solution toward an inefficient equilibrium.

Sufficiency of the first-order conditions

Optimization problems are particularly easy to solve when they are convex. In that case the first-order conditions are necessary and sufficient to characterize a solution, and standard algorithms can converge rapidly to the global maximum. The next two propositions describe conditions under which $\mathcal{R}$ is convex.

Proposition 2. *Let $\varepsilon_j = \varepsilon$ and $\alpha_j = \alpha$ for all j. If $\Omega_{ij} = d_i e_j$ for some vectors d and e then the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution to $\mathcal{R}$.*

A similar result holds for a different set of Ω matrices. Define $\bar{\Omega} = \omega (O_n - I_n)$ where O_n is the $n \times n$ matrix full of ones, I_n is the $n \times n$ identity matrix and $\omega > 0$. The matrix $\bar{\Omega}$ describes a network of potential connections in which firms are connected to each other, but not with themselves, with the same intensity ω. The following proposition shows that $\mathcal{R}$ is easy to solve when Ω is close to $\bar{\Omega}$.

Proposition 3. *Let $\sigma = \varepsilon_j$ for all j. Suppose that the $\{\beta_j\}_{j \in \mathcal{N}}$ are not too far from each other and that the matrix Ω is close enough to $\bar{\Omega}$. Then there exists a threshold $\bar{f} > 0$ such that if $f_j < \bar{f}$ for all j the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution* *to $\mathcal{R}$.*[13]

Propositions 2 and 3 establish sufficient conditions under which a feasible point θ^* that satisfies the first-order conditions and the complementary slackness condition solves $\mathcal{R}$.[14] As a result, standard numerical algorithms, such as gradient ascent, can rapidly solve $\mathcal{R}$ even for economies with thousands of firms.

Equivalence of the solutions

The next proposition establishes conditions under which a solution to $\mathcal{R}$ also solves $\mathcal{P}$.

Proposition 4. *If $\theta^* \in \{0,1\}^n$ solves $\mathcal{R}$, then θ^* also solves $\mathcal{P}$.*

This result follows directly from the fact that the feasible set of $\mathcal{R}$ contains the feasible set of $\mathcal{P}$ and that both of their objective functions coincide over $\{0,1\}^n$ by construction.

Together, Propositions 2 to 4 offer a convenient way to solve $\mathcal{P}$. First, find a solution θ^* to $\mathcal{R}$ using a standard algorithm for convex problems. If θ^* belongs to $\{0,1\}^n$ then it also solves $\mathcal{P}$. This last condition can of course be tested in practice, but the whole solution approach would not be very useful if θ^* rarely belonged to $\{0,1\}^n$. Fortunately, condition ($\star$) is such that solutions to $\mathcal{R}$ are naturally pushed toward $\{0,1\}^n$. To understand why, consider a stylized version of the first-order condition associated with the operating status θ_j of some firm j. That condition can be written as

$$\text{Marginal Benefit}_j\left(\theta_j, F_j\left(\theta\right)\right) - \text{Marginal Cost}_j\left(\theta_j, G_j\left(\theta\right)\right) = \overline{\mu}_j - \underline{\mu}_j \tag{11}$$

where $\overline{\mu}_j$ and $\underline{\mu}_j$ are the Lagrange multipliers associated with the constraints $\theta_j \leq 1$ and $\theta_j \geq 0$, respectively, and where both the "Marginal Benefit" and the "Marginal Cost" of increasing θ_j are functions that depend not only on θ_j itself, but also on the operating status of the other firms through some functions F_j and G_j.[15] These functions depend on the whole vector θ, instead of the specific θ_j, and we therefore refer to them as *aggregates.*

The following proposition shows how setting the shape parameters a_j and b_{ij} to their ($\star$) values affects the "Marginal Benefit" and the "Marginal Cost" functions in (11).

Proposition 5. *Under the condition ($\star$), the marginal benefit and the marginal cost of increasing θ_j only depend on θ_j through the aggregates F_j and G_j.*

This proposition shows that it is possible to find values for a_j and b_{ij} such that the "Marginal Benefit" and the "Marginal Cost" no longer depend on θ_j directly. In this case, θ_j affects the first-order condition (11) only through the aggregates F_j and G_j. These aggregates are summations

[13] To be precise, let $\bar{\beta}$ be a $n \times 1$ vector with identical elements. Then there exists a ball $\mathcal{B} = \left\{(\Omega,\beta) : \left\|(\Omega,\beta) - \left(\bar{\Omega},\bar{\beta}\right)\right\| < \delta\right\}$ for $\delta > 0$ such that the statement holds for $(\Omega,\beta) \in \mathcal{B}$.

[14] See footnote 29 for some intuition for the restrictions needed by Propositions 2 and 3.

[15] The "Marginal Benefit" and "Marginal Cost" functions are defined in the proof of Proposition 5 in Appendix H.

over many firms and, as n increases, they become more and more independent of θ_j itself, and so does the left-hand side of (11).[16,17] This pushes solutions to $\mathcal{R}$ toward $\{0,1\}^n$. To understand why, consider for instance a gradient ascent algorithm that begins at $\theta_j = 1/2$. If the marginal benefit of increasing θ_j is larger than the marginal cost, the planner increases θ_j slightly. But since the marginal benefit and marginal cost themselves are independent of θ_j, the marginal benefit remains larger and the planner keeps increasing θ_j until it reaches 1, at which point the Lagrange multiplier $\overline{\mu}_j$ becomes positive to enforce the constraint $\theta_j \leq 1$. The opposite happens if the marginal benefit is initially lower than the marginal cost. As a result, the solution θ^* to $\mathcal{R}$ is pushed toward $\{0,1\}^n$, and Proposition 4 guarantees that θ^* also solves $\mathcal{P}$.

3.3 Example with two firms

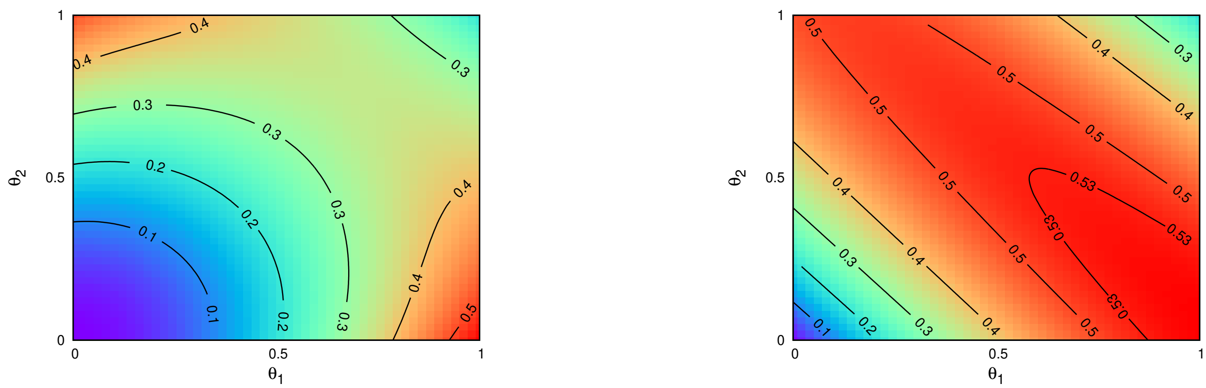
To better understand how the solution approach works, it is helpful to consider a simple economy with two firms $j \in \{1,2\}$ and a complete set of potential connections between them ($\Omega = O_2$). The objective function $V(\theta)$ of the relaxed planner's problem without any reshaping ($a_j = 1, b_{ij} = 0$) is shown in Figure 3a, where warmer colors represent higher utility levels for the planner. The horizontal and vertical axes refer to the operating decisions θ_1 and θ_2. We see that V is shaped like a saddle with local maxima at $(\theta_1, \theta_2) = (1,0)$ and $(0,1)$, and local minima at $(0,0)$ and $(1,1)$. The global maximum is at $(1,0)$.

Since V is not concave, first-order conditions are not sufficient to characterize the global maximum—they are indeed satisfied at both $(0,1)$ and $(1,0)$. As a result, this problem cannot be solved reliably with standard algorithms. Starting from an initial point, these algorithms move locally by following the steepest slope, so they can easily converge to the local maximum at $(0,1)$.

Figure 3b shows the objective function $V_R(\theta)$ of the same optimization problem but, this time, reshaped according to condition ($\star$). Three things are worth noticing. First, V and V_R coincide, by construction, at the corners $\{0,1\}^2$. As a result, the ranking of these corners, in terms of utility, is the same in both problems. Second, the reshaping procedure stretches the objective function so that V_R is concave. The first-order conditions are therefore sufficient to characterize the global maximum. Third, the procedure did not create another maximum somewhere inside $[0,1]^2$, and V_R's maximum is also the maximum of V. As a result, starting from any initial θ_0 in $[0,1]^2$, a simple gradient ascent algorithm will converge to the global maximum at $(1,0)$.

[16] The proof of Proposition 5 shows that F_j and G_j are functions of Q and $B_j = \left(\sum_{i\in\mathcal{N}} \theta_i^{b_{ij}} \Omega_{ij} q_i^{\varepsilon_j - 1}\right)^{1/(\varepsilon_j - 1)}$. When Ω is very sparse, it is possible that q_i depends strongly on θ_j itself, in which case B_j would remain affected by θ_j even when n is large. These cases are easily identified in practice: the numerical solution to $\mathcal{R}$ will be outside of $\{0,1\}^n$. Appendix B.3 shows that the solution approach performs well even in these more pathological cases. Appendix B.4 shows the performance of the algorithm when Ω is sparse.

[17] This argument might seem to suggest that the solution approach only works for very large n, but that is not the case. Table 1 below shows that it also performs well in economies with only a few firms. Even when n is small, F_j and G_j do not vary much with θ_j under condition ($\star$).



(a) The objective function $V(\theta)$ of the relaxed (but not reshaped) problem is not concave

(b) The objective function $V_R(\theta)$ of the relaxed and reshaped problem is concave

Notes: The parameters in this example are $n = 2$, $\alpha_j = 0.5$, $L = 1$, $f = 0.45$, $\sigma = \varepsilon_j = 5$, $z_1 = 1$, $z_2 = 0.95$, $\Omega_{ij} = 1$ for all i, j.

Figure 3: Reshaping the planner's problem in a simple economy

3.4 Numerical tests

The theoretical results of the last section give us *sufficient* conditions under which reshaping the planner's problem provides the solution to $\mathcal{P}$, but these conditions are not necessary. In this section, we show through numerical simulations that the solution approach also works well when these conditions are not satisfied. We first consider economies with only a few firms and then present results for economies with a large number of firms.

Economies with few firms

With a small number of firms, it is possible to find the true solution to the planner's problem by comparing the utility provided by the 2^n possible vectors $\theta \in \{0,1\}^n$, as in the exhaustive search algorithm described above. We can then compare this allocation to the solution of the relaxed problem with and without reshaping. Appendix B provides the details of the simulations. They involve a broad range of economies with firms that differ along all the dimensions of heterogeneity allowed by the model. They also cover matrices Ω with different shapes and various degrees of sparsity.

The results for economies with up to $n = 14$ firms are presented in Table 1. We see that reshaping the planner's problem (first two columns) attributes the correct status θ to more than 99.9% of the firms across simulations. It also finds output levels that are within 0.001% of their correct values. In contrast, without reshaping the problem (last two columns), over 15% of the firms can be assigned the wrong status θ and the average error in output can reach above 0.9%, a large number when studying aggregate fluctuations. The table also shows that the performance of the reshaping algorithm stays relatively constant as n increases, in contrast to the non-reshaped solution which performs worse as the number of firms increases.[18]

Table 1: Testing the solution approach for small n

	With reshaping		Without reshaping	
n	Correct θ	Error in C	Correct θ	Error in C
4	99.9%	0.000%	91.5%	0.502%
6	99.9%	0.000%	88.1%	0.692%
8	99.9%	0.000%	86.5%	0.791%
10	99.9%	0.001%	85.2%	0.855%
12	99.9%	0.001%	84.5%	0.903%
14	99.9%	0.001%	84.0%	0.928%

Notes: See Appendix B for the details of the simulations.

Economies with many firms

When n is large, finding the true solution to $\mathcal{P}$ through an exhaustive search would take an infeasibly long time. We can, however, verify whether there exist welfare-improving deviations from the solutions to the relaxed problems. To do so, we change the operational status θ_j of each firm to see if it improves the utility of the planner. We keep repeating this procedure as long as there are deviations to be found. We then compare this deviation-free solution to the original one. The precise algorithm is described in Appendix G.3.

Since this procedure is computationally costly, we only consider economies that follow the calibration of Section 6. The results are presented in Table 2. Again, the reshaping approach performs very well. After all the possible deviations are accounted for, more than 99.9% of the firms have kept the same operating status θ_j and aggregate output has changed by a negligible amount.[19] In contrast, without reshaping more than 30% of the firms are assigned the wrong operating status, and the error in aggregate output amounts to 0.56%. While this test does not guarantee that the solution approach finds the correct efficient allocation, it provides a good indication that there are no obvious mistakes in its solution.

Appendix B provides several additional exercises to further test the robustness of the solution method. It considers economies 1) with very sparse matrices Ω, 2) in which the production network is created through individual link formation, and 3) for which the solution to $\mathcal{R}$ is not in $\{0,1\}^n$. These additional tests show that the solution method performs well in a broad set of economic environments.

[18]Table 1 shows the fraction of firms with the correct status θ_j. We can also consider the fraction of economies for which the whole vector θ is correct. Table 8 in Appendix B.2 presents these results. On average, 99.7% of economies have the correct vector θ with reshaping. Without reshaping that number is 19.9%.

[19]When the reshaping approach fails it is often because it gets the wrong operating status for a firm that is fairly isolated from the rest of the network. Since, these firms are in general small, they only have little influence on aggregate production, which explains why the error in output is very small in Table 2.

Table 2: Testing the reshaping approach for n large

	With reshaping		Without reshaping	
n	Correct θ	Error in C	Correct θ	Error in C
1000	99.9%	$< 0.001\%$	66.5%	0.56%

Notes: Parameters as in the calibrated economy (see Section 6.2). We simulate 100 different matrices Ω and, for each Ω, draw 100 productivity vectors z. We run the procedure described in Appendix G.3 on each of them and report average results. $x < 0.001\%$ indicates that $x > 0$ but that proper rounding would yield 0.

4 Reshaping and equilibria

The previous section described how we can reshape the planner's problem to find the efficient allocation. In the current section, we focus instead on equilibrium allocations. We are mainly concerned with two questions. First, can we decentralize the efficient allocation as an equilibrium and, second, can we also use the reshaping solution method to find equilibrium allocations that are inefficient? In reality, distortions like market power or coordination failure might lead to inefficiencies, and it would be useful if the solution method could also find these allocations.

In a production network setting, the properties of an equilibrium depend critically on how extensively firms are allowed to interact with each other. At one extreme, we can allow for rich interactions, even between firms that are far apart in the network. Oberfield (2018) considers such an equilibrium. In a nutshell, firms are facing contractual obligations to purchase and deliver goods, and an equilibrium is an allocation in which no groups of firms want to deviate from the terms of their contracts. The rich interactions imply that firms can internalize any externalities that they impose on each other and, as a result, such a *stable* equilibrium is efficient. This setting therefore provides a decentralization of the efficient allocation as the outcome of individual decisions and market forces. The details of that equilibrium definition, and the formal result about efficiency, can be found in Appendix D.

In contrast, we can also think of a different equilibrium concept, closer to the standard monopolistic competition benchmark, in which firms simply maximize their individual profits without internalizing their impact on other producers. To explore both efficient and inefficient equilibria in that setting, we consider two versions of that equilibrium that differ in how prices are set in firm-to-firm transactions. In the first version, firms have some amount of market power, and offsetting subsidies are assumed to be in place. We show that the equilibrium can be efficient in this case, and so we refer to this version as the "undistorted equilibrium". In the second version, which we refer to as the *distorted equilibrium*, prices in firm-to-firm transactions are set by a take-it-or-leave-it offer by the purchasing firm. The idea is to capture what might be an important source of distortions in reality: the presence of superstar firms such as Walmart and IKEA that have a lot of pricing power with their suppliers.[20] In this case, the entry/exit decisions are inefficient, but we show in this section that the reshaping methodology can also be used as a tool to find such equilibria. In the next section, we describe how cascades of firm shutdowns propagate differently in the two versions of the equilibrium.

4.1 Equilibrium definitions

We begin by describing some elements that are common to both version of the equilibrium. The representative household owns the firms, supplies labor and purchases consumption goods from the individual firms. It takes all prices as given and maximizes the utility function (1) subject to the budget constraint

$$\sum_{j=1}^{n} P_j c_j \leq WL + \Pi + T, \tag{12}$$

where P_j is the price of good j, W is the wage, Π is profits and T is a lump-sum transfer from the government. The household's maximization problem yields the demand curve for good j,

$$c_j = \beta_j C \left(\frac{P_j}{\overline{P}} \right)^{-\sigma}, \tag{13}$$

where $\overline{P} = \left(\sum_j \beta_j P_j^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$ is the price index.

The decisions made by the firms take place over two sub-periods. In the first sub-period, each firm j decides whether to pay a sunk cost $f_j L$ to enter production. In the second sub-period, firms that have paid the fixed cost can produce. They hire labor, purchase intermediate inputs and sell their own good to the household and to other firms.

Before describing how prices are set, it is convenient to first define a firm's marginal cost of production in the second sub-period. If a firm j hires labor at a wage W and purchases input i at a price Q_{ij}, its marginal cost δ_j is the outcome of the cost-minimization problem[21]

$$\delta_j := \min_{x,l} \sum_{i=1}^{n} Q_{ij} x_{ij} + W l_j, \tag{14}$$

subject to $y_j \geq 1$, where y_j is given by the production function (2).[22]

We now describe how prices are set. Under both version of the equilibrium, firms take into account the demand curve (13) when selling to the household. They therefore have some amount of monopoly power and earn positive profits from these transactions. In contrast, the way prices

[20] Bloom and Perry (2001) document that small suppliers perform relatively worse when they have Walmart as a customer. There is also plenty of anecdotal evidence that Walmart squeezes the profit margins of its suppliers (see for instance Van Riper, 2199).

[21] We let prices Q_{ij} depends on the buyer j for now but later show that this will not be the case in equilibrium.

[22] Since firms have constant returns to scale (once f_j has been paid), δ_j is independent of the amount produced.

in firm-to-firm transactions are set vary between the two equilibrium definition. In an undistorted equilibrium, firms also have some amount of market power when setting these prices. Specifically, firm j selling goods to firm k faces the demand curve

$$x_{jk} = \Omega_{jk} X_k \left(\frac{Q_{jk}}{\overline{Q}_k} \right)^{-\varepsilon_k}, \text{ where } X_k = \left(\sum_j \Omega_{jk}^{\frac{1}{\varepsilon_k}} x_{jk}^{\frac{\varepsilon_k - 1}{\varepsilon_k}} \right)^{\frac{\varepsilon_i}{\varepsilon_i - 1}} \tag{15}$$

is firm k's bundle of intermediate inputs, Q_{jk} is the transaction price and $\overline{Q}_k = \left(\sum_j \Omega_{jk} Q_{jk}^{1-\varepsilon_k} \right)^{\frac{1}{1-\varepsilon_k}}$ is firm k's price index. In this case, we can write the problem of a firm in the second sub-period as maximizing profits

$$\pi_j^{\text{undist}} = P_j c_j + \sum_{k=1}^{n} \left(1 + s_{jk}^x\right) Q_{jk} x_{jk} - \sum_{i=1}^{n} Q_{ij} x_{ij} - W l_j - W f_j L \theta_j, \tag{16}$$

subject to a resource constraint

$$c_j + \sum_{k=1}^{n} x_{jk} \le y_j, \tag{17}$$

and to the demand curves (13) and (15). Since our goal here is to find an equilibrium that coincide with the efficient allocation, we also allow for a subsidy s^x that increases the revenues from firm-to-firm sales, and we set it to its efficient level $s_{jk}^x = \frac{1}{\varepsilon_k - 1}$ to compensate for j's market power when selling to k. As a result, firm j sells to the household at a price $P_j = \frac{\sigma}{\sigma-1}\delta_j$ and to an intermediate producer k at a price $Q_{jk} = \frac{\varepsilon_k}{\varepsilon_k - 1} \frac{1}{1+s_{jk}^x} \delta_j = \delta_j$.

In contrast, in the distorted equilibrium, the price of goods that are sold between intermediate producers is determined by a take-it-or-leave-it offer, made by the customer to the supplier. That offer specifies a price at which any amount of goods can then be purchased. Each offer is made taking all the other equilibrium prices as given. Profit maximization implies that the supplier is not willing to accept an offer with a price below its marginal cost of production, otherwise each unit sold would reduce profits. As a result, the customer offers a price exactly equal to that marginal cost so as to maximize its own profits.[23,24]

[23] Acemoglu and Tahbaz-Salehi (2020) use similar assumptions in an entry model with variable markups. These assumptions imply that the household and the firms might pay different prices for a given good. While this is a bit unappealing, this pricing structure has the advantage that trades between firms are undistorted by markups. To justify that buyers cannot resell the goods to the household (in which case the price difference would disappear), we can assume that each good must be customized by the seller for each specific buyer. Alternatively, goods might be perishable so that they can only be transported once.

[24] If the customer also provides intermediate goods to the supplier, the prices in these two transactions are set through two different offers that are handled independently by both parties.

The profit maximization problem of a firm j in the distorted equilibrium is therefore

$$\pi_j^{\text{dist}} = P_j c_j + \sum_{k=1}^{n} Q_{jk} x_{jk} - \sum_{i=1}^{n} Q_{ij} x_{ij} - W l_j - W f_j L \theta_j, \tag{18}$$

subject to the resource constraint (17), to the household's demand curve (13) and to the fact that $Q_{kl} = \delta_k$ for all k, l. Notice that there are no subsidies in (18). Our goal here is to show that the reshaping technique can help to find an inefficient equilibrium, and as a result, we do not introduce subsidies or taxes that could undo any inefficiencies.

The entry problem of the firm in the first sub-period is similar under both versions of the equilibrium. A firm j pays the fixed cost $W f_j L$ if and only if $\pi_j \geq 0$. To compute π_j, firms use the equilibrium prices to calculate their own marginal cost of production δ_j conditional on entry, which is given by (14). With δ_j, they can then compute their own prices and the demand for their goods from the household and the other firms. As a result, they can also compute π_j and make their optimal entry decision.[25]

We are now ready to define an undistorted equilibrium.

Definition 1. An *undistorted equilibrium* is a set of prices (P, Q, W) and an allocation (c, l, x, θ) such that: 1) given input prices and the demand curves (13) and (15), firms pick (c, l, x, θ) to maximize profits (16) subject to the resource constraint (17); 2) given prices the household maximizes utility (1) subject to (12), where $\Pi = \sum_j \pi_j$ and T pays for the subsidies; and 3) all markets clear.

The definition of a distorted equilibrium is the same except that firms are not subject to the demand curve (15). Instead, the prices Q_{jk} are equal to the marginal cost δ_j, given by (14).

4.2 Characterizing undistorted and distorted equilibria

We first characterize the equilibrium decisions under an exogenously given vector of entry decisions θ, and show that in this case the equilibrium allocations (c, l, x) under both the undistorted and distorted definitions are efficient. We then consider the equilibrium entry decisions of the firms and compare them to the efficient allocation.

Equilibrium decisions under a fixed θ

To better understand the links between an equilibrium and the efficient allocation, it is useful to characterize the vector of equilibrium unit costs δ_j. As discussed above, the pricing mechanisms imply that $Q_{jk} = \delta_j$ under both versions of the equilibrium. Together with (14) this implies that[26]

[25] To be precise, when contemplating whether to enter the economy or not, firms take as given the demand curves and the input prices that they face, so that they don't internalize the impact of their entry decision on C and $\overline{P}$ in (13), and on X_k and $\overline{Q}_k$, for all k, in (15). In the distorted equilibrium, the demand coming from intermediate producers is not straightforward to compute but since these goods are sold at marginal cost they do not affect profits.

[26] See the proof of Proposition 6 for a derivation of this result.

$$\delta_j = \frac{1}{z_j A} \left(\sum_{i \in \mathcal{N}} \Omega_{ij} \theta_i \delta_i^{1-\varepsilon_j} \right)^{\frac{\alpha_j}{1-\varepsilon_j}} W^{1-\alpha_j}. \tag{19}$$

This equation is essentially the same as (6), which pins down labor productivity in the efficient allocation. This implies that the equilibrium pricing mechanisms do not introduce wedges that would distort firm-to-firm transactions from their efficient price levels. As a result, the equilibrium decisions (c, l, x) coincide with the efficient allocation. The following proposition makes this point formally.[27]

Proposition 6. *For a given entry decision vector θ, distorted and undistorted equilibria are efficient. Furthermore, the equilibrium prices W and Q_{jk} are equal (up to a choice of numeraire) to the planner's Lagrange multipliers w and λ_j (associated with the resource constraints (5) and (4), respectively).*

This proposition establishes a connection between an equilibrium and the efficient allocation conditional on a given θ. It follows that for the whole equilibrium allocation, including the vector θ, to be efficient, entry decisions in the equilibrium and the efficient allocation must coincide. We now move on to characterize these decisions.

Entry in an undistorted equilibrium

By combining the pricing rules with (16), we can write the profits of a firm j in an undistorted equilibrium as

$$\pi_j^{\text{undist}} = \frac{\sigma}{\sigma - 1} \delta_j c_j + \sum_{k=1}^{n} \frac{\varepsilon_k}{\varepsilon_k - 1} \delta_j x_{jk} - \sum_{i=1}^{n} \delta_i x_{ij} - W l_j - W f_j L \theta_j. \tag{20}$$

Comparing this equation with the condition (10) that captures the entry margin in the reshaped planner's problem, we see that both equations are essentially the same under the specific shape parameters given by ($\star$) (recall that $\delta_j = \lambda_j$ by Proposition 6). This suggests that entry decisions in the undistorted equilibrium coincide with those of the reshaped planner's problem. The following proposition establishes this result.

Proposition 7. *A vector $\theta = \{0, 1\}^n$ that satisfies the first-order conditions of the reshaped problem $\mathcal{R}$ with shape parameters ($\star$) is an undistorted equilibrium.*

Two direct consequences from this proposition are worth highlighting. First, whenever a solution $\theta = \{0, 1\}^n$ of the reshaped problem coincides with the efficient allocation, that allocation is also an undistorted equilibrium. As a result, we can think of the efficient allocation as arising from

[27] To be precise, the word "efficient" in Proposition 6 refers to a planner's problem with the same fixed θ.

market forces and individual agents interacting through decentralized markets. Second, and most importantly, if we solve the reshaped problem numerically and find a point $\theta = \{0,1\}^n$ that satisfies the first-order conditions, we know that this point corresponds to an undistorted equilibrium, even though it might not coincide with the efficient allocation. This is reassuring from a practical point of view: even in a complicated economy that might not satisfy the conditions of Propositions 2 or 3, we can be sure that the allocation provided by the reshaping method is economically meaningful.[28,29]

Entry in a distorted equilibrium

The reshaping methodology can also be used to characterize an equilibrium that is distorted away from the efficient allocation. To see this, we can once again combine the firm's profits (18) together with the pricing rules to write profits in a distorted equilibrium as

$$\pi_j^{\text{dist}} = \frac{\sigma}{\sigma - 1}\delta_j c_j + \sum_{k=1}^{n} \delta_j x_{jn} - \sum_{i=1}^{n} \delta_i x_{ij} - W l_j - W f_j L \theta_j. \tag{21}$$

This equation is similar to the profits of the firm in the undistorted equilibrium, given by (20), with the notable exception that selling one unit of good to another firm brings in only the marginal cost δ_j, instead of $\frac{\varepsilon_k}{\varepsilon_k - 1}\delta_j$. These sales therefore generate no profits, and so losing a customer has no direct impact on entry decisions. This has important implications for the propagation of cascades, as we will see in the next section. The absence of a markup in firm-to-firm transactions also implies less profit than in the undistorted equilibrium and weaker incentives to operate. As a result, entry decisions are in general inefficient, but the reshaping method can still be used to characterize the equilibrium. Comparing (21) with the planner's equivalent condition (10) suggests to use the shape parameters

$$a_j^d = \frac{1}{\sigma - 1} \quad \text{and} \quad b_{ij}^d = -\frac{1}{\sigma - 1} \tag{22}$$

for that purpose instead of those given by ($\star$). The following proposition establishes that result.

Proposition 8. *A vector $\theta = \{0,1\}^n$ that satisfies the first-order conditions of the reshaped problem $\mathcal{R}$ with shape parameters (22) is a distorted equilibrium.*

This result shows that reshaping the planner's problem can be a useful tool to find not only

[28] Of course, if the conditions of Propositions 2 or 3 are satisfied, then we know that this point also corresponds to the efficient allocation. In this case, there is a unique undistorted equilibrium.

[29] In view of Proposition 7, we can think of the reshaping method as a tool to find undistorted equilibria. When the conditions of Propositions 2 or 3 are met, there is only one equilibrium and it is efficient. The conditions imposed by those propositions on Ω are important for the uniqueness. When Ω allows for isolated groups of firms, multiple equilibria can arise. Consider for instance a pair of firms (firms 1 and 2) such that 1's only potential input is 2, and 2's only potential input is 1. If f is small enough, there will be two equilibria: one in which both firms operate and one in which both firms do not. Indeed, if both firms are not operating, there is no incentives for either of them to deviate and operate. The deviating firm would have access to no input and would not be able produce. The conditions on Ω imposed by Propositions 2 and 3 rule out these isolated groups of firms, which pushes the reshaping method to find the efficient allocation.

the efficient allocation but also equilibrium allocations that are distorted. It also emphasizes the importance of the shape parameters a_j and b_{ij} for that purpose.

Some remarks about the equilibrium definitions

Before exploring the forces at work in the economy, some comments about the equilibrium are in order. First, we have assumed that firms behave atomistically when making decisions, in the sense that they take equilibrium quantities and the behavior of other producers as given. This assumption is common in the literature, and might be reasonable for large economies like the United States. For economies with a small number of firms, however, richer strategic interactions might play a more important role and deviations from efficiency might occur.[30] Second, we have only considered one source of inefficiency, which is that producers might have market power in firm-to-firm transactions. While this is arguably one of the main frictions in these markets, other distortions or market imperfections might certainly be at work in reality. Our goal here is not to provide an exhaustive study of these frictions, but rather to investigate how one plausible source of distortions might affect operation decisions. Third, Propositions 7 and 8 cast the complicated problem of finding an equilibrium in an economy with discrete decisions as that of finding a stable point in a continuous optimization problem, which is computationally straightforward. Fourth, under both the undistorted and distorted versions of the equilibrium multiple equilibria can arise.[31] While this multiplicity might be interesting on its own, it also raises the question of how an equilibrium is selected. In contrast, the efficient allocation is in general unique, which makes it a natural benchmark to study. As we have seen, the efficient allocation can also be thought of as arising from market forces, as either an undistorted equilibrium or a stable equilibrium (as in Appendix D). In what follow, we will therefore emphasize the efficient allocation, but we also highlight some discrepancies with the distorted equilibrium. In the quantitative section of the paper, we will see that both the efficient allocation and the distorted equilibrium behave in comparable ways, which suggests that the dominant economic forces are similar in both allocations.

5 Complementarities, cascades and clustering

We now explore how the economic forces at work in the environment influence the structure of the production network and the propagation of shocks. Complementarities between the operating decisions of nearby firms play an important role here. They lead to clustering of economic activity, cascades of firm shutdowns, and can trigger large reorganizations of the production network in response to small shocks. In what follows, we first describe the origin of these complementarities

[30] However, the stable equilibrium of Appendix D has rich firm-to-firm interactions and is efficient.

[31] This is common in models with complementarities and fixed costs (Schaal and Taschereau-Dumouchel, 2015). This multiplicity opens the door to coordination failures, as firms might coordinate on a suboptimal equilibrium.

and then turn to their impact on the economy. We describe some mechanisms in terms of the equilibrium allocation and other in terms of the planner's problem, depending on which perspective is more convenient. In the latter case, the same forces also operate in an undistorted equilibrium (Proposition 7) and in a stable equilibrium (Proposition 13).

5.1 Upstream and downstream complementarities

In the model, nearby firms tend to operate, or not, together. To highlight the origin of these complementarities, it is helpful to consider how the profits of a firm are affected by the operating status of its neighbors. In an undistorted equilibrium, we can write the operating profits (16) of a firm j under a given vector θ as

$$\pi_j^{\text{undist}}(\theta) = \frac{1}{\sigma-1}\delta_j(\theta)\, c_j(\delta_j(\theta)) + \sum_{k=1}^{n} \frac{1}{\varepsilon_k - 1}\theta_k \delta_j(\theta)\, x_{jk}(\delta_j(\theta)) - W f_j L, \tag{23}$$

where the demand curves are defined, as before, as

$$c_j(\delta_j) = \beta_j C \left(\frac{\sigma}{\sigma-1}\frac{\delta_j}{\overline{\overline{P}}}\right)^{-\sigma} \text{ and } x_{jk}(\delta_j) = \Omega_{jk} X_k \left(\frac{\delta_j}{\overline{\overline{Q}}_k}\right)^{-\varepsilon_k}, \tag{24}$$

and where the unit cost $\delta_j(\theta)$ is given by (19). The terms $\delta_j c_j$ and $\delta_j x_{jk}$ in (23) are the cost of goods sold to the household and to other intermediate producers. When adjusted for mark ups, these terms contribute to the profits of the firm.

We can use these equations to describe how the operating decision of a neighboring firm i affects j's own incentives to operate. To do so, we consider the impact on π_j^{undist} of a change from $\theta = (\theta_1, \ldots, \theta_i = 0, \ldots, \theta_n)$ to $\theta = (\theta_1, \ldots, \theta_i = 1, \ldots, \theta_n)$. In principle, this change in θ would affect several equilibrium quantities such as the wage rate W and aggregate consumption C. In practice, a single element of θ will have a negligible impact on these objects in an economy with a large number of firms, and we therefore take them as constant for our analysis. This partial equilibrium analysis allows us to sharply characterize the main forces that affect a firm's operating decisions.[32] Later on, we also provide results in which we do not take any quantities as given.

The equations (23)–(24) capture two important channels through which the operating status of a neighbor affects a firm's own operation decision. The first channel operates upstream, from suppliers to customers, and involves the first two terms in (23). Consider for instance the impact of a newly operating firm i that is directly upstream from j. Because of (19), this additional supplier implies that j is able to produce at a lower unit cost δ_j and to sell at a lower price. Its

[32]The notation in (23)–(24) highlights which quantities are affected by θ. To be precise, we consider the impact of a change in θ on j's profits π_j^{undist} keeping three sets of quantities fixed: 1) the wage level W, 2) the demand curves when selling to the household and to other producers (which in turn depend on aggregate consumption C, the aggregate price index $\overline{P}$, and for any of j's customer k the demand X_k and the price index $\overline{Q}_k$ conditional on operating), and 3) the unit costs δ_i, $i \neq j$.

good thus becomes more attractive and it is able to sell more units. This, in turn, leads to higher profits (notice that both $\delta_j c_j$ and $\delta_j x_{jk}$ in (23) are deceasing functions of δ_j) and j is more likely to operate as a result. This first channel therefore highlights a complementarity between firms' entry/exit decisions that operates from suppliers to customers.

A second channel at work in (23) operates upstream, from buyers to suppliers. The second term in (23) captures the importance for j's profits of the demand from other intermediate producers. If a firm k that is downstream from j begins operating ($\theta_k = 1$) this creates more demand for j's goods which results in higher profits and, as a result, more incentives to operate. This second channel therefore captures upstream complementarities in operating decisions.

These complementarities work differently in the distorted equilibrium. In that case, we can write j's profits as

$$\pi_j^{\text{dist}}(\theta) = \frac{1}{\sigma - 1}\delta_j(\theta)\, c_j(\delta_j(\theta)) - W f_j L, \tag{25}$$

where the key difference with the undistorted equilibrium is that, since the sale price is equal to the marginal cost of production, the firm does not receive any profits from selling to other intermediate producers. This has important implications for the complementarities in operating decisions. Indeed, the second channel mentioned above, which works through the demand of intermediate producers, is absent here, and so the operating status of a firm k that is downstream from j has no direct impact on j's decision to operate. Only j's suppliers, through their impact on δ_j, have any direct effect.

The following proposition formalizes this discussion.

Proposition 9. *Taking equilibrium quantities as given, the following holds.*

1. *In an undistorted equilibrium, operating a firm that is directly upstream or downstream from j increases π_j^{undist}.*
2. *In a distorted equilibrium, operating a firm that is directly upstream from j increases π_j^{dist}.*

As we will see, this proposition has important implications for the structure of the production network and the propagation of firm-level shocks through the production network.[33]

5.2 The structure of the production network

We know describe how the mechanisms of the model affect the structure of the production network through their impact on operating decisions.

[33] Proposition 9 does not hold when equilibrium quantities are not taken as given, in the sense of footnote 32. For instance, if there are only a few firms in the economy and that the operating cost f is high, operating an extra firm might put enough pressure on the wage that the operating profits of its neighbors would fall.

The impact of productivity on operating decisions

The productivity vector z plays an important role for the operating decisions of the firms. From (19) we see that a more productive firm benefits from a smaller unit cost δ_j. As a result, it sells more to the household and to other intermediate inputs, makes more profits and is more likely to operate.

A similar logic applies to groups of connected firms, but in that case an increase in the productivity of one producer j also affects the operating decisions of its neighbors. Consider for instance one of j's customer k. Since j is more productive it can now sell at a lower price which, in turn, reduces k's unit cost δ_k. It follows that k's operating profits increase as well, and that an increase in productivity has a positive impact on the incentives to operate of downstream firms. A similar mechanism applies to j's suppliers in the undistorted equilibrium. When j becomes more productive, it sells more units and therefore demands more from its suppliers. Given that each unit is sold at a mark up, this leads to higher operating profits for the suppliers. It follows that an increase in productivity has a positive impact on the incentives to operate of upstream firms in the undistorted equilibrium. This channel is absent from the distorted equilibrium since firm-to-firm sales are priced at cost in that case.

The following proposition formalizes that discussion in the context of the efficient allocation.

Proposition 10. *Let $\mathcal{J} \subset \mathcal{N}$ be a group of firms. Denote by $\theta^+ \in \{0,1\}^n$ the operating vector when the firms in $\mathcal{J}$ operate ($\theta_j^+ = 1$ for $j \in \mathcal{J}$). Similarly, let $\theta^- \in \{0,1\}^n$ be the operating vector when the firms in $\mathcal{J}$ do not operate ($\theta_j^- = 0$ for $j \in \mathcal{J}$). For all $j \notin \mathcal{J}$, assume $\theta_j^+ = \theta_j^-$. Denote by z^+ and z^- two productivity vectors such that $z_j^+ \geq z_j^-$ for all $j \in \mathcal{J}$ and $z_j^+ = z_j^-$ for $j \notin \mathcal{J}$. Then*

$$C_{z^+}\left(\theta^+\right) - C_{z^+}\left(\theta^-\right) \geq C_{z^-}\left(\theta^+\right) - C_{z^-}\left(\theta^-\right),$$

where $C_z\left(\theta\right)$ denotes consumption under the productivity vector z and the operating vector θ.

This result shows that the benefit to the planner, in terms of utility, of operating a group of firms increases with the productivity of the group's members. It follows that groups of firms with higher productivity are more likely to operate in the efficient allocation. The same forces are at work in undistorted and stable equilibria.

Highly connected groups of firms are more likely to operate

One immediate consequence of Proposition 9 is that firms with more operating suppliers and customers are more likely to operate in equilibrium. It follows that groups of firms that have a large number of potential connections in Ω are more likely to operate. Operating an extra firm in such a group adds an active neighbors to many producers which increases their incentives to operate. If some of these producers become active they, in turn, add an active neighbor to multiple firms, and so on.

To better understand this dynamic, it is useful to consider an example.

Example. Suppose that the planner wants to operate a total of n identical firms and can do so by operating groups of m firms that are fully-connected (so that each producer supplies to the $m-1$ other firms in the group) and with no connections outside the group.[34] Which value m is preferred by the planner? We can easily compute the productivity q of each firm in an m-group. Assuming that firms share the same productivity $z=1$, we find $q=(m-1)^{\frac{\alpha}{1-\alpha}\frac{1}{\varepsilon-1}}$ from (6). From (8), we can compute GDP as

$$C = n^{\frac{1}{\sigma-1}} (m-1)^{\frac{\alpha}{1-\alpha}\frac{1}{\varepsilon-1}} (1-fn) L. \tag{26}$$

This expression is increasing in m, and so the planner prefers, all else equal, to operate a unique tightly connected group of firms instead of a large number of connected pairs of firms. Equation (26) also shows that this preference for more connected groups is stronger when the share α of intermediate inputs in production is large and when the elasticity of substitution between these inputs ε is low. Under these conditions, a firm values additional suppliers more, which magnifies the impact of an additional connection.

The following proposition formalizes the intuition from that example and shows that the benefit to the planner from operating a given group of firms is greater when there are more potential connections between them. These groups are then more likely to operate.

Proposition 11. *Let $\mathcal{J} \subset \mathcal{N}$ be a group of firms. Denote by $\theta^+ \in \{0,1\}^n$ the operating vector when the firms in $\mathcal{J}$ operate ($\theta_j^+ = 1$ for $j \in \mathcal{J}$). Similarly, let $\theta^- \in \{0,1\}^n$ be the operating vector when the firms in $\mathcal{J}$ do not operate ($\theta_j^- = 0$ for $j \in \mathcal{J}$). For all $j \notin \mathcal{J}$, assume $\theta_j^+ = \theta_j^-$. Denote by Ω^- a network of potential connections and let Ω^+ be identical to Ω^- except that it has an additional connection between two firms in $\mathcal{J}$.*[35] *Then*

$$C_{\Omega^+}\left(\theta^+\right) - C_{\Omega^+}\left(\theta^-\right) \geq C_{\Omega^-}\left(\theta^+\right) - C_{\Omega^-}\left(\theta^-\right),$$

where $C_\Omega(\theta)$ denotes consumption under the potential network Ω and the operating vector θ.

Clustering of economic activity

The complementarities between firms also push for the clustering of economic activity in the efficient allocation. Since having multiple suppliers and customers is valuable, the planner prefers

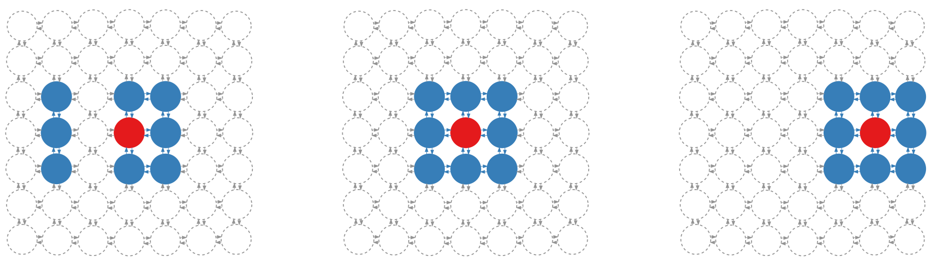
[34] For simplicity, assume that n divisible by m. While the current example might seem abstract it fits in the framework outlined in the last section. The whole economy would be formed of groups of n firms which various degree of clustering. The exercise would then be about which of these groups the planner prefers to operate.

[35] Formally, there are two firms $k \in \mathcal{J}$ and $l \in \mathcal{J}$ such that $\Omega_{ij}^+ = \Omega_{ij}^-$ for all pairs $(i,j) \neq (k,l)$ and $\Omega_{kl}^+ > \Omega_{kl}^-$.

to create networks in which many firms have many neighbors. Organizing production in this way involves building tightly connected clusters of operating firms. In this type of configuration, each firm j benefits from the presence of multiple suppliers, which increases its labor productivity q_j, and contributes to the labor productivity of its many customers. As a result, these clusters reinforce the productivity of all its members and makes organizing production in this way particularly productive.

Once again, an example is helpful to understand the mechanisms at work.

Example. Consider an economy in which firms are located on a grid, as shown in Figure 4. Firms are identical except for their position in the Ω network and that the red firm is slightly more productive.[36] In this case, it is never efficient to have more than one cluster of active firms, as in panel (a). Indeed, by grouping the two clusters together (panel b), the planner provides additional suppliers to some producers, which increases their labor productivity q and, as result, aggregate consumption. Clustering activity is therefore efficient.



(a) Multiple clusters are inefficient

(b) Efficient clustering

(c) Small change, large reorganization

Figure 4: Clustering in a grid network

In Figure 4, firms are essentially identical but the tendency to cluster remains even when firms differ, for instance in terms of their productivity z. Figure 5 shows the efficient allocation in the same economy as that of Figure 4 but with idiosyncratic productivity shocks z. In this case, the planner tends to cluster activity around the most productive firm, as in the first and second panel. If two distant firms have high productivity, organizing multiple clusters might be the optimal way of organizing production, as in the third panel.

The role of the elasticities of substitution

The elasticities of substitution in the aggregators for final consumption (σ) and intermediate inputs (ε) play an important role in shaping the production network. Figure 6 shows the efficient network in four economies that differ only in terms of σ and ε. In panel (a) both elasticities are large. Since firms are essentially producing the same good, the planner prefers to focus production

[36] The slightly more productive firm is needed to break the symmetry and position the cluster of operating firms. Without it, this configuration is still optimal but other solutions exist with the cluster simply translated on the grid.

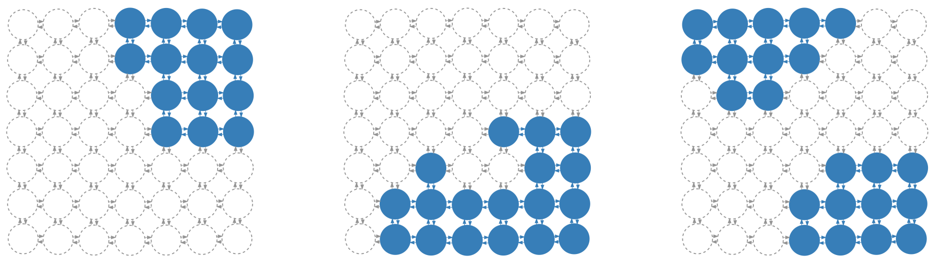
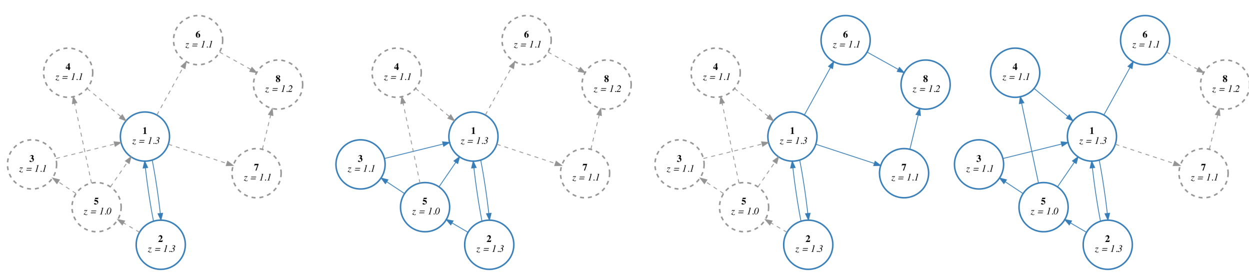


Figure 5: Clustering with three random draws of productivity z

in the hands of a small group of very productive producers (firms 1 and 2). The situation is different in panel (b) where the elasticity ε is small. In this case, goods are poor substitutes when they serve as intermediate inputs and additional suppliers are more valuable. The planner therefore provides additional inputs to firm 1 to increase its labor productivity q. If, instead, ε remains large but σ is small, as in panel (c), goods are poor substitutes in the consumption aggregator. The household prefers to enjoy a wide variety of products and, as a result, the planner operates producers that are downstream from firm 1. These firms can then take advantage of 1's high labor productivity to provide the household with cheap additional goods. When both elasticities are small, as in panel (d), the planner moves on both margins to operate some additional downstream and upstream producers.



(a) Large ε, Large σ (b) Small ε, Large σ (c) Large ε, Small σ (d) Small ε, Small σ

Figure 6: The impact of σ and ε on the production network

5.3 The impact of shocks

We now investigate the impact of idiosyncratic shocks in this economy. We first describe how cascades of firm shutdowns can occur, and how even a small productivity shock can trigger an important reorganization of the production network. Finally, we look at how the endogenous reorganization of the network affects GDP.

Cascades of firm shutdowns

One immediate consequence of the complementarities in operating decisions is that cascades of firm shutdowns can occur. Consider for instance a firm j in an undistorted equilibrium that stops production after suffering from a severe z shock. In response, its first neighbors, having lost a supplier or a customer, see their operating profits decline (Proposition 9) and are also more likely to shut down. But then j's second neighbors are also losing a neighbor and are at a greater risk of shutting down themselves. Since the same logic applies to further neighbors of j, the initial z shock can trigger a wave of shutdowns that spreads through the production network, both upstream and downstream.

The magnitude of these cascades depends on the strategic importance of firm j in the production network. If j purchases from many firms or has many customers, its exit affects the profits of many neighbors and is likely to trigger multiple shutdowns. On the other hand, Proposition 9 makes clear that firms with multiple neighbors have, all else equal, higher profits and would therefore remain in operation even after large adverse shocks. The model therefore predicts a negative correlation between the likelihood of a firm shutting down and the magnitude of the cascade it triggers upon exit. We will see in the next section that this correlation is also visible in U.S. data.

Cascades of shutdowns can also arise in a distorted equilibrium but the propagation mechanism is different. In that case, Proposition 9 tells us that complementarities operate only from supplier to customer, which implies that the exit of a firm might trigger the exit of its customers, but not the other way around.

Small shocks can lead to large reorganizations

One perhaps unusual feature of the model is that a small change in the environment can trigger a large reorganization of the network. To understand this process, it is helpful to consider the incentives of the planner. When designing the network, the planner compares the 2^n vectors θ in the set $\{0,1\}^n$ and selects the one providing the highest welfare. As, say, a firm's TFP z declines there is a point at which the planner shuts that firm down. But because of the complementarities between neighbors, it might be better to shut down the whole cluster around that firm and to move production elsewhere. In this case, the network might go through a large reorganization.

Figure 4 provides an example. Recall that in panel (b), all firms are identical, except for their potential links in the Ω network and the fact that the red firm is slightly more productive. In panel (c), another firm (in red) becomes slightly more productive than its peers. While the change in z is negligible, it triggers a large reorganization of the network, with several firms becoming active or inactive. Although output is barely affected by this reorganization, firm-level distributions can change substantially. In Appendix E, we provide another example in which the dispersion in labor productivity, output and employment across firms collapses after a small shock.

The reorganization of the network amplifies good shocks and dampens bad ones

Lastly, we look at the impact of the endogenous formation of the network on aggregate fluctuations. For that purpose, we consider the impact of a productivity shock on consumption when the network is flexible, in the sense that the vector θ can react to the shock, and when the network is kept fixed. The following proposition shows that the reorganization of the network amplifies the impact of positive shocks and mitigates the impact of negative shocks.

Proposition 12. *Let $\theta^*(z)$ be the efficient allocation under z and let $C(\theta, z)$ be consumption under (θ, z). Then the response of consumption after a change in productivity from z to z' is such that*

$$\underbrace{C\left(\theta^*\left(z'\right), z'\right) - C\left(\theta^*\left(z\right), z\right)}_{\text{Change in consumption under the flexible network}} \geq \underbrace{C\left(\theta^*\left(z\right), z'\right) - C\left(\theta^*\left(z\right), z\right)}_{\text{Change in consumption under the fixed network}}.$$

Under a flexible network, the planner is free to reorganize the production network to take advantage of the new productivity vector z'. For instance, clusters of firms that were built around formerly productive firms can be dismantled and the freed resources can be reallocated to producers that are now more productive. As a result, the negative impact of an adverse shock can be limited, and the positive impact of a beneficial shock can be amplified.

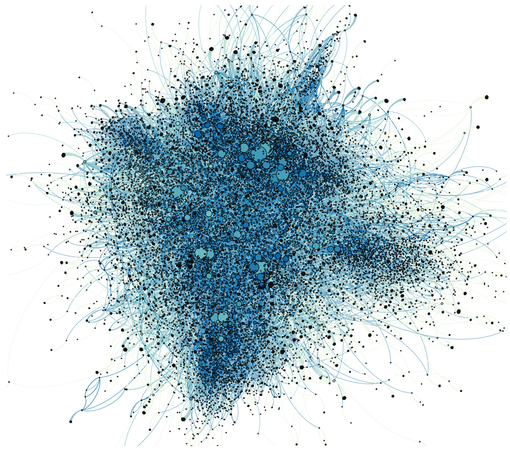
6 Quantitative exploration

This section provides a basic calibration of the model and shows that it captures salient features of the data such as cascades of firm shutdowns and movements in the structure of the production network over the business cycle. While we mostly focus on the efficient allocation, we can think of this allocation as the outcome of either an undistorted equilibrium or as a stable equilibrium, as defined in Appendix D. Toward the end of this section, we also consider a distorted equilibrium to evaluate the quantitative importance of the inefficiencies for the structure of the production network and the propagation of cascades.

6.1 Data

We calibrate the model using detailed firm-level input-output data from the Factset Revere Supply Chain Relationships dataset, which provides annual data from 2003 to 2016. These data are gathered by analysts from a variety of sources such as 10-Q and 10-K fillings, annual reports, investor presentations, websites, press releases, etc. In an average year, the sample includes almost 13,000 private and public firms and more than 40,000 relationships. Figure 7 shows the firm-level production network of the U.S. economy for a typical year.

We can use these data to evaluate how much of the variation in the production network can be explained by the firms' extensive margin of operation. In an average year, more than 40% of



Notes: Each circle is a firm, and the size of the circle reflects the number of connections. Darker colors denote higher local clustering coefficients. In 2016, Walmart had the largest indegree (448) and Microsoft had the largest outdegree (332). Image generated using Gephi with the Yifan Hu layout. There are 20,702 firms and 62,474 links. Vector image, zoom in.

Figure 7: 2016 Factset Revere U.S. firm-level production network

all link destructions occur at the same time as either the supplier or the customer (or both) stops producing. This number suggests that the firms' extensive margin plays a sizable role in shaping input-output relationships.[37]

To verify the robustness of some empirical patterns, we also rely on Compustat as another source of annual data. Compustat gathers information about a firm's major customers, defined as buyers of more than 10% of total sales, from annual financial statements. Since firms are not required to report smaller customers, we rarely see a firm supplying to more than 10 clients in the data.[38] Another limitation is that the names of the customers are self-reported, so General Motors might enter the database as "General Motors", "GM", "General Mtrs", etc. To address this issue, Cohen and Frazzini (2008) (CF) and Atalay et al. (2011) (AHRS), have used a combination of automatic algorithms and manual matching to properly identify each firm and to construct the production networks. Their samples cover the periods 1980 to 2004 and 1976 to 2009.

The Compustat dataset cover fewer firms than Factset—about 1,300 firms and 1,500 relationships in an average year—but over a longer time period. As it is available since 1976 it provides a more accurate picture of the evolution of the production network over the business cycle.

6.2 Parametrization

We normalize $A = 1$ and $L = 1$. The log of the productivities z_{it} are drawn from independent AR(1) processes with persistence ρ_z and a standard deviation σ_z for the ergodic distribution. Foster

[37]The analogous exercise for link creations finds a similar number. To remove high-frequency gaps in the data, we assume that a link is created during the first year it appears in the dataset and is destroyed during its last. A firm is considered as shutting down during the last year that it is in the sample.

[38]As a result, the tail of the outdegree distribution in Compustat is likely to be artificially thinner. Factset also relies on these data but supplements them using a variety of other sources.

et al. (2008) and Bartelsman et al. (2013) find that firm-level physical productivity in the United States has a standard deviation of 0.39 and a persistence of 0.81. We set $\sigma_z = 0.44$ and $\rho_z = 0.935$ so that measured TFP in the model matches these targets.[39] Since the model itself is static, the persistence in z_{it} is the only inter-temporal linkage in the economy. The idiosyncratic shocks to z_{it} will also generate aggregate fluctuations because of the finite number of firms.

There is no consensus in the literature about the cost of overhead labor f. Since employment in management occupations is about 5% of total employment, we set f so that $f \times n = 5\%$. For the number of firms, we set $n = 1000$ as a good trade-off between realism and computation time.[40]

For the share of intermediate goods, we follow Jorgenson et al. (1987) and Jones (2011) and set $\alpha = 0.5$. The empirical literature provides little guidance about the elasticity of substitution between intermediate inputs at the firm level. We therefore rely on Broda and Weinstein (2006) who estimate an elasticity of substitution between product varieties using import data. As these data do not differentiate between items used for consumption and as intermediate inputs, their estimates capture a mix of σ and ε. We set $\sigma = \varepsilon = 5$ as an average of their estimates and describe below how changes in these parameters affect the results.

We construct Ω by assuming that the number of *potential* incoming and outgoing connections, for any given firm, is drawn from a bivariate power law of the first kind. This family of distributions is entirely described by a single shape parameter ξ.[41] We set $\xi = 1.78$ so that the distribution of *active* incoming connections generated by the model is close to its empirical counterpart in the Factset data.[42] These two distributions are well approximated by power laws, with an exponent parameter of 0.97 for the empirical distribution (see Section 6.3 below). We therefore target that moment in the calibration. This indirect inference approach ensures that the calibrated economy is consistent with a key feature of the empirical production network.[43] To ensure that the results do not hinge on one particular matrix Ω, we randomly draw 20 different Ω's and, for each of them, simulate the economy for 100 periods.[44] The reported results are averages over these simulations.

[39] Firm-level TFP in the model is measured as in Foster et al. (2008) and Bartelsman et al. (2013). That measure controls, in particular, for the usage of intermediate inputs.

[40] See Appendix F.7 for simulations with $n = 20,000$ firms and aggregate shocks. The results are similar.

[41] The probability that a firm has x_{in} and x_{out} inbound and outbound links in Ω is $\xi\,(\xi - 1)\,(x_{in} + x_{out} - 1)^{-\xi - 1}$. The algorithm to construct Ω is in Appendix.

[42] We target moments from Factset, instead of Compustat, as it is the most comprehensive data source for linkages. We also target the indegree distribution, instead of its outdegree counterpart, as it is less affected by the 10% reporting threshold described in Section 6.1. Appendix F.1 shows how the production network is affected by ξ.

[43] All power law exponents are computed using the estimator of Gabaix and Ibragimov (2011).

[44] We discard and redraw simulations for which iterating on the first-order conditions does not converge to a point θ in $\{0, 1\}^n$. This rarely happens and, overall, the rejected networks do not look different. Keeping all the simulations in the sample yields very similar results.

6.3 Calibrated economy

Table 3 shows how the calibrated network compares to the U.S. data. We focus on five key moments to describe the overall structure of the network. The first three moments are the indegree, outdegree and eigenvector centrality distributions. In the model and in the data, these distributions are close to power laws so that their exponent parameters provide a good description of the full distributions. These exponents have an important influence for the aggregate impact of idiosyncratic shocks (Acemoglu et al., 2012). The fourth and fifth moments are the global clustering coefficient and the average distance between firms. Both measures describe how tightly connected firms are with one another—a key metric given the importance of clustering for productivity.[45]

We see from Table 3 that the calibrated economy (column 1), despite its simplicity, fits the Factset data (column 3) relatively well but there are some discrepancies with the Compustat datasets (columns 4 and 5), which is not surprising given its coverage. These discrepancies are particularly large when looking at the outdegree distribution and the clustering coefficient—a consequence of the 10% truncation threshold described above.[46]

Table 3: Production network in the calibrated economy and in the data

	(1)	(2)	(3)	(4)	(5)
	Model		Dataset		
	Calibrated	Neutral	Factset	Compustat	
				AHRS	CF
Power law exponents					
Indegree distribution	0.97	1.18	0.97	1.13	1.32
Outdegree distribution	0.92	1.15	0.83	2.24	2.22
Centrality distribution	1.16	1.23	0.59	0.08	0.06
Global clustering coefficient (normalized)	3.45	2.08	3.46	0.08	0.09
Average distance	2.64	3.04	4.81	1.06	1.04

Notes: To focus on the right tail, we truncate the eigenvector centrality distribution below the first quartile. Global clustering coefficients are computed on the undirected graph and multiplied by the square roots of the number of nodes. See footnote 45 for details. The average distance and the eigenvector centrality are computed on the undirected graph.

Figure 8 shows the degree distributions in the model and in the Factset data for 2016, the most recent year in the sample. To highlight the shape of these distributions, the figure uses a log-log scale and plots the complementary cumulative distributions (CCDF) on the vertical axis.

[45] The global clustering coefficient is computed on the undirected graph. It equals three times the number of triangles (three fully connected nodes) divided by the number of triplets (three connected nodes). In power law graphs, that coefficient declines naturally with n. Following Ostroumova Prokhorenkova and Samosvat (2014), we therefore normalize the means of the coefficients by multiplying them by the square root of the number of nodes. This normalization allows for a better comparison of networks across datasets.

[46] For a better comparison with Compustat, we can truncate the model-generated data using the 10% threshold. In that case, the exponent of the indegree, outdegree and centrality distributions are 1.06, 1.08 and 1.15. The clustering coefficient is 1.41 and the average distance is 3.14.

The roughly linear shapes confirm that they are close to power laws. As we can see, the model fits both distributions reasonably well.

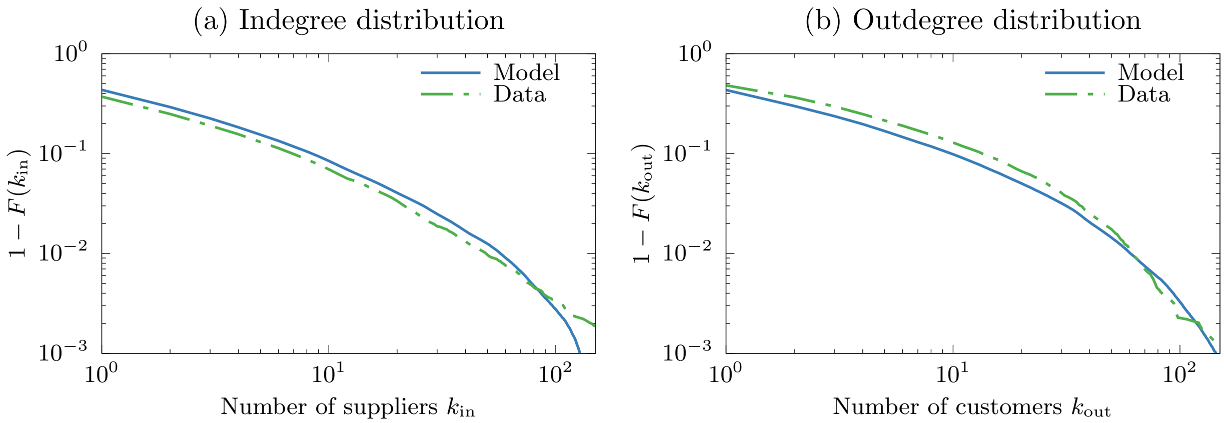


Figure 8: Distribution of the number of suppliers and customers

6.4 Comparison with a neutral network

To highlight which features of a network are desirable for efficiency, we can compare the calibrated network, which has been designed optimally by the planner, to a *neutral* benchmark built randomly by operating each firm with some probability $p > 0$.[47] All other quantities—except for the network itself—are chosen optimally by the planner. Since it is completely random, any discrepancies between the neutral benchmark and the efficient network are design decisions taken by the planner to improve efficiency.

The first two columns of Table 3 show how both networks differ. The power law exponents are smaller in the efficient network, indicating thicker tails than in the neutral benchmark. The efficient network therefore features a larger share of highly connected suppliers and customers, as well as relatively more high-centrality firms. The clustering coefficient is also larger, and the average distance is smaller, in the efficient network. These moments highlight that the planner prefers to organize production in tightly connected clusters of firms. Building the network in this way takes full advantage of the gains from input variety present in the environment.

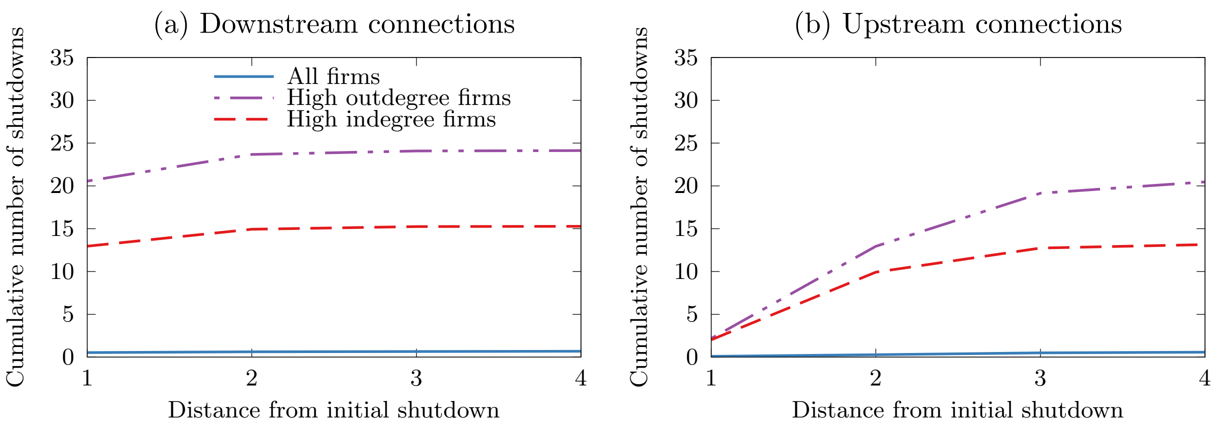
6.5 Cascades of firm shutdowns

We can use the calibrated model to evaluate how cascades of firm shutdowns arise and propagate through the network. To do so, we select a random active firm and set its productivity z to zero so that it stops production. We then compute the new efficient allocation and count the neighbors

[47] p is set so that both networks have the same number of active firms.

of the firm that shut down. Figure 9 shows the outcome of this exercise. The left panel looks at the firm's downstream neighbors, and the vertical axis shows the cumulative number of shutdowns as we move away from the shuttered firm. The right panel provides the same information but for upstream neighbors. The figure also differentiates between cascades originating from the average firm, and from firms with a high number of neighbors (above 99th percentile).[48]

We see that the shutdown of an average firm is likely to only create a small cascade: about 0.6 of its downstream neighbors, and even fewer of its upstream neighbors, shut down. But as we move to high-degree firms the cascades become larger: for important suppliers about 24 downstream neighbors are wiped by the cascade and the production network is extensively reorganized. Figure 9 also shows that cascades mostly propagate downstream, from customer to customer, instead of upstream. This is a consequence of the gains from input variety, embedded in equation (6), which make losing a supplier particularly costly in terms of productivity. That mechanism also implies that the exit of high outdegree producers is more damaging than that of high indegree ones.[49]



Notes: Cumulative number of exits at different distances from shuttered firm. "High degree" refers to the firms above the 99th percentile. Simulations of 100 randomly drawn matrices Ω, for each of which 1000 cascades are created.

Figure 9: Cumulative cascades by degree of originator

Welfare cost of cascades. We can also use the calibrated model to evaluate the welfare cost of cascades. Table 4 shows that the model generates a positive correlation between the size of a cascade and its negative impact on GDP. While the exit of an average firm has a negligible impact, a cascade that originates from a high-degree firm is responsible for a 2.4 percent drop in output on average. Firms with high outdegree—the star suppliers—have a disproportionate impact on aggregate output upon shutting down. Since they help to improve the productivity of many producers, their exit lowers the aggregate productivity of the network substantially.

Table 4 also shows that the elasticity of substitution ε between intermediate inputs matters for

[48] Appendix F.3 looks at cascades in the undirected network.

[49] Appendix F.2 shows that cascades that originate from high-degree firms have a much larger impact when $\varepsilon = 3$.

the impact of the cascades on GDP. When inputs are poor substitutes, the exit of a high-degree firm creates a bigger cascade (as seen in Figure 14 in the appendix) and these cascades have a larger negative effect on GDP.[50,51]

	Impact on GDP	
	Benchmark	$\varepsilon = 3$
Average firm	-0.1%	-0.1%
High indegree	-1.8%	-4.7%
High outdegree	-2.5%	-4.9%
High degree	-2.4%	-5.1%

Notes: "High degree" refers to firms above the 99th percentile. Simulations of 100 randomly drawn matrices Ω, for each of which 1000 cascades are created.

Table 4: Correlation between output drop and firm degree

Cascades in the model and the data. In the data, many firms are simultaneously hit by (unobserved) shocks and multiple cascades might overlap, so that there is no straightforward way to use the exercise above to evaluate the empirical performance of the model. We can, however, use simple regressions to capture the impact of an exiting firm on its neighbors. Since these regressions can be run in the data and in the calibrated economy, they provide a good test of the model's ability to generate realistic cascades.

Specifically, we compute the fraction of each firm j's neighbors that exit in a given period and regress that number on whether j itself shuts down. We run separate regressions for upstream and downstream neighbors at various distances from j. To be precise, denote by DX_{jdt} and UX_{jdt} the fraction of firm j's downstream and upstream neighbors located at a distance d that exit between t and $t+1$. We regress

$$\mathrm{DX}_{jdt} = \alpha^D + \beta_d^D \mathrm{Exit}_{jt} + \mathrm{Controls}_{jt} + \varepsilon_{jdt} \tag{27}$$

and

$$\mathrm{UX}_{jdt} = \alpha^U + \beta_d^U \mathrm{Exit}_{jt} + \mathrm{Controls}_{jt} + \varepsilon_{jdt} \tag{28}$$

where Exit_{jt} equals 1 if j exits between t and $t+1$ and 0 otherwise.[52] The coefficients β_d^D and β_d^U provide information about how cascades propagate in this economy. They capture the increase in

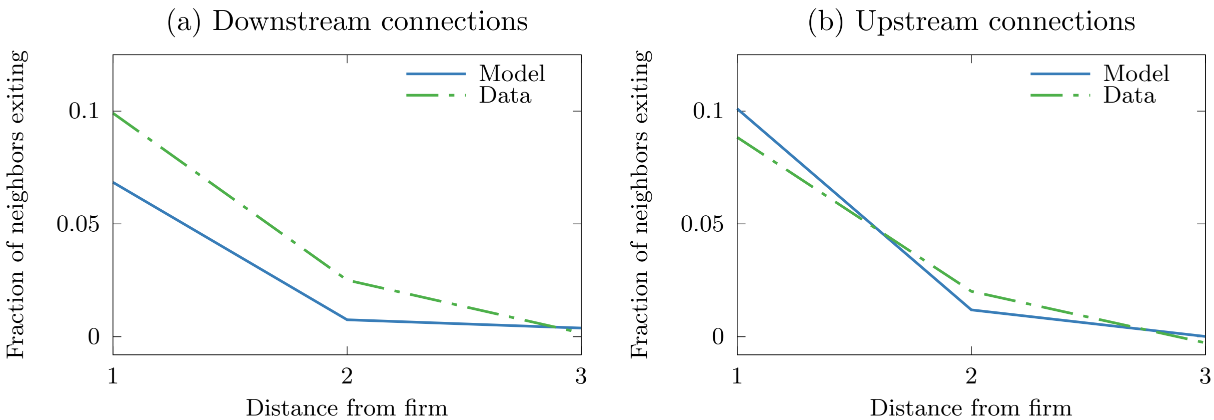
[50] In empirical work, Barrot and Sauvagnat (2016) have found that, similarly, a shock to a producer of a highly specific input is more damaging to its customers.

[51] Appendix F.4 shows the impact of the same shocks when the network is kept fixed. In that case, the impact on GDP is larger, particularly for high-degree firms and when the elasticity ε is small. This highlights the importance of the discrete adjustment margin (the vector θ) for the response to shocks.

[52] In the data, we consider that a firm shuts down during its last year in the sample. We use SDC Platinum to exclude mergers and acquisitions. The controls in (27) and (28) include the in- and outdegree of firm j.

shutdown probability associated with the exit of a neighboring firm located at a distance d.[53]

Figure 10 shows the coefficients estimated from the Factset data (green dashed lines). We see that the shutdown of a firm is associated with about a 10% increase in the probability that one of its direct suppliers or customers also exits. This number falls to about 2% for the second neighbors and keeps declining afterwards. The model (solid blue lines) is roughly able to match these patterns, suggesting that it broadly captures the joint operating decisions of nearby firms.[54]



Notes: Factset data. Estimated coefficients from regressing the fraction of exiting neighbors on whether a firm exits. Indegree and outdegree controls are included. The distance is the smallest number of connections between two firms.

Figure 10: Cascades of firm shutdowns in the model and in the data

While the shutdown of any firm has the potential to push a neighbor out of business, the exit of well-connected producers generally trigger larger cascades. To see this, we can first measure the size of a cascade as the total number of shutdowns, summed up to the second neighbors, associated with the exit of a firm, and then compare this statistic across firms with different numbers of neighbors.

The results are presented in Table 5. The first column shows that, in the data, firms that are above the 90th percentile of the degree distribution are associated with cascades that are about three times larger than those associated with the average firm. High-degree firms are, however, less likely to actually shut down in response to shocks, as the third column shows. In the data, an average firm has a 11.8% chance of exiting in a given year, while this number drops to 2.5% for a high-degree firm.[55]

The model does well in terms of the size of the cascades and is also able to roughly replicate

[53]Regressions (27) and (28) can suffer from endogeneity issues, such that β_d^D and β_d^U should not be interpreted as capturing a causal relationship. Nonetheless, they describe correlation patterns between the operating status of nearby firms and we can compare these patterns in the model and the data.

[54]One possibility is that the regressions (27) and (28) capture common shocks across firms instead of the propagation over the network. For instance, since trading partners are likely to be geographically close to each other, a local shock could directly affect both of them at the same time. To alleviate this concern, we run the same regressions on supplier/customer pairs located in different zip codes. Reassuringly, the results are essentially the same.

[55]The elasticities of substitution play an important role in determining whether a firm shuts down after an adverse z shock. A smaller ε makes the network more rigid and firms are less likely to shut down. A higher elasticity σ, as it makes inputs in the consumption aggregator more substitutable, leads to a higher likelihood of exit.

the exit probabilities. In the model, high-degree firms are particularly valuable to the planner and are therefore kept in operation even after severe shocks. When they do shut down, however, the planner reorganizes the whole cluster of producers that was built around them, which explains the large cascades that they trigger.

Table 5: High-degree firms are more resilient but create larger cascades

	Size of cascades		Probability of exit	
	Data	Model	Data	Model
Average firm	0.9	1.1	11.8%	11.3%
High-degree firm	3.0	4.3	2.5%	1.7%

Notes: "High degree firms" are above the 90th percentile of the degree distribution. "Size of cascades" is the sum of exiting firms up to the second neighbors downstream and upstream, computed by multiplying the regression coefficients in Figure 10 by the number of neighbors at the corresponding distance.

We can also look at these cascades in the distorted equilibrium, in which entry and exit decisions are inefficient. In Appendix F.5, we see that the distorted equilibrium cascades are similar to those in the efficient allocation except that they propagate relatively more downstream. This is because of the difference in complementarities described in Proposition 9 and Section 5.3. Equilibrium cascades also lead to larger welfare losses.

6.6 Aggregate fluctuations

There is a finite number of firms in the economy, so the idiosyncratic productivity shocks create aggregate fluctuations. Since these shocks also affect the production network, aggregate output is endogenously correlated with the structure of the network. We investigate that correlation in this section, and we also consider how the endogenous reorganization of the network amplifies or dampens fluctuations in macroeconomic aggregates.[56,57]

Comovements

Table 6 shows the correlations between GDP and the structure of the network in the calibrated economy and in the data. We see that in the model the exponent parameters of the degree and eigenvector centrality distributions are negatively correlated with output, which indicates thicker right tails, and thus an abundance of well-connected and high-centrality firms, during expansions. The economy also features more clustering and smaller average distances during booms. These correlations are similar in the data, although there are some discrepancies across datasets. The

[56]Proposition 14 in the appendix shows that aggregate shocks to A would have no effect on the network. We can therefore abstract from them to explore the interaction between GDP and the shape of the network.

[57]Appendix F.7 provides a version of the calibrated economy with a much larger number of firms and aggregate shocks. The results are similar.

model is closest to the Factset data, which provides the most comprehensive link coverage.[58,59]

Table 6: Correlation between network moments and GDP

	Model	Datasets		
		Factset	Compustat	
			AHRS	CF
Power law exponents				
Indegree distribution	−0.53	−0.87	−0.35	−0.12
Outdegree distribution	−0.63	−0.97	−0.31	−0.11
Centrality distribution	−0.10	−0.15	0.28	−0.37
Global clustering coefficient	0.60	0.76	0.18	0.11
Average distance	−0.82	−0.69	0.18	0.00

Notes: All time series are in logs. In the data, output is annual real GDP, detrended linearly in sample. Since there are only 13 years in the Factset data we use the CBO 10-year projection for real GDP growth at the beginning of the sample in 2003 (2.58%) to detrend the series. To focus on the right tail, we truncate the eigenvector centrality distribution below the first quartile. The global clustering coefficient, the average distance and the eigenvector centrality are computed on the undirected graph.

These patterns can be explained through the lens of the model. When well-positioned firms receive good shocks, the planner builds highly-connected clusters around them. As discussed before, these clusters are particularly productive, which generates the observed correlations between output, clustering and the degree distributions. Inversely, during recessions it might be too costly to organize these productive clusters—perhaps because a few critical firms face low z shocks. As a result, production is more dispersed and output is lower.

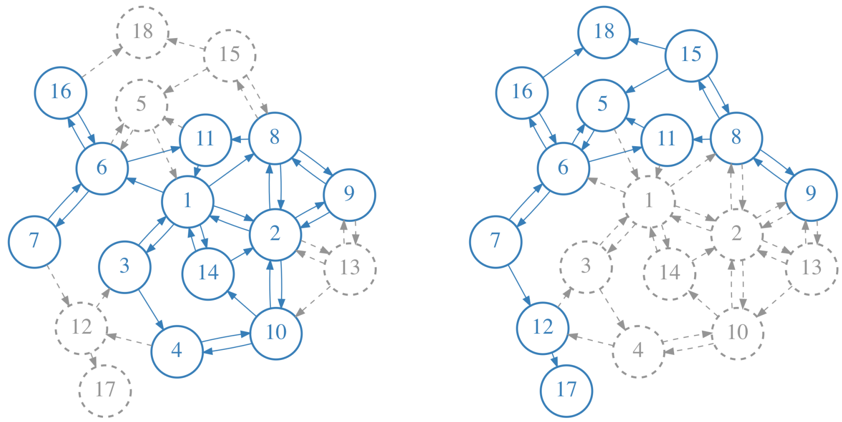
Figure 11 provides an example of these mechanisms in a simple economy under two realizations of the idiosyncratic shocks z. In panel (a), the shocks are such that the planner organizes a cluster around firms 1 and 2. Firms in this cluster have many connections and take advantage of the gains from input variety to improve their productivity. As a result, the clustering coefficient is large, the degree distributions have thick tails, as seen in panel (c), and the economy is booming. Panel (b) shows the same economy but under a realization of shocks such that operating a cluster around firms 1 and 2 is not efficient. In this case, economic activity moves to the outskirts of the network, where there are fewer connections between firms. As a result, the clustering coefficient is smaller, the degree distributions have thinner tails and the economy is in a downturn.

Level and volatility of output

We now consider the impact that the endogenous formation of the network has on the level and the volatility of aggregate output. To do so, it is useful to compare the efficient allocation, in which

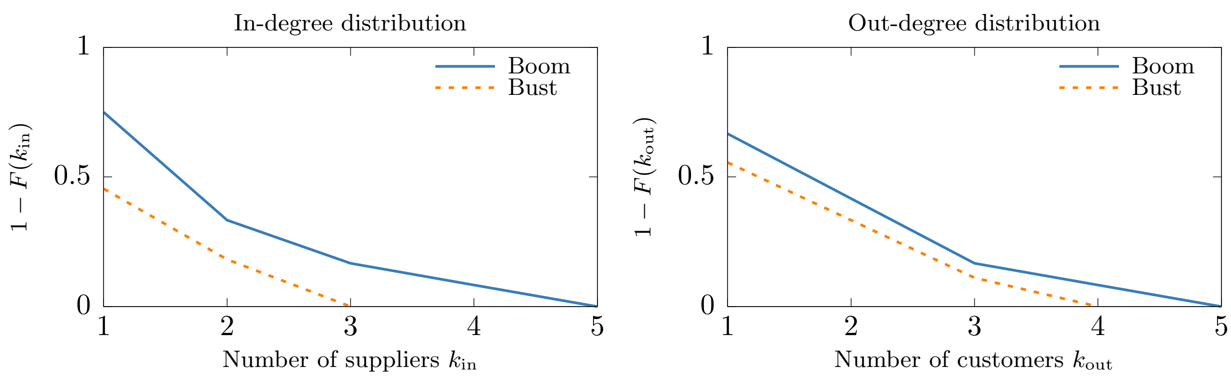
[58] If we truncate the model-generated data for a better comparison with Compustat, we find that the correlations between output and the indegree, outdegree and centrality distribution exponents are −0.48, −0.56 and −0.03. The correlations with the clustering coefficient and the average distance are 0.80 and −0.76.

[59] Table 21 in the appendix shows that these correlations are similar under different elasticities σ and ε.



(a) Boom: $C = 3.3$, clustering coefficient $= 0.37$

(b) Bust: $C = 2.9$, clustering coefficient $= 0.14$



(c) Degree distributions

Figure 11: Booms and busts in the production network

the network is constantly reorganized in response to shocks, to an alternative economy in which the network is designed efficiently in the first period but then kept completely fixed afterward. The differences between these two economies capture the role played by the endogenous response of the network to shocks.[60]

We find large differences between these two economies. First, aggregate output is 11% lower when the network is kept fixed, which suggests that frictions that might impede the reorganization of the network can have large welfare consequences. Second, aggregate output is 17% more volatile when the network is fixed, which shows the importance of the endogenous evolution of the network for the aggregation of firm-level shocks into macroeconomic fluctuations.[61,62]

[60]The differences between these economies are also informative about the importance of the discrete margin of adjustment, the vector θ, for the level and volatility of GDP.

[61]These numbers are similar in an economy with a larger number of firms and aggregate shocks (Appendix F.7). This suggest that the network adjustment margin (or the discrete margin) remains important even in large economies. See Appendix F.8 for the impact of the elasticities of substitution on the difference between the flexible and the fixed network economies.

[62]To get a sense of magnitude for this last number we can also consider an economy in which, in addition to the

To understand how the reorganization of the network dampens fluctuations, it helps to think of the planner as choosing, for each productivity vector z, the best network θ out of 2^n possibilities. Total production $C(z)$ can therefore be written as

$$C(z) = \max_{k \in \{1,\ldots,2^n\}} C_k(z)$$

where $C_k(z)$ is GDP under the kth network. By itself, each network k is associated with a probability distribution for C_k where the randomness comes from the underlying shocks z. The mean and the variance of these distributions vary with k, but for the networks that are actually selected by the planner the differences are limited and the distributions overlap substantially. Figure 12 provides an example with five potential networks. Each line represents the PDF of GDP under a fixed network k. The figure also shows the output produced by each network under four different productivity vectors z. Each z is associated with a symbol whose locations on the graph represent the output $C_k(z)$ of network k under shock z. For instance, we see that network $k = 1$ performs poorly under the ■ shock, while network $k = 5$ performs well. The symbols in blue indicate which network performs the best under a given z and is therefore chosen by the planner. We see that, for any fixed network, the PDFs are spread out and the variance of output is relatively large. In contrast, the output produced by the best network—the blue symbols in the right tails of the distributions—are close to each other indicating that the variance of output $C_{k^*}(z)$ under the efficient network $k^*(z) = \arg\max C_k(z)$ is relatively small.[63]

6.7 The inefficient equilibrium and the data

So far, the quantitative exploration of the model has focused on the efficient allocation, but in reality market power and other distortions might play a role in shaping the production network. To have a better sense of the quantitative importance of these frictions, we now solve for the distorted equilibrium introduced in Section 4. Remember that in that equilibrium entry decisions and the overall structure of the production network are distorted away from the efficient allocation.[64]

Table 7 presents various moments that describe the network in the distorted equilibrium. It also

network itself, all the inputs of the firms are kept fixed. In this case output volatility doubles compared to the flexible network benchmark. On its own, the endogenous formation of the network is therefore able to explain about one fifth of the reduction in volatility generated by all the adjustment margins together.

[63] This intuition is reminiscent of results from extreme value theory that show that the variance of the maximum of a large number m of independent normal random variables declines rapidly with m. This result does not hold when the underlying random variables have fat tails. Acemoglu et al. (2015) and Baqaee and Farhi (2017a) explore network models in which GDP can have fat tails. In this case, it might be possible for the flexible network economy to be *more* volatile than its fixed-network counterpart.

[64] One potential issue here is the presence of multiple equilibria. To evaluate their importance, we have solved the model starting from different initial conditions. We do occasionally find multiple equilibria but they are in general very similar to each other.

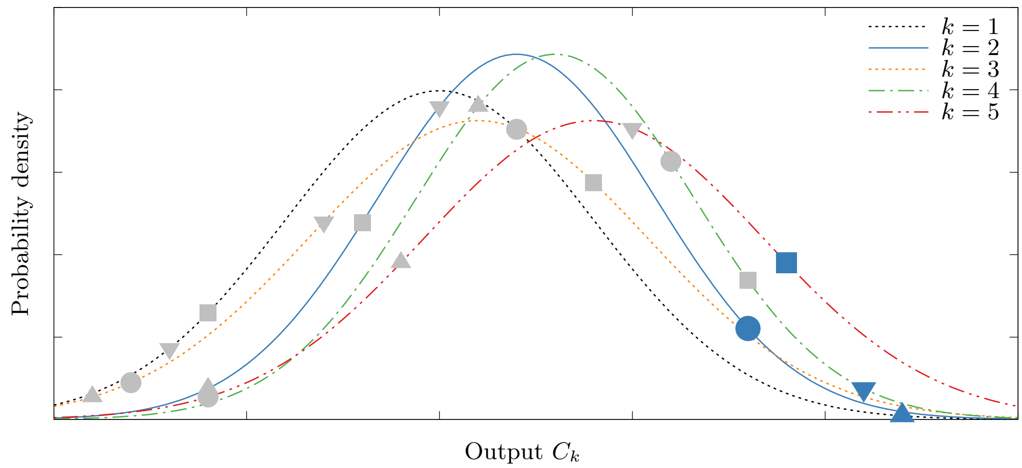


Figure 12: Distribution of output C_k under different networks k

reports the equivalent numbers in the efficient allocation for comparison. The first two columns show that the power law exponents of the degree and centrality distributions are roughly similar under both allocations. It is notable, however, that these exponents are all smaller in the efficient allocation, such that the equilibrium features fewer highly-connected firms. Similarly, firms in the equilibrium are less clustered together and they are further away from each other. In Section 5 we explored how organizing clusters of firms was an effective way of improving the productivity of the network. Table 7 suggests that the distortions at work in the environment impede the emergence of that type of organization in equilibrium.

The last two columns of Table 7 describe how the networks evolve over the business cycle. As we can see, the correlations are similar in the two allocations. They are however smaller (in absolute value) in the equilibrium, indicating that this network is more rigid and less able to adapt to take advantage of changing economic conditions.

Overall, Table 7 shows that while the efficient and the distorted networks are not identical, the differences are somewhat limited. Both allocations are also similar in terms of macroeconomic aggregates: GDP is 0.9% lower and 0.2% more volatile in the equilibrium. This suggests that the key forces that shape the efficient allocation are at work in the distorted equilibrium—not only qualitatively but also quantitatively. While other distortions might lead to more important departures from the efficient allocation, our findings in this section suggest that the efficient allocation might be able to provide a reasonable approximation to some distorted economies.

Table 7: Production networks in the efficient allocation and the distorted equilibrium

	Average		Correlation with GDP	
	Efficient	Equilibrium	Efficient	Equilibrium
Power law exponents				
indegree distribution	0.97	1.02	−0.53	−0.18
outdegree distribution	0.92	0.95	−0.63	−0.46
Centrality distribution	1.16	1.22	−0.10	0.03
Global clustering coefficient	3.45	2.99	0.60	0.40
Average distance	2.64	2.66	−0.82	−0.65

Notes: "Equilibrium" refers to the distorted equilibrium. To focus on the right tail, we truncate the eigenvector centrality distribution below the first quartile. Global clustering coefficients are computed on the undirected graph and multiplied by the square roots of the number of nodes. See footnote 45 for details. The average distance and the eigenvector centrality are computed on the undirected graph.

7 Conclusion

This paper proposes a theory of network formation that operates through the firms' extensive margin of production. Because of complementarities between firms' operating decisions, production tends to be organized in tightly connected clusters centered around the most productive firms. The complementarities also give rise to cascades of firm shutdowns. As in the data, highly connected firms are more resilient to shocks but trigger larger cascades upon shutdown. The theory also predicts how the shape of the network comoves with GDP. Expansions feature more clustering among firms, as well as thicker tails for the degree distributions. These correlations are also present in U.S. data. Finally, the theory predicts that the optimal reorganization of the network is responsible for a substantial decline in the size of aggregate fluctuations. These findings highlight the importance of endogenizing the production network to better understand the origin of aggregate fluctuations.

The paper considers the role of some form of market power on the equilibrium allocation. Clearly many other types of externalities, coordination problems or market frictions might be at work in reality. Studying the importance of these distortions and figuring out which ones, if any, matter quantitatively is an important topic for future research. Exploring fully granular equilibrium definitions could also provide interesting insights into these models.

The model also opens the door to future research projects such as the study of industry clusters, like Silicon Valley for technology, and New York or London for finance. In particular, the theory could be used to quantify how well-connected, high-productivity firms affect other participants' incentives to join the cluster, and therefore shed light on how these clusters are created in the first place. Finally, the model could be extended to provide a quantitative framework that could guide policymakers when considering whether they should bail out distressed firms. For that purpose, including a broader set of inefficiencies might be desirable.

References

Acemoglu, D., and P. D. Azar (2018): "Endogenous Production Networks," Working Paper 24116, National Bureau of Economic Research.

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80, 1977–2016.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015): "Microeconomic Origins of Macroeconomic Tail Risks," Working Paper 20865, National Bureau of Economic Research.

Acemoglu, D., and A. Tahbaz-Salehi (2020): "Firms, Failures, and Fluctuations: The Macroeconomics of Supply Chain Disruptions," *Working Paper*.

Atalay, E. (2017): "How Important Are Sectoral Shocks?" *American Economic Journal: Macroeconomics*, 9, 254–80.

Atalay, E., A. Hortaçsu, J. Roberts, and C. Syverson (2011): "Network structure of production," *Proceedings of the National Academy of Sciences*, 108, 5199–5202.

Baqaee, D. R. (2018): "Cascading Failures in Production Networks," *Econometrica*, 86, 1819–1838.

Baqaee, D. R., and E. Farhi (2017a): "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem," Working Paper 23145, National Bureau of Economic Research.

——— (2017b): "Productivity and Misallocation in General Equilibrium.," Working Paper 24007, National Bureau of Economic Research.

Barrot, J.-N., and J. Sauvagnat (2016): "Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks," *The Quarterly Journal of Economics*, 131, 1543–1592.

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.

Bigio, S., and J. La'O (2016): "Financial Frictions in Production Networks," Working Paper 22212, National Bureau of Economic Research.

Bloom, P. N., and V. G. Perry (2001): "Retailer power and supplier welfare: The case of Wal-Mart," *Journal of Retailing*, 77, 379–396.

Boehm, J., and E. Oberfield (2018): "Misallocation in the Market for Inputs: Enforcement and the Organization of Production," Working Paper 24937, National Bureau of Economic Research.

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.

Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2017a): "The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy," working paper.

Caliendo, L., F. Parro, and A. Tsyvinski (2017b): "Distortions and the Structure of the World Economy," Working Paper 23332, National Bureau of Economic Research.

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 A. Tahbaz-Salehi (2018): "Production networks: A primer."

Carvalho, V. M., and N. Voigtländer (2015): "Input Diffusion and the Evolution of Production Networks," Working Paper 20025, National Bureau of Economic Research.

Chahrour, R., K. Nimark, and S. Pitschner (2019): "Sectoral Media Focus and Aggregate Fluctuations," *Available at SSRN 3477432*.

Cohen, L., and A. Frazzini (2008): "Economic Links and Predictable Returns," *The Journal of Finance*, 63, 1977–2011.

Dupor, B. (1999): "Aggregation and irrelevance in multi-sector models," *Journal of Monetary Economics*, 43, 391 – 409.

Elliott, M., B. Golub, and M. V. Leduc (2020): "Supply network formation and fragility," *Available at SSRN 3525459*.

Foerster, A. T., P.-D. G. Sarte, and M. W. Watson (2011): "Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production," *Journal of Political Economy*, 119, 1–38.

Foster, L., J. Haltiwanger, and C. Syverson (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *The American Economic Review*, 98, 394–425.

Gabaix, X. (2011): "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79, 733–772.

Gabaix, X., and R. Ibragimov (2011): "Rank - 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," *Journal of Business and 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 and Co.

di Giovanni, J., A. A. Levchenko, and I. Mejean (2014): "Firms, Destinations, and Aggregate Fluctuations," *Econometrica*, 82, 1303–1340.

Grassi, B. (2017): "IO in I-O: Size, Industrial Organization and the Input-Output Network Make a Firm Structurally Important," working paper.

Hatfield, J. W., S. D. Kominers, A. Nichifor, M. Ostrovsky, and A. Westkamp (2013): "Stability and Competitive Equilibrium in Trading Networks," *Journal of Political Economy*, 121, 966–1005.

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.

Jackson, M. O., and A. Wolinsky (1996): "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71, 44 – 74.

Jones, C. I. (2011): "Intermediate Goods and Weak Links in the Theory of Economic Development," *American Economic Journal: Macroeconomics*, 3, 1–28.

Jorgenson, D. W., F. Gollop, and B. Fraumeni (1987): *Productivity and U.S. Economic Growth*: Harvard University Press, Reprinted- Universe, 1999.

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.

König, M. D., A. Levchenko, T. Rogers, and F. Zilibotti (2018): "Aggregate Fluctuations in Adaptive Production Networks," working paper.

Kopytov, A., K. Nimark, B. Mishra, and M. Taschereau-Dumouchel (2022): "Endogenous Production Networks under Supply Chain Uncertainty," *Available at SSRN 3936969.*

Li, D., and X. Sun (2006): *Nonlinear integer programming* Volume 84: Springer Science.

Lim, K. (2018): "Endogenous Production Networks and the Business Cycle," working paper.

Liu, E. (2019): "Industrial policies in production networks," *The Quarterly Journal of Economics*, 134, 1883–1948.

Long, J. B., and C. I. Plosser (1983): "Real Business Cycles," *Journal of Political Economy*, 91, 39–69.

Murray, W., and K.-M. Ng (2010): "An algorithm for nonlinear optimization problems with binary variables," *Computational Optimization and Applications*, 47, 257–288.

Oberfield, E. (2018): "A Theory of Input-Output Architecture," *Econometrica*, 86, 559–589.

Ostroumova Prokhorenkova, L., and E. Samosvat (2014): *Global Clustering Coefficient in Scale-Free Networks*, 47–58, Cham: Springer International Publishing.

Ozdagli, A., and M. Weber (2017): "Monetary Policy through Production Networks: Evidence from the Stock Market," working paper.

Schaal, E., and M. Taschereau-Dumouchel (2015): "Coordinating Business Cycles," working paper.

Tintelnot, F., A. K. Kikkawa, M. Mogstad, and E. Dhyne (2018): "Trade and Domestic Production Networks," Working Paper 25120, National Bureau of Economic Research.

Van Riper, T. , "The Wal-Mart Squeeze," *Forbes*, https://www.forbes.com/2007/04/23/walmart-suppliers-margins-lead-cx_tvr_0423walmart.html.

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, 10.1002/j.2158-1592.2008.tb00081.x.

Zurich Insurance Group (2015): "Potential business impact of loss of the main supplier for small and medium enterprises (SMEs) in 2015."

Online Appendix

A Efficient allocation in an economy with a fixed network

This section shows how to derive all the quantities in the efficient allocation for a given network θ. Lemma 1 already gives an expression for C. From an intermediate step in its proof, we can find the efficient allocation for labor l. Define Γ as the matrix with typical element

$$\Gamma_{jk} = \frac{\alpha_k \Omega_{jk} q_j^{\varepsilon_k - 1}}{\sum_{i \in \mathcal{N}} \Omega_{ik} q_i^{\varepsilon_k - 1}}.$$

Then, (41) defines a linear system in l. Inverting this system yields

$$l = \left[(I_n - \Gamma) \operatorname{diag} \left(\frac{1}{1-\alpha} \right) \right]^{-1} \left(\beta \circ \left(\frac{q}{Q} \right)^{\circ(\sigma-1)} \frac{C}{Q} \right), \tag{29}$$

where β denotes the vector of intensities β_j, $\operatorname{diag}\left(\frac{1}{1-\alpha}\right)$ denotes a matrix with the vector $\frac{1}{1-\alpha}$ on the diagonal and $\circ$ denotes Hadamard (element-by-element) operations such that $(A \circ B)_{ij} = (A)_{ij} \circ (B)_{ij}$ and $\left(A^{\circ B}\right)_{ij} = (A)_{ij}^{(B)_{ij}}$ for arbitrary matrices A and B. We q_j and l_j in hand, we can use the first-order condition (34) we can then write y_j as

$$y_j = (1 - \alpha_j)^{-1} q_j l_j.$$

Finally, we can rewrite the first-order condition (36) as

$$x_{ij} = \left(\frac{q_i}{q_j} \right)^{\varepsilon_j} \alpha_j \left(A z_j \theta_j \left(\frac{1}{q_j} \right)^{1-\alpha_j} \right)^{\frac{\varepsilon_j - 1}{\alpha_j}} \Omega_{ij} y_j$$

when $q_i, q_j > 0$ and $x_{ij} = 0$ otherwise.

B Additional numerical tests

This section provides the details of the numerical simulations of Section 3.4 as well as several additional exercises to show the robustness of the solution approach.

B.1 Details of the simulations of Table 1

The numerical simulations of Table 1 involve a large number of economies that are generated randomly from a broad set of parameters.

Aggregate parameters. The aggregate parameters are selected from: $n \in \{4, 6, 8, 10, 12, 14\}$ for the number of firms and $\sigma \in \{4, 6, 8\}$ for the elasticity of substitution in the consumption aggregator. The matrix Ω is such that each firm has on average 3, 4, ... up to n *potential* incoming connections (non-zero Ω_{ij}).[65] We restrict the matrices Ω to have empty diagonals, as in the data. Each non-zero element in Ω is drawn from $\Omega_{ij} \sim \text{iid } U([0, 1])$. Appendix G.1 describes the precise algorithm used to build Ω.

Firm-level parameters. The firm-level parameters are drawn from: $\log(z_k) \sim \text{iid } \mathcal{N}(0, 0.25^2)$ for the productivities, $f_j \sim \text{iid } U([0, 0.2/n])$ for the fixed costs, $\alpha_j \sim \text{iid } U([0.25, 0.75])$ for the intermediate input shares, $\varepsilon_j \sim \text{iid } U([4, 8])$ for the elasticities of substitutions between intermediate inputs, and $\beta_j \sim \text{iid } U([0, 1])$ for the factor intensities in the production of the final good.

Procedure. For every potential combination of the aggregate parameters we simulate 500 economies. In each case, the matrix Ω and the individual characteristics of the firms are drawn from the distributions described above. We then use the exhaustive search algorithm described in Appendix G.2 to compute the true solution to $\mathcal{P}$. We also use the algorithm of Appendix G.4 to compute solutions to the reshaped and non-reshaped versions of the planner's problem. These two solutions are then compared to the true solution and the results are reported in Table 1. We exclude from the simulations pathological cases in which the algorithms find an aggregate consumption of 0.[66] For the benchmark tests, an economy is kept in the sample only if the first-order conditions of the reshaped problem yield a solution in $\{0, 1\}^n$. Appendix B.3 shows that the algorithm performs well when these simulations are kept in the sample.

B.2 Alternative measure of success

Table 1, in the main text, presents the fraction of firms with the correct operating status θ_j. One can also use the fraction of economies in which the full vector θ is correct as an alternative measure of success. Table 8 provides that information. We see that with reshaping the algorithm finds the correct vector θ 99.7% of the time on average, which is much better than without reshaping, where the equivalent number is 19.9%.

The table also shows that the performance of the reshaping method degrades slightly as n increases. This is not surprising since the likelihood that any one firm will be assigned the wrong status increases with n. Since the importance of any one firm for aggregate outcomes also declines with n, the error in C remains small.

[65] The corresponding average numbers of *active* incoming connections are 2.1, 3.0, 3.8, 4.5, 5.3 and 5.8, respectively. See Appendix B.4 below for tests on very sparse matrices Ω.

[66] This happens, for instance, when Ω is so sparse that a closed-loop of suppliers does not exist.

Table 8: Using the full vector θ as a measure of success

	With reshaping		Without reshaping	
n	Correct vector θ	Error in C	Correct vector θ	Error in C
4	100%	0%	68.9%	0.502%
6	99.9%	0.000%	47.1%	0.692%
8	99.8%	0.001%	32.3%	0.791%
10	99.7%	0.001%	21.6%	0.855%
12	99.7%	0.001%	15.3%	0.903%
14	99.6%	0.001%	10.6%	0.928%

Notes: Same simulations as Table 1. See Section B.1 in the Appendix for details.

B.3 When the solution to the reshaped problem is not in $\{0, 1\}^n$

The results presented in Table 1 exclude economies in which $\mathcal{R}$'s first-order conditions are such that $0 < \theta_j < 1$ for at least one firm j, which happens in less than a tenth of the simulations. This section shows that, even when these cases are not excluded from the sample, the solution approach performs well. To see this, Table 9 shows the outcome of the same simulations as Table 1 but without excluding the economies such that $\theta \notin \{0, 1\}^n$. We see that on average the error in aggregate output C is less than 0.008% and about 99.7% of firms are assigned the correct status. In contrast, without reshaping the average error in output is 0.867%—108 times more.[67]

Table 9: Testing the reshaping on small networks without exclusions

	With reshaping		Without reshaping	
n	Correct θ	Error in C	Correct θ	Error in C
4	99.9%	0.000%	91.5%	0.499%
6	99.9%	0.007%	88.2%	0.689%
8	99.7%	0.011%	86.5%	0.789%
10	99.7%	0.006%	85.3%	0.852%
12	99.7%	0.007%	84.6%	0.901%
14	99.7%	0.007%	84.0%	0.926%

Notes: Same simulations as Table 1 but without excluding economies such that $\theta \notin \{0, 1\}^n$. See Section B.1 in the Appendix for details.

A similar exercise can be done for economies with a large number of firms. This exercise is analogous to that of Table 2 and is presented in Table 10. We see that even when the first-order conditions of $\mathcal{R}$ yield a solution $\theta \notin \{0, 1\}^n$, the error in aggregate output is negligible. This last test suggests that the solution approach works particularly well in realistic economies with a large number of firms.

[67] Note that we can see whether $\theta \in \{0, 1\}^n$ or not when solving the problem. If extreme precision is needed, we can be extra careful when $\theta \notin \{0, 1\}^n$ and use additional tests (for instance we can look for beneficial deviations) to check the robustness of the solution.

Table 10: Testing the reshaping on large networks without exclusions

	With reshaping		Without reshaping	
n	Correct θ	Error in C	Correct θ	Error in C
1000	99.9%	$< 0.001\%$	66.5%	0.56%

Notes: Same as Table 2 except all simulations are kept in the sample. $x < 0.001\%$ indicates that $x > 0$ but that proper rounding would yield 0, and similarly for $x > 99.9\%$.

B.4 Performance with very sparse matrices Ω

Table 11 shows the same simulations as Table 1 but with matrices Ω that are drawn so that firms have only 1 or 2 *potential* incoming connections on average. As a result, the networks of potential connections described by the matrices Ω are extremely sparse. The algorithm still performs well, with an average error in aggregate output that is 66 times smaller than when the problem is not reshaped.

Table 11: Testing the reshaping on sparse networks Ω

	With reshaping		Without reshaping	
n	Correct θ	Error in C	Correct θ	Error in C
8	99.6%	0.008%	91.0%	0.508%
10	99.5%	0.007%	91.1%	0.496%
12	99.6%	0.007%	90.8%	0.503%
14	99.5%	0.009%	90.5%	0.531%

Notes: Same simulations as Table 1 but with matrices Ω in which firms have on average only 1 or 2 potential connections.

B.5 Same parametrization as the calibrated economy

For the tests in this section, we set the parameters of the model to their calibrated values, which will be described in Section 6, except for the number of firms n which must remain small so that the true solution of the planner's problem can be found. Table 12 presents the results. We see that the solution approach still performs well, although the errors in C and θ are larger than in the general simulations of Table 1.[68] More importantly, the error in C are slowly declining as n increases which is reassuring about the algorithm's performance for realistic n. In contrast, without reshaping the performance of the algorithm degrades quickly as n increases.

[68] The calibrated matrices Ω are very sparse, which makes these economies more challenging for the algorithm.

Table 12: Testing the solution approach with calibrated parameters

	With reshaping		Without reshaping	
n	Correct θ	Error in C	Correct θ	Error in C
8	98.9%	0.008%	88.8%	0.252%
10	98.9%	0.008%	86.9%	0.294%
12	98.9%	0.007%	85.8%	0.317%
14	99.0%	0.007%	84.5%	0.345%
16	99.0%	0.006%	83.6%	0.373%
18	99.0%	0.006%	83.2%	0.380%
20	99.0%	0.006%	82.5%	0.395%

Notes: Parameters as in the calibrated economy of Section 6 except for n. For each n, 15,000 random economies are simulated. See also Appendix B for more details about the procedure.

B.6 Formation of the network link by link

This appendix provides the results of two exercises that show that reshaping the planner's problem is also useful when the production network is constructed link by link instead of through the extensive margin of the firms. In both exercises, the economy contains m real firms that are always active ($f_j = 0$).[69] Any two of these real firms are connected to each other by a link: for any ordered pair of real firms i, j with $i \neq j$, there exists a "link firm" k such that $\Omega_{ik} > 0$ and $\Omega_{kj} > 0$). There are no other connections in Ω. These link firms will operate or not as a function of economic conditions.

Individual link formation in small networks

When the number m of real firms is small, we can use the same approach as in Section 3.4 and find the true solution to the planner's problem by comparing the welfare provided by each possible network θ (see algorithm in Appendix G.2). There are at most $m\,(m-1)$ links in an economy, in which case the utility provided by $2^{m(m-1)}$ networks must be compared. Since this quantity grows rapidly with m, Table 13 shows the results of these tests when there are only $m \in \{3, 4, 5\}$ real firms. As before, the outcome of this exhaustive search is compared to the allocation found by reshaping the planner's problem.

We see from Table 13 that the reshaping algorithm works well. Over all the simulations, more than 99.7% of the links are assigned the proper operating status θ and the errors in aggregate output are small. Without reshaping, large fractions of the links are assigned the wrong operating status and the error in aggregate output can be sizable.

[69] This assumption is made to focus on the link formation aspect of the problem. We have experimented with economies in which $f_j > 0$ for the real firms and the results are similar.

Table 13: Individual links formation with few firms

Number of firms		With reshaping		Without reshaping	
Real firms m	Link firms $n-m$	Correct θ	Error in C	Correct θ	Error in C
3	6	99.9%	0.001%	90.9%	0.25%
4	12	99.8%	0.004%	85.9%	0.39%
5	20	99.7%	0.004%	82.0%	0.52%

Notes: Real firms: $f_j = 0$, $\alpha_j = 0.5$, $\sigma = \varepsilon_j = 6$, $\sigma_z = 0.25$. Link firms: $f_{link} \sim$ iid $U([0.0, 0.1/n])$, $\alpha_{link} \sim$ iid $U([0.5, 1.0])$ and $\sigma_{z_{link}} = 0.25$. For simplicity all non-zero Ω_{ij} are set to 1. For each m, 500 economies are generated randomly and the algorithm of Section G.4 is used to solve the planner's problem. An economy is kept in the sample only if the first-order conditions converge to a point in $\{0,1\}^n$. More than 80% of the economies are kept in the sample.

Individual link formation in large networks

For economies with a large number of firms, the true solution to the planner's problem is unknown but we can check whether there exist welfare-improving deviations from the allocation found using the reshaped problem. The procedure is the same as in the Section 3.4. The parameters of the tests are the same as in Table 13 but the economies feature $m \in \{10, 25, 40\}$ real firms and $n \in \{100, 625, 1600\}$ total firms (real plus links). The results are presented in Table 14. Reshaping the planner's problem yields solutions with few welfare-improving deviations so that the vast majority of links are assigned the correct status and the errors in aggregate output are negligible. In contrast, a large fraction of the links are assigned the wrong status and the errors in aggregate output are significant when the problem is not reshaped.

Table 14: Individual links formation with a large number of firms

Number of firms		With reshaping		Without reshaping	
Real firms m	Link firms $n-m$	Correct θ	Error in C	Correct θ	Error in C
10	90	99.9%	0.002%	76.8%	0.66%
25	600	$>$ 99.9%	$<$ 0.001%	74.0%	0.73%
40	1560	$>$ 99.9%	$<$ 0.001%	73.4%	0.74%

Notes: The parameters of these tests, except for m, are as in Table 13. An economy is kept in the sample only if the first-order conditions converge to a point in $\{0,1\}^n$. $x < 0.001\%$ indicates that $x > 0$ but that proper rounding would yield 0, and similarly for $x > 99.9\%$. For each m, 500 economies are generated randomly and the algorithm of Section G.4 is used to solve the planner's problem.

One potential concern of using the reshaping method in this context is that the first-order conditions often converge on a vector θ such that $\theta_j \notin \{0,1\}$ for at least one firm.[70] There are two reasons for this. First, as the total number of firms increases (up to $n = 1600$ for the economies with $m = 40$) it's more likely that at least one firm ends up with $\theta_j \notin \{0,1\}$. Second, the matrices Ω considered here are extremely sparse. As a result, the forces pushing the first-order conditions

[70] In the simulations of Table 14, the first-order conditions converge to a point $\theta_j \in \{0,1\}$ *for all* j in 43% of the simulations for $m = 10$, 12% for $m = 25$ and 5% for $m = 40$.

to hit the bounds are weakened (see footnote 16). In practice, however, these issues have limited implications. Only a small fraction of the links end up away from the $\{0,1\}$ bounds, and their impact on aggregate output is minimal. Table 15 shows the outcome of the same simulations but without excluding any simulations. We see that the results are essentially unchanged and that the solution approach also performs well in these situations.

Table 15: Individual links formation with a large number of firms and without exclusions

Number of firms		With reshaping		Without reshaping	
Real firms m	Link firms $n-m$	Correct θ	Error in C	Correct θ	Error in C
10	90	99.8%	0.004%	76.9%	0.65%
25	600	$>99.9\%$	$<0.001\%$	74.3%	0.71%
40	1560	$>99.9\%$	$<0.001\%$	73.4%	0.73%

Notes: The parameters of these tests are the same as in Table 13. No economies are excluded from the sample. $x<0.001\%$ indicates that $x>0$ but that proper rounding would yield 0, and similarly for $x>99.9\%$. For each m, 500 economies are generated randomly and the algorithm of Section G.4 is used to solve the planner's problem.

C Undistorted and distorted equilibrium

This appendix contains analytical results related to the equilibrium definitions of Section 4.

C.1 Marginal cost of production

The marginal cost of production δ_j of firm j is given by

$$\delta_j := \min \sum_{i\in\mathcal{N}} Q_i x_{ij} + W l_j,$$

subject to

$$\frac{A}{\alpha_j^{\alpha_j}\left(1-\alpha_j\right)^{1-\alpha_j}} z_j \theta_j \left(\sum_{i\in\mathcal{N}} \Omega_{ij}^{\frac{1}{\varepsilon_j}} x_{ij}^{\frac{\varepsilon_j-1}{\varepsilon_j}}\right)^{\frac{\varepsilon_j}{\varepsilon_j-1}\alpha_j} l_j^{1-\alpha_j} \geq 1.$$

Denote ζ the Lagrange multiplier on the constraint. The first-order conditions are given by

$$\left(1-\alpha_j\right)\zeta y_j = W l_j$$

$$\zeta \frac{A}{\alpha_j^{\alpha_j}\left(1-\alpha_j\right)^{1-\alpha_j}} \theta_j z_j \alpha_j \left(\sum_{i\in\mathcal{N}} \Omega_{ij}^{\frac{1}{\varepsilon_j}} x_{ij}^{\frac{\varepsilon_j-1}{\varepsilon_j}}\right)^{\alpha_j \frac{\varepsilon_j}{\varepsilon_j-1}-1} \Omega_{kj}^{\frac{1}{\varepsilon_j}} x_{kj}^{\frac{\varepsilon_j-1}{\varepsilon_j}} l_j^{1-\alpha_j} = Q_k x_{kj}.$$

The last equation can be written

$$\sum_i Q_i x_{ij} = \zeta \alpha_j y_j,$$

so that

$$\sum_{i\in\mathcal{N}} Q_i x_{ij} + W l_j = \zeta\alpha_j y_j + (1-\alpha_j)\,\zeta y_j$$
$$= \zeta.$$

Note that this implies that δ_j is equal to the Lagrange multiplier on the constraint (17) in the profit maximization problem of the firm.

D Stable equilibrium

In this section, we consider an alternative equilibrium concept based on the stability of an allocation. Namely, we consider an environment in which firms are facing contractual obligations to purchase and deliver goods. An equilibrium, in that context, is an allocation in which there are no groups of firms that wants to change the terms of the contracts.[71] One of the result of this section is that the efficient allocation can be decentralized as a stable equilibrium.

We first describe the contractual environment. Define a *contract* between two firms i and j as a pair $\{x_{ij}, T_{ij}\}$ where x_{ij} is a quantity shipped from i to j, and T_{ij} is a payment from j to i. An *arrangement* is a collection of contracts between all possible pairs of firms $\{x_{ij}, T_{ij}\}_{i,j\in\mathcal{N}^2}$.

Under a given arrangement, a firm j must supply and purchase the prescribed quantities, but it can decide on a price p_j to charge the household, an amount c_j to sell to the final good producer, how much labor l_j to employ, and its operating status θ_j. It makes these decisions to maximize profits

$$\pi_j = p_j c_j - w l_j + \sum_{i\in\mathcal{N}} T_{ji} - \sum_{i\in\mathcal{N}} T_{ij} - w\theta_j f_j L, \tag{30}$$

where w is the wage, and subject to a technology constraint

$$c_j + \sum_{k\in\mathcal{N}} x_{jk} \leq y_j, \tag{31}$$

where y_j satisfies (2), and to the usual demand curve

$$c_j = \beta_j C\,(p_j/P)^{-\sigma} \tag{32}$$

where $P = \left(\sum_j \beta_j P_j^{1-\sigma}\right)^{1/(1-\sigma)}$ is the price index. When making decision each firm takes w, C and P as given.

We say that an allocation is *feasible* if all the technology constraints (31) and the labor resource

[71] This equilibrium concept has proven particularly convenient in network economies (Jackson and Wolinsky, 1996; Hatfield et al., 2013). The approach followed here is most closely related to Oberfield (2018).

constraint $\sum_j l_j + \sum_j \theta_j f_j L \leq L$ are satisfied.

A *coalition* is a set of firms J. All coalitions behave atomistically, in the sense that they take aggregate consumption C, the wage w and the aggregate price level P as given. A *deviation* for a given coalition J consists of (i) dropping any contracts that involve at least one firm in J and (ii) altering the terms of any contract involving a buyer and a supplier that are both members of the coalition. Finally, a *dominating deviation* for a given coalition is a deviation that delivers at least the same amount of profits to all members of the coalition and strictly greater profits to at least one member.

We can now define a stable equilibrium in this environment.

Definition 2. A stable equilibrium is an arrangement $\{x_{ij}, T_{ij}\}_{i,j \in \mathcal{N}^2}$, firms' choices $\{p_j, c_j, l_j, \theta_j\}_{j \in \mathcal{N}}$ and a wage w such that (i) given the wage, total profits, and prices, the consumption choices $\{c_j\}_{j \in \mathcal{N}}$ maximize the utility of the representative household; (ii) for each $j \in \mathcal{N}$, $\{p_j, c_j, l_j, \theta_j\}$ maximizes the profits of j given the arrangement, the wage, the household's demand and the technology constraint; (iii) labor and final goods markets clear; (iv) there are no dominating deviations available to any coalition; and (v) the equilibrium allocation is feasible.

The following proposition shows how equilibria and the efficient allocation are related.

Proposition 13. *Every stable equilibrium is efficient.*

Proof. All proofs are in Appendix H. □

This proposition shows that every equilibrium allocation is a solution to the planner's problem $\mathcal{P}$.[72] As a result, solutions to $\mathcal{P}$ implicitly characterize equilibrium outcomes in this economy.

E Additional example of clustering

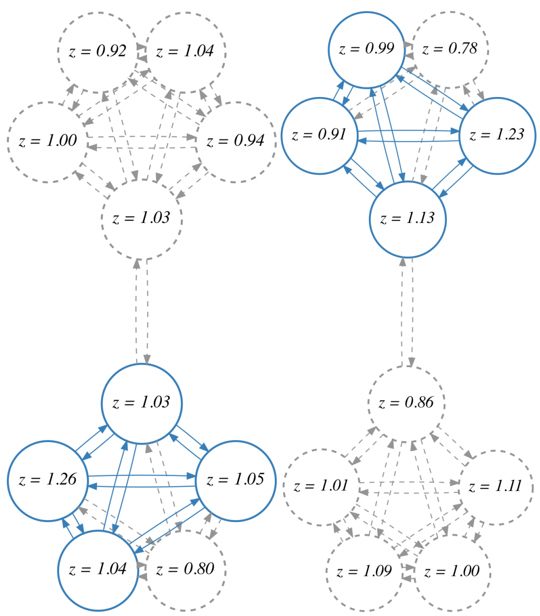
In this section, we provide another example of clustering and how small changes in the environment can trigger large reorganizations of the network.

When z varies across firms, the efficient network tends to cluster economic activity around the most productive firms. The left panel of Figure 13 presents an example of this process in an economy in which the network Ω consists of two groups of five fully-connected firms. The two groups are linked together by a single pair of connections. The figure shows the same economy under two randomly drawn vectors z. We can see that the planner clusters economic activity in either the top or the bottom group of firms. Operating a few firms in each group would be a poor way to organize production. In addition, the active cluster tends to be the one with the highest-z

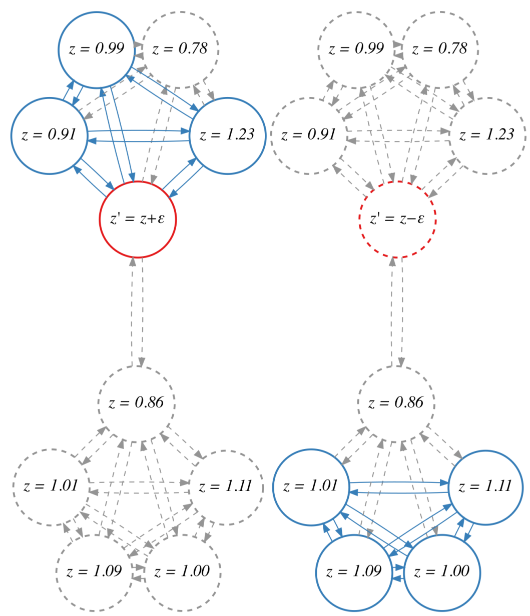
[72] Note that since the household is not part of any coalition, the grand coalition does not seek to maximize social welfare.

firm. By organizing production around these high-performing firms, the planner magnifies their impact on their neighbors and on aggregate output.

Figure 13b also provides an example of how a small shock can trigger a large reorganization. Both economies are identical except for the productivity z of the red firm which is slightly larger in the economy on the left. While the drop in z from left to right is negligible, it triggers a large reorganization of the network. Aggregate output, however, is barely affected. Indeed, the planner reorganizes the network precisely to limit the negative impact of the shock on output. But while aggregate output barely changes, firm-level distributions can change substantially. In this example, for instance, the dispersion in labor productivity, output and employment across firms collapses after the shock. A negligible shock can therefore have a large impact on cross-producer distributions.



(a) Clustering of activity



(b) Small shock, large reorganization

Figure 13: Economic forces shape the network

F Appendix for Section 6

F.1 Sensitivity to different Ω

To see how the matrix Ω affects the efficient allocation, we simulate the model under different parametrization for Ω. We still assume that Ω is drawn from a bivariate power law of the first kind but we vary its exponent to $\xi = 1.7$ and $\xi = 1.9$. The results are presented in Table 16. We see from the table that changing ξ has a direct impact on the degree distributions and the global clustering coefficient in the efficient network. Under $\xi = 1.7$, the distribution of the number of potential connections features thicker tails such that Ω offers a lot of options for the planner to create highly-connected firms and dense clusters of producers. The planner takes advantage of these possibilities to increase welfare: the mean of aggregate output is 17.1 under $\xi = 1.7$, but only 15.4 under $\xi = 1.9$.

Table 16: Impact of Ω on the production network

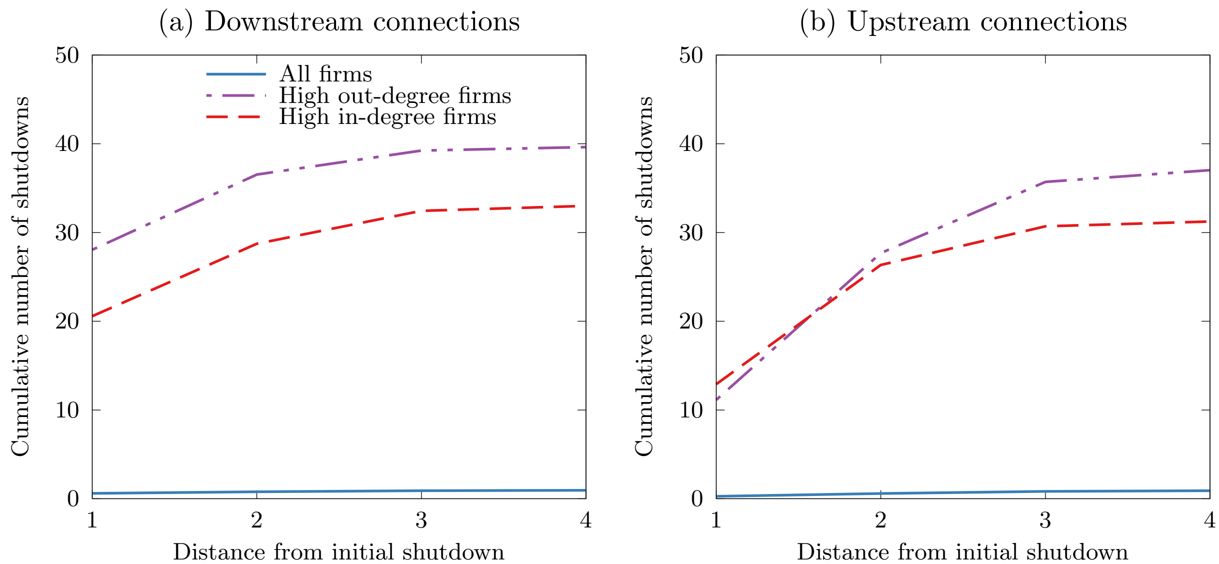
	Power law exponent		Clustering coefficient
Network	indegree	outdegree	
$\xi = 1.7$	0.89	0.86	4.20
$\xi = 1.78$ (benchmark)	0.97	0.92	3.45
$\xi = 1.9$	1.11	1.05	2.30

Notes: The parameters are the same as in the benchmark calibration except for the distribution from which Ω is drawn (see text).

F.2 Cascades with $\varepsilon = 3$

We can use the model to evaluate how cascades propagate in counterfactual economies with different parameters. One parameter with an important role for the cascades is the elasticity of substitution between intermediate inputs, ε. Figure 14 shows the outcome of the same exercise as that of Figure 9 but in an economy in which $\varepsilon = 3$, such that intermediate inputs are worse substitutes than in the calibrated economy. By comparing the figures, we see that the lower elasticity affects the cascades in two important ways: 1) shocks to high degree firms now trigger larger cascades (notice the different scales) and, 2) these cascades have more substantial upstream propagation compared to those in the benchmark economy.

Why cascades are larger when they originate from high-degree firms is easy to understand. With ε small, intermediate inputs are poor substitutes and losing a supplier has a larger negative impact on a firm's productivity, which leads to more shutdowns. To understand why it also makes cascades have more upstream propagation, it is useful to think about the planner's incentives to operate a firm in this economy. Since the elasticity of substitution in the consumption aggregator, σ, is relatively large, equation (7) implies that firms with high productivity q are particularly valued by the household. But because ε is small, these high-q firms are likely to get their high productivity from a large number of suppliers. As a result, if one of the high-q firm shuts down, its many suppliers are no longer useful (they don't contribute much to Q) and the planner is likely to shut them down as well, thereby triggering an upstream cascade. In contrast, in the benchmark economy, where $\varepsilon = \sigma = 5$, the planner puts a higher value on the direct contribution to final consumption of these many suppliers, and they are therefore more likely to remain if one of their large customers shuts down.



Notes: Cumulative number of exits at different distances from shuttered firm. "High degree" refers to the firms above the 99th percentile. Simulations of 100 randomly drawn matrices Ω, for each of which 1000 cascades are created.

Figure 14: Cumulative cascades by degree of originator, $\varepsilon = 3$.

F.3 Cascades in the undirected network

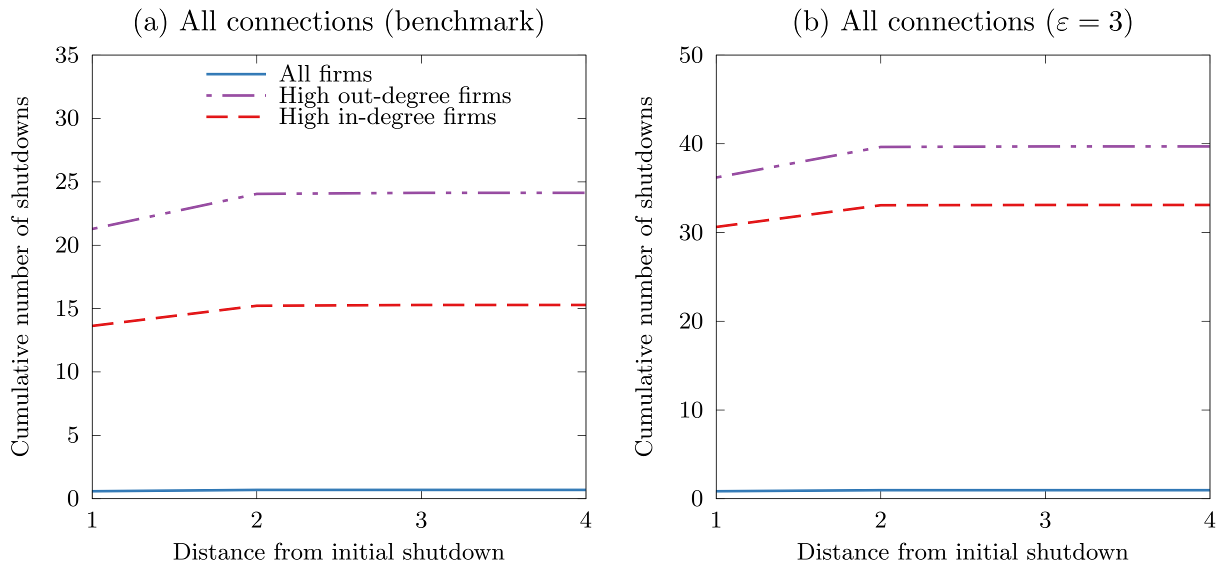
Section 6.5 describes the propagation of cascades of shutdowns by considering upstream and downstream firms separately. Here, we reproduce the same figures but by including all connections together instead. Figure 15 corresponds to the exercises of Figures 9 and 14 in the main text. Figure 16 corresponds to the exercise of Figure 10 in the main text.

F.4 Cost of cascades in a fixed network

In this section we evaluate the welfare cost of cascades in an economy with a *fixed* production network. The results are presented in Table 17 which shows the outcome of the same exercise as Table 4, except that the operating status of all firms is kept constant. We see that shocks have a more severe impact in this case. This effect is particularly strong for shocks to high-degree firms in the economy with $\varepsilon = 3$.

F.5 Cascades in the distorted equilibrium

In this section, we consider how cascades propagate in the distorted equilibrium. To do so, we proceed as in the exercise of Figure 9, where we looked at the downstream and upstream propagation that follows from the exit of a single producer. In Figure 17 we report the ratio of downstream to upstream cumulative shutdowns in the efficient allocation (solid blue line) and the inefficient equilibrium (dashed purple line). As we can see, cascades in the equilibrium propagate much more downstream than upstream. This different is particularly important when we consider



Notes: Cumulative number of exits at different distances from shuttered firm. "High degree" refers to the firms above the 99th percentile. Simulations of 100 randomly drawn matrices Ω, for each of which 1000 cascades are created.

Figure 15: Cumulative cascades by degree of originator, all connections.

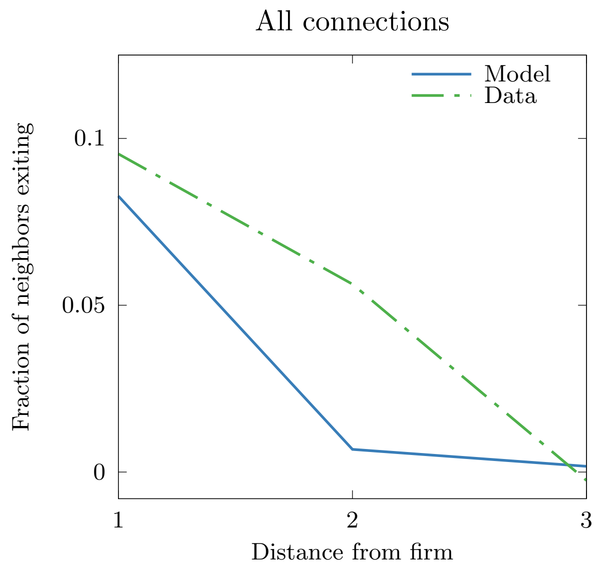
	Impact on output	
	Benchmark	$\varepsilon = 3$
Average firm	-0.1%	-0.1%
High indegree	-1.9%	-4.8%
High outdegree	-2.6%	-5.1%
High degree	-2.5%	-5.3%

Notes: The operating status of firms is kept fixed. "High degree" refers to firms above the 99th percentile. Simulations of 100 randomly drawn matrices Ω, for each of which 1000 cascades are created.

Table 17: Correlation between output drop and firm degree in a fixed network

all cascades (panel a), but also when we restrict the sample to cascades originating from high-degree firms (panels b and c). Additional downstream propagation is also visible in the efficient allocation, although the effect is less strong here, implying that there is significantly more upstream propagation in the efficient allocation. This is in line with Proposition 9 and the overall discussion in Section 5.3, which describe that because of the pricing distortions the equilibrium features less upstream propagation of shocks.

We can also look at the welfare cost of the cascades in the efficient allocation. Table 18 shows the same exercise as Table 4 but for the distorted equilibrium. We see that cascades have a larger impact on output in the distorted equilibrium. In that case, the reorganization of the network after the shock is suboptimal and welfare is adversely affected as a result.



Notes: Factset data. Estimated coefficients from regressing the fraction of exiting neighbors on whether a firm exits. Time fixed effects and indegree and outdegree controls are included. The distance is the smallest number of connections between two firms.

Figure 16: Cascades of firm shutdowns in the model and in the data, all connections

	Impact on output	
	Efficient	Equilibrium
Average firm	−0.11%	−0.13%
High indegree	−1.79%	−1.83%
High outdegree	−2.47%	−2.56%
High degree	−2.36%	−2.47%

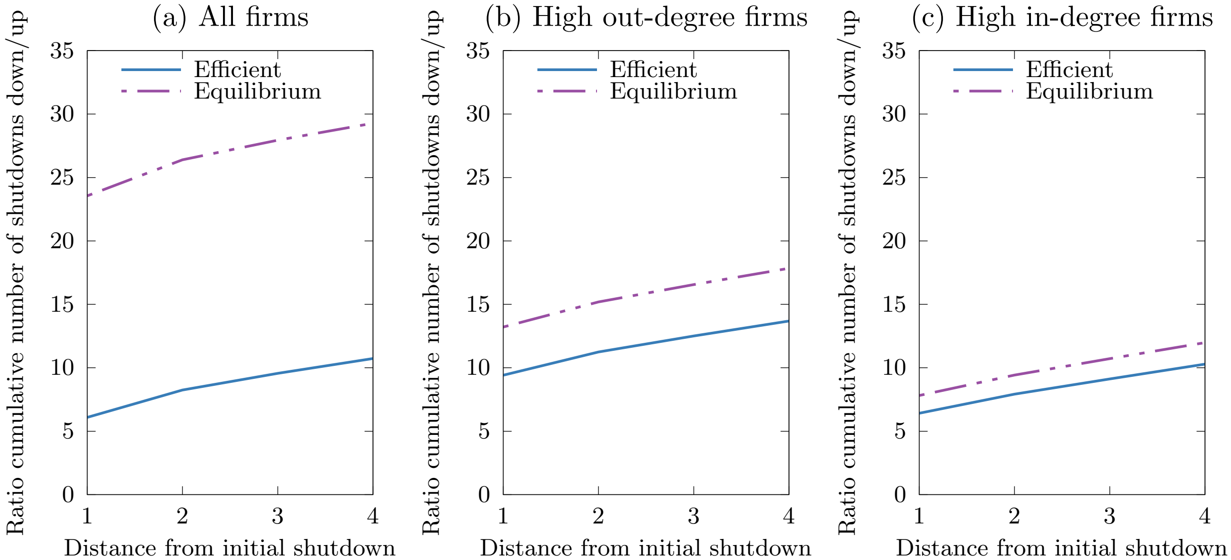
Notes: "High degree" refers to firms above the 99th percentile. Simulations of 100 randomly drawn matrices Ω, for each of which 1000 cascades are created.

Table 18: Correlation between output drop and firm degree in the efficient allocation and the distorted equilibrium

F.6 Invariance of the network to aggregate shocks

There are no aggregate productivity shocks in the calibrated economy but one might think that these shocks can also affect the structure of the network and should therefore be included in the analysis. Fortunately, the following proposition shows that this is not the case: the network is completely unaffected by changes in aggregate productivity A. We can therefore abstract from aggregate shocks to explore the interaction between aggregate fluctuations and the shape of the network.

Proposition 14. *If $\alpha_j = \alpha$ for all $j \in \mathcal{N}$, then the efficient network θ does not depend on aggregate productivity A.*



Notes: Cumulative number of exits at different distances from shuttered firm. The figure reports the downstream to upstream ratio of these numbers in the efficient allocation and the distorted equilibrium. "High degree" refers to the firms above the 99th percentile. Simulations of 100 randomly drawn matrices Ω, for each of which 1000 cascades are created.

Figure 17: Ratio of downstream to upstream cumulative cascades in the efficient allocation and the distorted equilibrium

F.7 Large number of firms and aggregate shocks

To investigate how the model behaves under a more realistic parametrization, we simulate the calibrated economy with $n = 20,000$ firms and with aggregate shocks to total factor productivity A. The number of firms was chosen to roughly match the number of operating firms in the Factset data. We assume that $\log(A_t)$ follows an AR(1) process with an autocorrelation of 0.9 and and a standard deviation parameter set to match empirical estimates about the impact of aggregate shocks on volatility.[73,74]

Table 19 shows the correlations between aggregate output and the shape of the network. We see that the numbers are broadly similar to those of the benchmark calibration. We also compute the difference in output volatility between the flexible and fixed networks in this setting. We find that the flexible network economy is about 11% less volatile. Finally, aggregate output is also 13%

[73] Atalay (2017) generalizes an empirical strategy introduced by Foerster et al. (2011) to evaluate the impact of aggregate shocks on aggregate fluctuations. He finds that they account for 17% of volatility. We parametrize the stochastic process followed by $\log(A_t)$ to match that estimate.

[74] The proof of Proposition 14 shows that we can write the planner's problem as

$$\max_{\theta\in\{0,1\}^n} A^{\frac{1}{1-\alpha}} Q\left(1-\sum_{j\in\mathcal{N}}\theta_j f_j\right) L, \text{ where } q_j = z_j\theta_j\left(\sum_{i\in\mathcal{N}}\Omega_{ij} q_i^{\varepsilon_j-1}\right)^{\frac{\alpha}{\varepsilon_j-1}}, \forall j. \tag{33}$$

Since A only enters as a multiplicative constant in the objective function, it has no impact on the optimal θ and thus on the production network. (33) also shows that it is straightforward to compute the variance of output with aggregate shocks. If the variance of log output is x without fluctuations in A, then the overall variance of log output would simply be $x + (1-\alpha)^{-2} y$ if we added shocks with a variance of y to $\log(A)$.

larger under the flexible network, roughly the same number as in the benchmark economy. This suggest that the importance of the discrete margin of adjustment matters even in large economies.

Table 19: Correlation between with aggregate output with $n = 20,000$ firms and aggregate shocks

	Power law exponents		Clustering coefficient
Network	indegree	outdegree	
Model with $n = 20,000$ firms and aggregate shocks	-0.75	-0.83	0.79
Benchmark model	-0.53	-0.63	0.60

Notes: All time series are in logs. The parameters of the economy are as in the benchmark calibration except as mentioned in the text. Since these simulations are computationaly intensive, we simulate four economies instead of twenty in the benchmark exercises.

F.8 Mean and standard deviation of GDP with different parameters

The elasticities of substitution σ and ε matter for the differences between the flexible and the fixed networks. As Table 20 shows, these differences become larger when ε and σ are large. In this case, the network is very flexible so that small shocks can create big movement. Preventing these movement can lead to a steep decline in output and a large increase in volatility. In contrast, when the elasticities are low, the network is very rigid so that the differences between the flexible and fixed networks are minimal.[75]

Elasticities ε and σ	7	5	3
Change in mean C from fixed network	20	11%	1%
Change in std C from fixed network	-34%	-17%	-1%

Table 20: Flexible vs fixed network under different elasticities

F.9 Correlations between output and the network with different parameters

In this section we look at the comovement between the production network and aggregate output under different parameter values for σ and ε. Table 21 shows the results. As we can see, the correlations are roughly stable across parameters.

[75] The correlations in Table 6 stay roughly the same when the elasticities σ and ε change.

Table 21: Correlations between network moments and aggregate output with different σ and ε

Elasticity consumption bundle σ	3	5	5	7
Elasticity inter. inputs bundle ε	3	3	5	7
Power law exponents				
Indegree distribution	−0.68	−0.56	−0.53	−0.05
Outdegree distribution	−0.71	−0.78	−0.63	−0.25
Centrality distribution	0.08	0.38	−0.10	0.14
Global clustering coefficient	0.61	0.77	0.60	0.44
Average distance	−0.66	−0.83	−0.82	−0.75

Notes: All time series are in logs. Output is detrended linearly in sample. Power law exponents are estimated following Gabaix and Ibragimov (2011). To focus on the right tail, we truncate the eigenvector centrality distribution below the first quartile. The global clustering coefficient, the average distance and the eigenvector centrality are computed on the undirected graph.

G Algorithms

This appendix describes the various algorithms used in the paper.

G.1 Construction of the matrix Ω for the numerical tests.

This algorithm constructs the matrices Ω used in the numerical tests of Section 3.4 in the main text and of Sections B.1 and B.3 in the Appendix. Consider an economy with n firms, each with m incoming potential connections on average. Set $p = m/(n-1)$ and $\Omega_{ij} = 0$ for all $i, j \in \mathcal{N}^2$.[76]

1. Draw $\Omega_{ij} \sim$ iid Bernoulli(p), for all $i, j \in \mathcal{N}^2$.
2. For each $i, j \in \mathcal{N}^2$ such that $\Omega_{ij} = 1$, draw $\Omega_{ij} \sim$ iid $U[0, 1]$.
3. Set $\Omega_{ii} = 0$ for all $i \in \mathcal{N}$.

G.2 Exhaustive search

This algorithm performs an exhaustive search of the 2^n vectors $\theta \in \{0, 1\}^n$. It is used in Section 3.4 in the main text as well as in Sections B.1, B.3 and B.6 in the Appendix.

1. Order in an arbitrary way all the possible $\theta \in \{0, 1\}^n$, from θ^1 to θ^{2^n}.
2. For each $p \in \{1, \ldots, 2^n\}$, use equations (6) and (8) to compute the aggregate consumption associated with θ^p.
3. The vector θ that provides the highest aggregate consumption corresponds to the efficient allocation.

[76]The division by $n-1$ is needed because the diagonal is forced to be empty.

This algorithm is guaranteed to find the global maximum of $\mathcal{P}$ but it is infeasible for large n given the speed at which the number of vectors in $\{0,1\}^n$ grows with n.

G.3 Deviation-free allocation

This algorithm starts from an allocation $\theta^0 \in \{0,1\}^n$ and looks for welfare-improving deviations. It is used in Sections 3.4, B.3 and B.6.

1. Initialize the 0-th iteration with θ^0.

2. For the p-th iteration, define $\tilde{\theta} = \theta^p$ and set $j = 1$.

   (a) If $\theta_j^p = 0$, set $\tilde{\theta}_j = 1$. If, instead, $\theta_j^p = 1$, set $\tilde{\theta}_j = 0$.

   (b) Using equations (6) and (8) compute the welfare associated with $\tilde{\theta}$.

   (c) If the welfare under $\tilde{\theta}$ is larger than the welfare under θ^p set $\theta^p = \tilde{\theta}$.

   (d) Set $j = j + 1$, set $\tilde{\theta} = \theta^p$ and repeat steps (a) through (d) until $j = n$.

3. Repeat step 2 above until no welfare-improving deviations are found for some θ^p.

G.4 Iterating on the first-order conditions

A convenient way to solve the reshaped planner's problem is to iterate on the first-order conditions of the log of the objective function of $\mathcal{R}$ while treating (9) as an inequality constraint. In what follows ζ_k is the Lagrange multiplier on the k-th inequality constraint (9), and $\underline{\mu}_j$ and $\overline{\mu}_j$ are the Lagrange multipliers on the constraint $\theta_j \geq 0$ and $\theta_j \leq 1$. 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 $k \in \mathcal{N}$.

2. For the p-th iteration:

   (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) With θ^p, iterate on (9) until convergence to find the vector q^p.

   (c) For each j, compute $B_j = \left(\sum_{i=1}^n \Omega_{ij} q_i^{\varepsilon_j - 1}\right)^{\frac{1}{\varepsilon_j - 1}}$ and $\Lambda_j = \frac{\theta_j}{B_j^{\varepsilon_j - 1}}$ if $B_j > 0$ and $\Lambda_j = 0$ otherwise.

   (d) Find $\frac{\zeta_k^p q_k^p}{\theta_k^p}$ by solving the following system of linear equations derived from the first-order conditions:

$$\frac{\beta_k \left(A z_k B_k^{\alpha_k}\right)^{\sigma-1}}{\sum_{j\in\mathcal{N}} \beta_j q_j^{\sigma-1}} + \sum_{j\in\mathcal{N}} \left(A z_k B_k^{\alpha_k}\right)^{\varepsilon_j - 1} \Omega_{kj} \alpha_j \Lambda_j \frac{\zeta_j q_j}{\theta_j} = \frac{\zeta_k q_k}{\theta_k}$$

   for each k, and where $\frac{\beta_k \left(A z_k B_k^{\alpha_k}\right)^{\sigma-1}}{\sum_{j\in\mathcal{N}} \beta_j q_j^{\sigma-1}}$ should be set to 0 if $\sum_{j\in\mathcal{N}} \beta_j q_j^{\sigma-1} = 0$.

(e) Compute $\Delta\mu_k$ using the following equation derived from the first-order conditions

$$\frac{f_k}{L - \sum_{j\in\mathcal{N}} f_j \theta_j} = a_k \frac{\zeta_k q_k}{\theta_k} + \sum_{j\in\mathcal{N}} b_{kj}\alpha_j \Omega_{kj} \Lambda_j \left(A z_k B_k^{\alpha_k}\right)^{\varepsilon_j - 1} \frac{\zeta_j q_j}{\theta_j} + \Delta\mu_k$$

for each $k \in \mathcal{N}$ and update to $\Delta\mu^{p+1} = \psi \Delta\mu + (1-\psi)\Delta\mu^p$ where $0 < \psi \leq 1$ is some parameter to control the speed of convergence.

3. Repeat step 2 above until convergence on $\Delta\mu$.

In practice, it is useful to slow down the updating rule by setting $\psi = 0.9$.

Notice that this algorithm imposes that $\theta \in \{0,1\}^n$ at every iteration. When the solution to $\mathcal{R}$ is not in $\{0,1\}^n$, the algorithm does not converge and the status θ of some firms keeps alternating between 0 and 1. In practice, we stop the algorithm when the distance between $\Delta\mu_k^{p+1}$ and $\Delta\mu_k^p$ starts to increase, which usually indicates that there will be no convergence. We then look at the set of firms for which θ keeps alternating (different sign for $\Delta\mu_k^{p+1}$ and $\Delta\mu_k^p$), and then pick the best $\theta \in \{0,1\}$ to maximize the planner's objective function.

G.5 Construction of the matrix Ω in the calibrated economy

The matrix Ω is constructed by assuming that the number of potential incoming and outgoing connections (x_{in}, x_{out}), for any given firm, is drawn from a bivariate power law of the first kind $\mathcal{K}$ for which the joint density over (x_{in}, x_{out}) is $g(x_{in}, x_{out}) = \xi(\xi - 1)(x_{in} + x_{out} - 1)^{-(\xi+1)}$. The full algorithm to construct the matrix is as follows:

1. Begin with $\Omega_{ij} = 0$ for all $i, j \in \mathcal{N}^2$.
2. For each firm $j \in \mathcal{N}$, draw from $\mathcal{K}$ a pair $\left(x_{in}^j, x_{out}^j\right)$ for the number of incoming and outgoing connections for j. Redraw until $\sum_j x_{in}^j = \sum_j x_{out}^j$ so that the total number of incoming connections is equal to the total number of outgoing connections in the economy.
3. For each $j \in \mathcal{N}$, create x_{in}^j incoming stubs and x_{out}^j outgoing stubs.
4. Randomly match each incoming stub to an outgoing stub. An incoming stub has the same probability of being matched with any outgoing stub. Set $\Omega_{ij} = 1$ where i is the firm associated with the outgoing stub and j is the firm associated with the incoming stub.
5. Since there are no self-link in the data, set $\Omega_{ii} = 0$ for all i.
6. Verify that each firm has at least one potential input, otherwise go back to step 1.

H Proofs

This section contains the proofs.

H.1 Efficient allocation in a fixed network

This section contains various results related to the economy with a fixed vector θ.

Labor productivity vector q

The proof of Proposition (1) relies on the following definitions from Kennan (2001).

Definition. A function $g : \mathbb{R}^n \to \mathbb{R}^n$ is radially quasiconcave ("R-concave") if $g(x) = 0$ and $x > 0$ and $0 \le \lambda \le 1$ implies $g(\lambda x) \ge 0$. If (in addition) $0 < \lambda < 1$ implies $g(\lambda x) > 0$, then g is strictly R-concave.

Definition. A function $g = (g_1, g_2, \ldots, g_n) : \mathbb{R}^n \to \mathbb{R}^n$ is quasi-increasing if $y_i = x_i$ and $y_j \ge x_j$ for all j implies $g_i(y) \ge g_i(x)$.

The following Lemma is used as an intermediate step to prove Proposition (1)

Lemma 3. *Denote by $\tilde{\mathcal{N}}$ any subset of $\mathcal{N}$ with $\tilde{n}$ firms and such that $\sum_{i \in \tilde{\mathcal{N}}} \Omega_{ij} > 0$ for all $j \in \tilde{\mathcal{N}}$. The function $g : \mathbb{R}^{\tilde{n}} \to \mathbb{R}^{\tilde{n}}$ defined, for all $j \in \tilde{\mathcal{N}}$, as*

$$g_j(p) = (z_j A)^{\varepsilon_j} \left(\sum_{i \in \tilde{\mathcal{N}}} \Omega_{ij} p_i^{\frac{\varepsilon_j - 1}{\varepsilon_j}} \right)^{\alpha_j \frac{\varepsilon_j}{\varepsilon_j - 1}} - p_j$$

is strictly R-concave.

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

$$\begin{aligned} g_j(\lambda p^*) &= \lambda^{\alpha_j} (z_j A)^{\varepsilon_j} \left(\sum_{i \in \tilde{\mathcal{N}}} \Omega_{ij} (p_i^*)^{\frac{\varepsilon_j - 1}{\varepsilon_j}} \right)^{\alpha_j \frac{\varepsilon_j}{\varepsilon_j - 1}} - \lambda p_j^* \\ &\ge \lambda (z_j A)^{\varepsilon_j} \left(\sum_{i \in \tilde{\mathcal{N}}} \Omega_{ij} (p_i^*)^{\frac{\varepsilon_j - 1}{\varepsilon_j}} \right)^{\alpha_j \frac{\varepsilon_j}{\varepsilon_j - 1}} - \lambda p_j^* \\ &\ge \lambda g_j(p^*) \\ &\ge 0 \end{aligned}$$

where the first inequality is strict for $0 < \lambda < 1$ since $0 < \alpha_j < 1$ and $\sum_{i \in \tilde{N}} \Omega_{ij} > 0$ by assumption. □

Proposition 1. In the efficient allocation, the labor productivity vector q satisfies

$$q_j = z_j \theta_j A \left(\sum_{i \in \mathcal{N}} \Omega_{ij} q_i^{\varepsilon_j - 1} \right)^{\frac{\alpha_j}{\varepsilon_j - 1}}, \tag{6}$$

for all j. Furthermore, there is a unique q that solves (6) such that $q_j > 0$ if j operates and has access to an operating cycle, and $q_j = 0$ otherwise.

Proof. We focus on the firms that have access to an operating cycle, as defined in Section 3, since the problem of the other firm is trivial. The first-order conditions of $\mathcal{P}$ with respect to l_j and x_{ij} are

$$wl_j = \lambda_j \left(1-\alpha_j\right) y_j \tag{34}$$

$$\lambda_i x_{ij} = \lambda_j \frac{A}{\alpha_j^{\alpha_j}\left(1-\alpha_j\right)^{1-\alpha_j}} \alpha_j z_j \theta_j \left(\sum_{k\in\mathcal{N}} \Omega_{kj}^{\frac{1}{\varepsilon_j}} x_{kj}^{\frac{\varepsilon_j-1}{\varepsilon_j}}\right)^{\frac{\varepsilon_j}{\varepsilon_j-1}\alpha_j-1} l_j^{1-\alpha_j} \Omega_{ij}^{\frac{1}{\varepsilon_j}} x_{ij}^{\frac{\varepsilon_j-1}{\varepsilon_j}}. \tag{35}$$

when firms i and j have access to an operating cycle.[77] Combining these conditions with the production function yields

$$x_{ij}\lambda_i^{\varepsilon_j} = \lambda_j^{\varepsilon_j}\alpha_j \left(A z_j \theta_j \left(\frac{\lambda_j}{w}\right)^{1-\alpha_j}\right)^{\frac{\varepsilon_j-1}{\alpha_j}} \Omega_{ij} y_j. \tag{36}$$

Plugging (34) and (36) back in the production function we find (6). For future calculations, it will be useful to rely on another expression for x_{ij} that we obtain by combining (35) with the production function:

$$x_{ij} = \Omega_{ij} X_j \left(\frac{\lambda_i}{\left(\sum_k \Omega_{kj}\lambda_k^{1-\varepsilon_j}\right)^{\frac{1}{1-\varepsilon_j}}}\right)^{-\varepsilon_j} \tag{37}$$

where $X_j = \left(\sum_k \Omega_{kj}^{\frac{1}{\varepsilon_j}} x_{kj}^{\frac{\varepsilon_j-1}{\varepsilon_j}}\right)^{\frac{\varepsilon_j}{\varepsilon_j-1}}$.

We follow Kennan (2001) to prove the uniqueness of q. Consider the change of variable $p_j = q_j^{\varepsilon_j}$, and let $\tilde{\mathcal{N}}$ be the set of firms that operate and that have access to an operating cycle. Denote the number of such firms by $\tilde{n}$. Clearly, $p_j = 0$ for $j \notin \tilde{\mathcal{N}}$. We can rewrite (6) as the following mapping from $\mathbb{R}^{\tilde{n}}$ to $\mathbb{R}^{\tilde{n}}$:

$$p_j = \left(z_j A\right)^{\varepsilon_j} \left(\sum_{i\in\tilde{\mathcal{N}}} \Omega_{ij} p_i^{\frac{\varepsilon_j-1}{\varepsilon_j}}\right)^{\alpha_j\frac{\varepsilon_j}{\varepsilon_j-1}}, \tag{38}$$

for all $j \in \tilde{\mathcal{N}}$. Denote the right-hand side of (38) by $f_j\left(p\right)$ and define $g : \mathbb{R}^{\tilde{n}} \to \mathbb{R}^{\tilde{n}}$ as $g\left(p\right) = f(p) - p$. A solution to (38) is therefore a vector p such that $g\left(p\right) = 0$. By Lemma 3, g is strictly R-concave. Note also that g is quasi-increasing.

Consider the mapping $h : \mathbb{R} \to \mathbb{R}^{\tilde{n}}$ defined as $h\left(s\right) = f\left(\mathbb{1}_{\tilde{n}} s\right)$ where $\mathbb{1}_{\tilde{n}}$ is the all-one vector

[77] A firm j without access to an operating cycle cannot produce and the planner simply sets $l_j = 0$, $y_j = 0$, and $x_{ij} = 0$ and $x_{ji} = 0$ for all i.

of size $\tilde{n}$. Then $h(s)$ is strictly concave, strictly increasing and differentiable with $h(0) = 0$, $\lim_{s\to 0} h'(s) = \infty$ and $\lim_{s\to\infty} h'(s) = 0$, in all dimensions.[78] As a result, there exist constants $\bar{p} > \underline{p} > 0$ such that $h(\underline{p}) > \underline{p}$ and $h(\bar{p}) < \bar{p}$. Then, Theorems 3.1 and 3.2 in Kennan (2001) apply: (38) has a unique positive fixed point p^* and there is therefore a unique positive q^* that satisfies (6). It is such that $q_j^* = \left(p_j^*\right)^{\frac{1}{\varepsilon_j}}$ if j operates and has access to an operating cycle, and $q_j^* = 0$ otherwise. Finally, Kennan (2001) also shows that the fixed point can be found my iterating on the mapping (6). Note also that the proof is essentially unchanged if we use the reshaped equation (9) instead of (6). □

Aggregate output in a fixed network

Lemma 1. In the efficient allocation, aggregate output is

$$C = Q\left(1 - \sum_{j\in\mathcal{N}} \theta_j f_j\right) L \tag{8}$$

where $Q = \left(\sum_{j\in\mathcal{N}} \beta_j q_j^{\sigma-1}\right)^{\frac{1}{\sigma-1}}$.

Proof. The first-order condition of $\mathcal{P}$ with respect to c_j is

$$c_j = \beta_j \left(\frac{q_j}{w}\right)^{\sigma} C. \tag{39}$$

Raising both sides to the power $\frac{\sigma-1}{\sigma}$ and summing across j's yields

$$w = Q. \tag{40}$$

Using the first-order conditions (34), (36) and (39) into the resource constraints (4), we find

$$0 \quad \geq \quad \beta_j \left(\frac{q_j}{Q}\right)^{\sigma-1} \frac{C}{Q} + \sum_{k\in\mathcal{N}} \alpha_k \frac{\Omega_{jk} q_j^{\varepsilon_k - 1}}{\sum_{i\in\mathcal{N}} \Omega_{ik} q_i^{\varepsilon_k - 1}} \frac{l_k}{1-\alpha_k} - \frac{l_j}{1-\alpha_j} \tag{41}$$

for all $j \in \mathcal{N}$. Summing across j's and simplifying yields (8). Note that once q is known we can find l by inverting (41). We can then find y and x using the first-order conditions (34) and (36). □

H.2 Reshaping the planner's problem

This section contains the proofs of results related to the reshaping solution method.

[78] Note that $\sum_{i\in\tilde{\mathcal{N}}} \Omega_{ij}$ must be strictly positive for all j, otherwise that j is not part of an operating cycle.

Let $V_R : [0,1]^n \to \mathbb{R}$ be the objective function of $\mathcal{R}$ defined as

$$V_R(\theta) = \left(\sum_{j \in \mathcal{N}} (q_j(\theta))^{\sigma-1} \right)^{\sigma-1} \left(1 - \sum_{j \in \mathcal{N}} f_j \theta_j \right) L \tag{42}$$

where q_j is implicitly defined by (9). Similarly, let $V_P : \{0,1\}^n \to \mathbb{R}$ be the objective function of $\mathcal{P}$.

Preliminary result

The proof of Proposition 3 relies on the following lemma.

Lemma 4. *Let $F = A - fB$ where $f > 0$, A is the all-one $n \times n$ matrix and B is an $n \times n$ matrix. If B is negative definite on the subspace $S : \sum_{i=1}^n x_i = 0$ then F is positive definite for $f > 0$ small enough.*

Proof. The negative definiteness of B on S implies that $x'Bx \leq -d\|x\|^2$ for $x \in S$ and some $d > 0$. We can write any vector z as $z = x + y$ where $x \in S$ and $y \perp S$. Then,

$$\begin{aligned} z'(A - fB)z &= n\|y\|^2 - fx'Bx - fy'By - 2fy'Bx \\ &\geq (n - 1/2)\|y\|^2 + df\|x\|^2 - 2f\|B\|\|x\|\|y\| \end{aligned}$$

for f small enough. For f small enough, this last expression is strictly convex in $(\|x\|, \|y\|)$ with a minimum of 0 at $(0,0)$ or, equivalently, at $z = 0$. Since $z'(A - fB)z > 0$ for any $z \neq 0$, it follows that F is positive definite. $\square$

First-order conditions for θ_k

Lemma 2. The first-order conditions of the reshaped planner's problem $\mathcal{R}$ can be written as

$$(1 + a_k)\lambda_k c_k + \sum_{j \in \mathcal{N}} (1 + a_k + b_{kj})\lambda_k x_{kj} - \sum_{j \in \mathcal{N}}^{n} \lambda_j x_{jk} - wl_k - w\theta_k f_k L = \theta_k \Delta\mu_k,$$

where $\Delta\mu_k = \overline{\mu}_k - \underline{\mu}_k$ is the difference between the Lagrange multipliers on the constraints $\theta_k \leq 1$ and $\theta_k \geq 0$, respectively.

Proof. Combining (8), (7) and (9), we can write the planner's reshaped problem $\mathcal{P}$ as

$$\max_{\theta, q} \left(\sum_j \beta_j q_j^{\sigma-1} \right)^{\frac{1}{\sigma-1}} \left(1 - \sum_{j \in \mathcal{N}} \theta_j f_j \right) L$$

subject to

$$q_j \leq Az_j\theta_j^{a_j}\left(\sum_{i\in\mathcal{N}}\Omega_{ij}\left(\theta_i^{b_{ij}}q_i\right)^{\varepsilon_j-1}\right)^{\frac{1}{\varepsilon_j-1}\alpha_j}.$$

Denoting by ζ_j the Lagrange multiplier on the jth constraint, the first-order condition with respect to q_k is □

$$\beta_k\left(\frac{q_k}{Q}\right)^{\sigma-1}\frac{C}{Q}-\frac{\zeta_k q_k}{Q}+\sum_j\alpha_j\frac{\zeta_j q_j}{Q}\frac{\Omega_{kj}\theta_k^{b_{kj}(\varepsilon_j-1)}q_k^{\varepsilon_j-1}}{\sum_{i\in\mathcal{N}}\Omega_{ij}\left(\theta_i^{b_{ij}}q_i\right)^{\varepsilon_j-1}}=0. \tag{43}$$

Notice that this equation, together with the reshaped version of (41), implies that

$$\zeta_k q_k = \frac{1}{1-\alpha_k}wl_k. \tag{44}$$

The first-order condition with respect to θ_k is

$$-Q\theta_k f_k L + a_k q_k\zeta_k + \sum_j\alpha_j\zeta_j q_j b_{kj}\frac{\Omega_{kj}\theta_k^{b_{kj}(\varepsilon_j-1)-1}q_k^{\varepsilon_j-1}}{\sum_{i\in\mathcal{N}}\Omega_{ij}\left(\theta_i^{b_{ij}}q_i\right)^{\varepsilon_j-1}}=\theta_k\Delta\mu_k \tag{45}$$

where $\Delta\mu_k=\overline{\mu}_k-\underline{\mu}_k$ is the difference between the Lagrange multipliers on $\theta_k\leq 1$ and $\theta_k\geq 0$. We can combine the last equation with (34), (36) and (44) to find an equation that governs the entry of firms

$$-Q\theta_k f_k L + a_k\lambda_k y_k + \sum_j b_{kj}x_{kj}\lambda_k = \theta_k\Delta\mu_k.$$

Using the resource constraint for good k, this last equation can be written

$$(1+a_k)\lambda_k c_k + \sum_{j\in\mathcal{N}}(1+a_k+b_{kj})\lambda_k x_{kj} - \sum_{j\in\mathcal{N}}^{n}\lambda_j x_{jk} - wl_k - w\theta_k f_k L = \theta_k\Delta\mu_k, \tag{10}$$

which is the result.

Sufficient conditions for Karush-Kuhn-Tucker

Proposition 2. Let $\varepsilon_j=\varepsilon$ and $\alpha_j=\alpha$ for all j. If $\Omega_{ij}=d_i e_j$ for some vectors d and e then the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution to $\mathcal{R}$.

Proof. Raise both sides of (9) to the power $\varepsilon-1$, multiply by $d_j\theta_j^{b(\varepsilon-1)}$ and sum across j's to find

$$\sum_{j\in\mathcal{N}}d_j\left(\theta_j^b q_j(\theta)\right)^{\varepsilon-1}=\left(\sum_{j\in\mathcal{N}}d_j e_j^{\alpha}(Az_j)^{\varepsilon-1}\theta_j^{a(\varepsilon-1)+b(\varepsilon-1)}\right)^{\frac{1}{1-\alpha}}$$

so that, once combined with (9), we find

$$q_j(\theta) = Az_j\theta_j^a d_j^{\frac{\alpha}{\varepsilon-1}} \left(\sum_{i\in\mathcal{N}} d_i e_i^{\alpha} (Az_i)^{\varepsilon-1}\theta_i^{a(\varepsilon-1)+b(\varepsilon-1)}\right)^{\frac{\alpha}{1-\alpha}\frac{1}{\varepsilon-1}}.$$

Computing the log of Q, we get

$$\begin{aligned}\log(Q) =& \frac{1}{\sigma-1}\log\left(\left(\sum_{j\in\mathcal{N}}\beta_j\left(z_j\theta_j^a A d_j^{\frac{\alpha}{\varepsilon-1}}\right)^{\sigma-1}\right)\left(\sum_{i\in\mathcal{N}} d_i e_i^{\alpha}(Az_i)^{\varepsilon-1}\theta_i^{a(\varepsilon-1)+b(\varepsilon-1)}\right)^{\frac{\alpha}{1-\alpha}\frac{\sigma-1}{\varepsilon-1}}\right)\\ =& \frac{1}{\sigma-1}\log\left(\sum_{j\in\mathcal{N}}\beta_j\left(z_j\theta_j^a A d_j^{\frac{\alpha}{\varepsilon-1}}\right)^{\sigma-1}\right) + \frac{1}{\varepsilon-1}\frac{\alpha}{1-\alpha}\log\left(\sum_{i\in\mathcal{N}} d_i e_i^{\alpha}(Az_i)^{\varepsilon-1}\theta_i^{a(\varepsilon-1)+b(\varepsilon-1)}\right)\end{aligned}$$

If $0 < a \leq (\sigma-1)^{-1}$ and $-a \leq b \leq (\varepsilon-1)^{-1} - a$ (and in particular if ($\star$) holds) the exponents on θ are all between 0 and 1 so that the summations in $\log(Q)$ are concave functions of θ. The log of a concave function is concave so $\log(Q)$ is also concave. Moving toward the full objective function, the term $1-\sum_{j\in\mathcal{N}}\theta_j f_j$ is concave and so is $\log V_R$. 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 $\log(V_R(\theta))$ on the set $\theta\in[0,1]^n$. Since log is an increasing transformation, a solution to this problem also solves $\mathcal{R}$. □

Proposition 3. Let $\sigma = \varepsilon_j$ for all j. Suppose that the $\{\beta_j\}_{j\in\mathcal{N}}$ are not too far from each other and that the matrix Ω is close enough to $\bar{\Omega}$. Then there exists a threshold $\bar{f} > 0$ such that if $f_j < \bar{f}$ for all j the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution to $\mathcal{R}$.

Proof. To simplify the notation, define $p_j = q_j^{\sigma-1}$ and let

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

$\mathcal{R}$ can then be written as

$$\min_{p\in P} -\frac{1}{\sigma-1}\log\left(\sum_{j\in\mathcal{N}}\beta_j p_j\right) - \log\left(1-\sum_{j\in\mathcal{N}} f_j g^j(p)\right)$$

where $P = \left\{p\in\mathbb{R}^n_{\geq 0} : p_j \leq z_j^{\sigma-1}\left(\sum_{i\in\mathcal{N}}\Omega_{ij}p_i\right)^{\alpha_j}, \forall j\right\}$.

Denote the objective function by Λ. Its Hessian matrix has typical element

$$\frac{\partial^2 \Lambda}{\partial p_k \partial p_l} = \underbrace{\frac{1}{\sigma - 1} \beta_k \beta_l \left(\sum_j \beta_j p_j \right)^{-2}}_{A_{kl}} + \underbrace{\frac{\sum_j f_j g^j_{kl}(p)}{1 - \sum_j f_j g^j(p)}}_{B_{kl}} + \underbrace{\frac{\left(\sum_j f_j g^j_k(p) \right) \left(\sum_j f_j g^j_l(p) \right)}{\left(1 - \sum_j f_j g^j(p) \right)^2}}_{C_{kl}}, \tag{46}$$

and define A, B and C as the matrices with typical elements A_{kl}, B_{kl} and C_{kl}.

We will show that in the limit as $\Omega \to \bar{\Omega}$ and $\beta_j \to \bar{\beta}$ for all j the Hessian is positive definite when the largest fixed cost $\max_j f_j$ is small enough. To do so, we will rely on Lemma 4 above. For that purpose, notice that, in the limit, A is a positive multiple of the all-one matrix.

Pick $\bar{f} > 0$ and $\tilde{f}_j \in [0,1]^n$ so that $f_j = \bar{f} \tilde{f}_j$ for all j. Taking the derivatives of g, we find

$$B_{kl} = \frac{1}{L - \sum_j f_j g^j(p)} \left(\underbrace{\sum_{j \in \mathcal{N}} f_j \frac{\alpha_j (\alpha_j + 1) p_j \Omega_{kj} \Omega_{lj}}{z_j^{\sigma-1} \left(\sum_i \Omega_{ij} p_i \right)^{\alpha+2}}}_{B^1_{kl}} \right. \tag{47}$$

$$\left. - \bar{f} \underbrace{\left(\tilde{f}_k \frac{\alpha_k \Omega_{lk}}{z_k^{\sigma-1} \left(\sum_i \Omega_{ik} p_i \right)^{\alpha_k+1}} + \tilde{f}_l \frac{\alpha_l \Omega_{kl}}{z_l^{\sigma-1} \left(\sum_i \Omega_{il} p_i \right)^{\alpha_l+1}} \right)}_{B^2_{kl}} \right). \tag{48}$$

B^1 is a Gramian matrix where B^1_{kl} is the scalar product of a pair of vectors v_k and v_l defined as

$$v_m = \left[\sqrt{\frac{f_1 \alpha_1 (\alpha_1+1) p_1}{z_1^{\sigma-1} \left(\sum_i \Omega_{i1} p_i \right)^{\alpha_1+2}}} \Omega_{m1} \quad \cdots \quad \sqrt{\frac{f_j \alpha_j (\alpha_j+1) p_j}{z_j^{\sigma-1} \left(\sum_i \Omega_{ij} p_i \right)^{\alpha_j+2}}} \Omega_{mj} \quad \cdots \quad \sqrt{\frac{f_n \alpha_n (\alpha_n+1) p_n}{z_n^{\sigma-1} \left(\sum_i \Omega_{in} p_i \right)^{\alpha_n+2}}} \Omega_{mn} \right]'.$$

Since Gramian matrices are positive semi-definite, so is B^1. For B^2, we will show that, in the limit, it is negative definite on the subspace $S : \sum_{i=1}^n x_i = 0$. Define b as a vector with typical element $b_j = \tilde{f}_j \alpha_j \left[z_j^{\sigma-1} \left(\sum_i \Omega_{ij} p_i \right)^{\alpha_j+1} \right]^{-1}$. We can write $B^2 = \Omega \text{diag}(b) + (\Omega \text{diag}(b))'$. Take any vector $x \in S$, then in the limit,

$$\begin{aligned} x' B^2 x &= x' \left[\omega (O_n - I_n) \text{diag}(b) + (\omega (O_n - I_n) \text{diag}(b))' \right] x \\ &= x' \left[-2\omega I_n \text{diag}(b) \right] x < 0 \end{aligned}$$

for any $x \neq 0$. The matrix B^2 is therefore negative definite on S. Using Lemma 4, $A - \bar{f} B^2$ is therefore positive definite for $\bar{f} > 0$ small enough. Finally, the matrix C in (46), is also a Gramian matrix and its contribution to the Hessian is thus positive semi-definite.

Putting the pieces together, we have shown that the Hessian of the objective function Λ is positive definite for $\max_j f_j$ small enough when $\Omega = \bar{\Omega}$ and $\beta = \bar{\beta}$. Now, each element of the Hessian is also a continuous function of (Ω, β) in a neighborhood of $\left(\bar{\Omega}, \bar{\beta}\right)$.[79] Since the eigenvalues are continuous functions of the elements of a matrix, they are also continuous functions of Ω and β. There is therefore a ball $\mathcal{B} = \left\{(\Omega, \beta) : \left\|(\Omega, \beta) - \left(\bar{\Omega}, \bar{\beta}\right)\right\| < \delta\right\}$ for $\delta > 0$ such that the Hessian is also positive definite for $(\Omega, \beta) \in \mathcal{B}$.[80] □

Equivalence of solutions

Proposition 4. If $\theta^* \in \{0,1\}^n$ solves $\mathcal{R}$, then θ^* also solves $\mathcal{P}$.

Proof. By construction, the objective function V_{RR} of $\mathcal{R}$ and the objective function V_{SP} of $\mathcal{P}$ coincide over $\{0,1\}^n$. Therefore $V_R(\theta^*) = V_P(\theta^*)$. Since the feasible set of $\mathcal{R}$, $[0,1]^n$, contains the feasible set of $\mathcal{P}$, $\{0,1\}^n$, it must be that $V_P(\theta^*) \geq V_P(\theta)$ for $\theta \in \{0,1\}^n$, otherwise θ^* would not be a solution to $\mathcal{R}$. θ^* therefore solves $\mathcal{P}$. □

Marginal benefit and cost of increasing θ_j

Proposition 5. Under the $(\star)$ condition, the marginal benefit and the marginal cost of increasing θ_j only depend on θ_j through the aggregates F_j and G_j.

Proof. Rewrite $\mathcal{R}$ as

$$\max_{\theta \in [0,1]^n} \left(\sum_{j \in \mathcal{N}} \beta_j q_j^{\sigma-1}\right)^{\frac{1}{\sigma-1}} \left(1 - \sum_{j \in \mathcal{N}} f_j \theta_j\right) L$$

subject to $q_j \leq A z_j \theta_j^{a_j} B_j^{\alpha_j}$ for all $j \in \mathcal{N}$ and where $B_j = \left(\sum_{i \in \mathcal{N}} \theta_i^{b_{ij}(\varepsilon_j - 1)} \Omega_{ij} q_i^{\varepsilon_j - 1}\right)^{\frac{1}{\varepsilon_j - 1}}$. This problem is equivalent to $\mathcal{R}$ since the inequality constraints always bind at the optimum. The first-order conditions with respect to θ_k and q_k are

$$f_k Q + \bar{\mu}_k - \underline{\mu}_k = \zeta_k \frac{\partial q_k}{\partial \theta_k} + \sum_{j \in \mathcal{N}} \zeta_j \frac{\partial q_j}{\partial B_j} \frac{\partial B_j}{\partial \theta_k}$$

$$\frac{\partial Q}{\partial q_k} \left(1 - \sum_{j \in \mathcal{N}} f_j \theta_j\right) L = \zeta_k - \sum_{j \in \mathcal{N}} \zeta_j \frac{\partial q_j}{\partial B_j} \frac{\partial B_j}{\partial q_k}$$

where ζ_j is the Lagrange multiplier on the j-th inequality constraints, and $\bar{\mu}_j$ and $\underline{\mu}_j$ are the

[79] Elements of the Hessian become infinite at the boundary of P where $\sum_i \Omega_{ij} p_i = 0$ for some j. However, these points are not relevant to the planner under our assumptions. When the fixed costs are small enough and when Ω is close enough to $\bar{\Omega}$, each firm is connected to at least one producing firm at the optimum. Therefore, we can exclude these points easily adding the constraints $\sum_i \Omega_{ij} p_i \geq D$ for some $D > 0$ small. These constraints will never bind and the solution to the planner's problem is therefore unchanged.

[80] Since all norms are equivalent in a finite dimensional space, there is no need to specify one here.

Lagrange multipliers on the constraint $\theta_j \leq 1$ and $\theta_j \geq 0$.[81] Combining these first-order conditions yields

$$f_k Q + \bar{\mu}_k - \underline{\mu}_k = \left(1 - \sum_{j \in \mathcal{N}} f_j \theta_j\right) L \frac{\partial Q}{\partial q_k} \frac{\partial q_k}{\partial \theta_k} + \sum_{j \in \mathcal{N}} \zeta_j \frac{\partial q_j}{\partial B_j} \left(\frac{\partial B_j}{\partial \theta_k} + \frac{\partial B_j}{\partial q_k} \frac{\partial q_k}{\partial \theta_k}\right).$$

For the result to hold, we need $\frac{\partial Q}{\partial q_k} \frac{\partial q_k}{\partial \theta_k}$ and $\frac{\partial B_j}{\partial \theta_k} + \frac{\partial B_j}{\partial q_k} \frac{\partial q_k}{\partial \theta_k}$ to depend on θ_k only through aggregates. We can write

$$\frac{\partial Q}{\partial q_k} \frac{\partial q_k}{\partial \theta_k} = \left(\sum_{j \in \mathcal{N}} \beta_j q_j^{\sigma-1}\right)^{\frac{1}{\sigma-1}-1} \beta_k \left(A z_k \theta_k^{a_k} B_k^{\alpha_k}\right)^{\sigma-2} \times a_k A z_k \theta_k^{a_k-1} B_k^{\alpha_k}$$

where we see that θ_k drops out of the equation if $a_k = \frac{1}{\sigma-1}$. Similarly, we can write

$$\begin{aligned}\frac{\partial B_j}{\partial \theta_k} + \frac{\partial B_j}{\partial q_k} \frac{\partial q_k}{\partial \theta_k} = &\left(\sum_{i \in \mathcal{N}} \theta_i^{b_{ij}(\varepsilon_j-1)} \Omega_{ij} q_i^{\varepsilon_j-1}\right)^{\frac{1}{\varepsilon_j-1}-1} \Big(b_{kj} \theta_k^{b_{kj}(\varepsilon_j-1)-1} \Omega_{kj} q_k^{\varepsilon_j-1} \\ &+ \theta_k^{b_{kj}(\varepsilon_j-1)} \Omega_{kj} \left(\varepsilon_j - 1\right) q_k^{\varepsilon_j-2} a_k A z_k \theta_k^{a_k-1} B_k^{\alpha_k}\Big)\end{aligned}$$

and, by taking into account that q_k depends directly on θ_k, we see that θ_k drops out of the equation if we also impose $b_{kj} = \frac{1}{\varepsilon_j-1} - \frac{1}{\sigma-1}$. The first-order condition therefore only depends on θ_k through the aggregates B_k and Q. □

H.3 Undistorted and distorted equilibria

Proposition 6. For a given entry decision vector θ, distorted and undistorted equilibria are efficient. Furthermore, the equilibrium prices W and Q_{jk} are equal (up to a choice of numeraire) to the planner's Lagrange multipliers w and λ_j (associated with the resource constraints (5) and (4), respectively).

Proof. We begin by computing the first-order conditions of the firms in equilibrium. Under both equilibrium definitions, the cost-minimization problem of a firm j is

$$\min_{x,l} \sum_{i=1}^{n} Q_{ij} x_{ij} + W l_j$$

subject to

$$\frac{A}{\alpha_j^{\alpha_j} \left(1-\alpha_j\right)^{1-\alpha_j}} z_j \theta_j \left(\sum_{i=1}^{n} \Omega_{ij}^{\frac{1}{\varepsilon_j}} x_{ij}^{\frac{\varepsilon_j-1}{\varepsilon_j}}\right)^{\frac{\varepsilon_j}{\varepsilon_j-1}\alpha_j} l_j^{1-\alpha_j} \geq y_j,$$

[81] The partial derivatives of q_j are to be understood for the binding inequality constraint, i.e. $\frac{\partial q_j}{\partial \theta_j} = A z_j a_j \theta_j^{a_j-1} A B_j^{\alpha_j}$ and $\frac{\partial q_j}{\partial B_j} = A z_j \theta_j^{a_j} A \alpha_j B_j^{\alpha_j-1}$.

for some fixed output level y_j. Denote δ_j the Lagrange multiplier on the constraint, which also corresponds to the marginal cost of production. The first-order conditions with respect to l_j and x_{kj} are

$$(1-\alpha_j)\,\delta_j y_j = W l_j \tag{49}$$

$$\delta_j \frac{A}{\alpha_j^{\alpha_j}\left(1-\alpha_j\right)^{1-\alpha_j}}\theta_j z_j \alpha_j \left(\sum_{i=1}^{n} \Omega_{ij}^{\frac{1}{\varepsilon_j}} x_{ij}^{\frac{\varepsilon_j-1}{\varepsilon_j}}\right)^{\alpha_j \frac{\varepsilon_j}{\varepsilon_j-1}-1} \Omega_{kj}^{\frac{1}{\varepsilon_j}} x_{kj}^{\frac{\varepsilon_j-1}{\varepsilon_j}} l_j^{1-\alpha_j} = Q_{kj} x_{kj}. \tag{50}$$

We now show that these equations imply that the ratio W/δ_j in the equilibrium is equal to q_j in the efficient allocation. In the undistorted equilibrium, the standard pricing equation applies such that $Q_{kj} = \frac{\varepsilon_j}{\varepsilon_j-1}\frac{1}{1+s_{kj}^x}\delta_k = \delta_k$, under the assumed subsidy. In the distorted equilibrium, the price Q_{kj} is equal to the marginal cost of production δ_k. As a results, the equations (49) and (50) coincide with their efficient allocation counterparts, (34) and (35), if we replace δ_j with λ_j and W with w. We can therefore follow the same steps and find that the equation

$$\frac{W}{\delta_j} = z_j \theta_j A \left(\sum_{i=1}^{n} \Omega_{ij}\left(\frac{W}{\delta_i}\right)^{\varepsilon_j-1}\right)^{\frac{\alpha_j}{\varepsilon_j-1}}, \tag{51}$$

holds. Since this equation is the same as (6), it follows that $\frac{W}{\delta_j} = q_j = \frac{w}{\lambda_j}$ for all j.

Next we turn to consumption c_j. Under both equilibrium definitions, the usual markup over marginal cost pricing rule applies and

$$P_j = \frac{\sigma}{\sigma-1}\delta_j,$$

and

$$c_j = \beta_j C\left(\frac{P_j}{\overline{P}}\right)^{-\sigma} = \beta_j C\left(\frac{\frac{\sigma}{\sigma-1}\delta_j}{\overline{P}}\right)^{-\sigma} \text{ and } \overline{P} = \left(\sum_{j=1}^{n}\beta_j P_j^{1-\sigma}\right)^{\frac{1}{1-\sigma}} = \frac{\sigma}{\sigma-1}\left(\sum_{j=1}^{n}\beta_j \delta_j^{1-\sigma}\right)^{\frac{1}{1-\sigma}}.$$

It will be convenient to normalize the aggregate price level so that $\overline{P} = \frac{\sigma}{\sigma-1}$ and $\left(\sum_j \beta_j \delta_j^{1-\sigma}\right)^{\frac{1}{1-\sigma}} = 1$, and we can write

$$c_j = \beta_j C\left(\delta_j\right)^{-\sigma}, \tag{52}$$

which is exactly the first-order condition (39) of the planner.

We now show that the equilibrium prices correspond to the planner's Lagrange multipliers.

Since $\frac{W}{\delta_j} = q_j$ we can write

$$1 = \left(\sum_{j=1}^{n} \beta_j \delta_j^{1-\sigma} \right)^{\frac{1}{1-\sigma}} = \left(\sum_{j=1}^{n} \beta_j \left(\frac{W}{q_j} \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} = W \times Q^{-1},$$

and since $Q = w$ by (40) we find that $W = w$. It follows immediately that $\delta_j = \lambda_j$.

We have shown that all first-order conditions are the same in the equilibria and the planner's problem, and that all prices are equal to their respective shadow value in the efficient allocation.[82] Since the resources constraints are also the same, the distorted and undistorted equilibria coincide with the efficient allocation, for an exogenously given entry vector θ. □

Proposition 7. A vector $\theta = \{0, 1\}^n$ that satisfies the first-order conditions of the reshaped problem $\mathcal{R}$ with reshaping parameters ($\star$) is an undistorted equilibrium.

Proof. Because of Proposition 6, we just need to show that the vector θ that solves the reshaped problem's first-order conditions is also an undistorted equilibrium.

Recall from (10) that the planner entry decision for firm k can be written as

$$(1 + a_k) \lambda_k c_k + \sum_{j=1}^{n} (1 + a_k + b_{kj}) x_{kj} \lambda_k - \sum_{j=1}^{n} \lambda_j x_{jk} - w l_k - w \theta_k f_k L = \theta_k \Delta \mu_k,$$

which, using the first-order conditions (34) and (35) and the resource constraint (4), boils down to

$$a_k \lambda_k c_k + \sum_{j \in \mathcal{N}} (a_k + b_{kj}) x_{kj} \lambda_k - w \theta_k f_k L = \theta_k \Delta \mu_k.$$

We can combine this equation with the planner's consumption decision (39) and the previously-derived equation for x_{kj} (37) to write

$$a_k \lambda_k \beta_k \lambda_k^{-\sigma} C + \sum_{j \in \mathcal{N}} (a_k + b_{kj}) \lambda_k \Omega_{kj} X_j \left(\frac{\lambda_k}{\left(\sum_i \Omega_{ij} \lambda_i^{1-\varepsilon_j} \right)^{\frac{1}{1-\varepsilon_j}}} \right)^{-\varepsilon_j} - w \theta_k f_k L = \theta_k \Delta \mu_k.$$

Now suppose that $\theta_k > 0$. We can divide this equation by θ_k, such that

$$a_k \beta_k \frac{\lambda_k^{1-\sigma}}{\theta_k} C + \sum_{j \in \mathcal{N}} (a_k + b_{kj}) \frac{\lambda_k^{1-\varepsilon_j}}{\theta_k} \Omega_{kj} X_j \left(\frac{1}{\left(\sum_i \Omega_{ij} \lambda_i^{1-\varepsilon_j} \right)^{\frac{1}{1-\varepsilon_j}}} \right)^{-\varepsilon_j} - w f_k L = \Delta \mu_k. \tag{53}$$

It turns out that this equation is also valid for $\theta_k = 0$. To see this, notice that in this case $\lambda_k \to \infty$

[82] Recall that for a given θ the problem of the social planner is convex such that first-order conditions are sufficient to characterize the efficient allocation.

(since $q_k = 0$), and that the ratios $\frac{\lambda_k^{1-\sigma}}{\theta_k}$ and $\frac{\lambda_k^{1-\varepsilon_j}}{\theta_k}$ remain finite. Indeed, from (6) we can write

$$\frac{1}{\theta_j}\left(\frac{w}{\lambda_j}\right)^{\sigma-1} = \frac{1}{\theta_j}\left(z_j\theta_j^{a_j}A\left(\sum_{i\in\mathcal{N}}\Omega_{ij}\left(\theta_i^{b_{ij}}q_i\right)^{\varepsilon_j-1}\right)^{\frac{\alpha_j}{\varepsilon_j-1}}\right)^{\sigma-1}$$

$$= \left(z_jA\left(\sum_{i\in\mathcal{N}}\Omega_{ij}\left(\theta_i^{b_{ij}}q_i\right)^{\varepsilon_j-1}\right)^{\frac{\alpha_j}{\varepsilon_j-1}}\right)^{\sigma-1} \tag{54}$$

such that $\frac{\lambda_k^{1-\sigma}}{\theta_k}$ is indeed well-defined when $\theta_k = 0$ under ($\star$). For $\frac{\lambda_k^{1-\varepsilon_j}}{\theta_k}$, note that since $\theta \in \{0,1\}^n$, we can write (6) as

$$q_j = z_j\theta_j^{\frac{1}{\varepsilon_j-1}}A\left(\sum_{i\in\mathcal{N}}\Omega_{ij}q_i^{\varepsilon_j-1}\right)^{\frac{\alpha_j}{\varepsilon_j-1}},$$

since $\varepsilon_j > 1$ for all j. Then this equation can be reorganized as

$$\frac{1}{\theta_j}\left(\frac{w}{\lambda_j}\right)^{\varepsilon_j-1} = \left(z_jA\left(\sum_{i\in\mathcal{N}}\Omega_{ij}\left(\frac{w}{\lambda_i}\right)^{\varepsilon_j-1}\right)^{\frac{\alpha_j}{\varepsilon_j-1}}\right)^{\varepsilon_j-1}, \tag{55}$$

and so $\frac{\lambda_k^{1-\varepsilon_j}}{\theta_k}$ is also well-defined when $\theta_k = 0$.

We now compare (53), to its equivalent in the undistorted equilibrium. Combining the profits (20) with the resource constraint (17) and the demand curves (52) and (15), we can write the entry decision of a firm j as

$$\frac{1}{\sigma-1}\delta_j\beta_jC\frac{\delta_j^{1-\sigma}}{\theta_j} + \sum_{i\in\mathcal{N}}\frac{1}{\varepsilon_i-1}\frac{\delta_j^{1-\varepsilon_j}}{\theta_j}\Omega_{ji}X_i\left(\frac{1}{\left(\sum_k\Omega_{ki}\delta_k^{1-\varepsilon_i}\right)^{\frac{1}{1-\varepsilon_i}}}\right)^{-\varepsilon_i} - Wf_jL \geq 0 \tag{56}$$

where we have used the fact that $Q_{ji} = \delta_j$. Here, notice that if $\theta_j = 1$ the expressions $\frac{\delta_j^{1-\sigma}}{\theta_j} = \delta_j^{1-\sigma}$ and $\frac{\delta_j^{1-\varepsilon_j}}{\theta_j} = \delta_j^{1-\varepsilon_j}$ correspond to the actual prices of the firm raised to the powers $1-\sigma$ or $1-\varepsilon_j$. If instead $\theta_j = 0$, these expressions correspond to the *counterfactual* prices of the firm (raised to the appropriate powers), if it were to enter. This can be seen from the right-hand sides of (54) and (55), and by remembering from the proof of Proposition (6) that the same equations hold with δ_j and W instead of λ_j and w. As a result, (56) is the correct equation to think of entry given our equilibrium concept.

Now, suppose that in the reshaped planner's problem $\theta_k = 1$, then in must be that $\Delta\mu_k \geq 0$ by the slackness condition, and so the left-hand side of (53) is positive (under the reshaping parameters $\star$). Then the left-hand side of (56) is also positive and the firm enters in the undistorted equilibrium.

Similarly, if $\theta_k = 0$, $\Delta\mu_k < 0$ by the complementary slackness conditions and the firm does not enter in equilibrium. $\square$

Proposition 8. A vector $\theta = \{0,1\}^n$ that satisfies the first-order conditions of the reshaped problem $\mathcal{R}$ with parameters (22) is a distorted equilibrium.

Proof. The steps are essentially the same as for Proposition 8, except that we use the reshaping parameters (22) instead and the entry equation (21) for the equilibrium. $\square$

H.4 Forces shaping the network

Proposition 9. Taking equilibrium quantities as given, the following holds.

1. In an undistorted equilibrium, operating a firm that is directly upstream or downstream from j increases π_j^{undist}.
2. In a distorted equilibrium, operating a firm that is directly upstream from j increases π_j^{dist}.

Proof. Consider a firm i that is directly upstream from j so that $\Omega_{ij} > 0$. A transition from $\theta_i = 0$ to $\theta_i = 1$ results in a decline in δ_j through the recursive structure of (19). This in turn, leads to an increase in $\delta_j c_j$ and $\delta_j x_{jk}$ for all k by (24). It follows that both π_j^{undist} and π_j^{dist} increase as a result of the change in θ_i. Now consider a firm k that is directly downstream from j so that $\Omega_{jk} > 0$. The transition from $\theta_k = 0$ to $\theta_k = 1$ leads to an increase in k's purchases of j's good (second term in (23)). Since in the undistorted equilibrium j earns a markup above its marginal cost for each unit sold, π^{undist} increases. We see from (25) that this channel is absent from π_j^{dist} and so the operating decisions of upstream firms have no direct impact on j's operating decision. $\square$

Proposition 10. Let $\mathcal{J} \subset \mathcal{N}$ be a group of firms. Denote by $\theta^+ \in \{0,1\}^n$ the operating vector when the firms in $\mathcal{J}$ operate ($\theta_j^+ = 1$ for $j \in \mathcal{J}$). Similarly, let $\theta^- \in \{0,1\}^n$ be the operating vector when the firms in $\mathcal{J}$ do not operate ($\theta_j^- = 0$ for $j \in \mathcal{J}$). For all $j \notin \mathcal{J}$, assume $\theta_j^+ = \theta_j^-$. Denote by z^+ and z^- two productivity vectors such that $z_j^+ \geq z_j^-$ for all $j \in \mathcal{J}$ and $z_j^+ = z_j^-$ for $j \notin \mathcal{J}$. Then

$$C_{z^+}\left(\theta^+\right) - C_{z^+}\left(\theta^-\right) \geq C_{z^-}\left(\theta^+\right) - C_{z^-}\left(\theta^-\right),$$

where $C_z(\theta)$ denotes consumption under the productivity vector z and the operating vector θ.

Proof. First, notice that $C_{z^+}(\theta^-) = C_{z^-}(\theta^-)$ since in either cases firms in $\mathcal{J}$ are not operating and the difference in their productivities is therefore irrelevant. We need to show that $C_{z^+}(\theta^+) \geq C_{z^-}(\theta^+)$. From (7), (8) and (6) we can write

$$C_z(\theta) = \left(\sum_{j=1}^{n} \beta_j q_j^{\sigma-1}\right)^{\frac{1}{\sigma-1}} \left(1 - \sum_{j=1}^{n} \theta_j f_j\right) L \tag{57}$$

where q_j is the (unique) fixed point of

$$q_j = z_j \theta_j A \left(\sum_{i=1}^{n} \Omega_{ij} q_i^{\varepsilon_j - 1} \right)^{\frac{\alpha_j}{\varepsilon_j - 1}}. \tag{6}$$

When comparing $C_{z^+}(\theta^+)$ and $C_{z^-}(\theta^+)$ the vector θ is the same and, therefore, so is $\sum_{j=1}^{n} \theta_j f_j$. The only difference in the objective function comes from the vector q. We now show that $q_{z^+}(\theta) \geq q_{z^-}(\theta)$. Recall from the proof of Lemma 1 that we can find the fixed point q by iterating on (6). If we start the iterations for $q_{z^+}(\theta^+)$ from $q_{z^-}(\theta^+)$, we find that for all j

$$\begin{aligned} q_{z^+,j}^{(1)}(\theta^+) &= z_j^+ \theta_j^+ A \left(\sum_{i=1}^{n} \Omega_{ij} \left(q_{z^-,i}(\theta^+) \right)^{\varepsilon_j - 1} \right)^{\frac{\alpha_j}{\varepsilon_j - 1}} \\ &\geq z_j^- \theta_j^+ A \left(\sum_{i=1}^{n} \Omega_{ij} \left(q_{z^-,i}(\theta^+) \right)^{\varepsilon_j - 1} \right)^{\frac{\alpha_j}{\varepsilon_j - 1}} = q_{z^-,j}(\theta^+), \end{aligned}$$

where $q_{\Omega^+,j}^{(n)}(\theta^+)$ denote the nth term in the iteration sequence. But since the right-hand side of (6) is an increasing function of q, we have that $q_{z^+,j}^{(n+1)}(\theta) \geq q_{z^+,j}^{(n)}(\theta)$ for all j and n, and it follows that $q_{z^+}(\theta) \geq q_{z^-}(\theta)$. Since $\sum_{j=1}^{n} \beta_j q_j^{\sigma-1}$ is increasing in q, it is then immediate by (57) that $C_{z^+}(\theta^+) \geq C_{z^-}(\theta^+)$. □

Proposition 11. Let $\mathcal{J} \subset \mathcal{N}$ be a group of firms. Denote by $\theta^+ \in \{0,1\}^n$ the operating vector when the firms in $\mathcal{J}$ operate ($\theta_j^+ = 1$ for $j \in \mathcal{J}$). Similarly, let $\theta^- \in \{0,1\}^n$ be the operating vector when the firms in $\mathcal{J}$ do not operate ($\theta_j^- = 0$ for $j \in \mathcal{J}$). For all $j \notin \mathcal{J}$, assume $\theta_j^+ = \theta_j^-$. Denote by Ω^- a network of potential connections and let Ω^+ be identical to Ω^- except that it has an additional connection between two firms in $\mathcal{J}$.[83] Then

$$C_{\Omega^+}(\theta^+) - C_{\Omega^+}(\theta^-) \geq C_{\Omega^-}(\theta^+) - C_{\Omega^-}(\theta^-),$$

where $C_\Omega(\theta)$ denotes consumption under the potential network Ω and the operating vector θ.

Proof. First note that $C_{\Omega^+}(\theta^-) = C_{\Omega^-}(\theta^-)$ since in either case the firms in $\mathcal{J}$ are not operating and so the extra link in $\Omega^+$is irrelevant. We therefore have to show that $C_{\Omega^+}(\theta^+) \geq C_{\Omega^-}(\theta^+)$. From (7), (8) and (6) we can write

$$C_\Omega(\theta) = \left(\sum_{j=1}^{n} \beta_j q_j^{\sigma-1} \right)^{\frac{1}{\sigma-1}} \left(1 - \sum_{j=1}^{n} \theta_j f_j \right) L \tag{58}$$

[83]Formally, there are two firms $k \in \mathcal{J}$ and $l \in \mathcal{J}$ such that $\Omega_{ij}^+ = \Omega_{ij}^-$ for all pairs $(i,j) \neq (k,l)$ and $\Omega_{kl}^+ > \Omega_{kl}^-$.

where q_j is the (unique) fixed point of

$$q_j = z_j \theta_j A \left(\sum_{i=1}^{n} \Omega_{ij} q_i^{\varepsilon_j - 1} \right)^{\frac{\alpha_j}{\varepsilon_j - 1}}. \tag{6}$$

When comparing $C_{\Omega^+}(\theta^+)$ and $C_{\Omega^-}(\theta^+)$ the vector θ is the same and, therefore, so is $\sum_{j=1}^{n} \theta_j f_j$. The only difference in the objective function comes from the vector q. We now show that $q_{\Omega^+}(\theta) \geq q_{\Omega^-}(\theta)$. Recall from the proof of Lemma 1 that we can find the fixed point q by iterating on (6). If we start the iterations for $q_{\Omega^+}(\theta^+)$ from $q_{\Omega^-}(\theta^+)$, we find that for all j

$$\begin{aligned} q^{(1)}_{\Omega^+,j}\left(\theta^+\right) &= z_j \theta_j^+ A \left(\sum_{i=1}^{n} \Omega_{ij}^+ \left(q_{\Omega^-,i}\left(\theta^+\right)\right)^{\varepsilon_j - 1} \right)^{\frac{\alpha_j}{\varepsilon_j - 1}} \\ &\geq z_j \theta_j^+ A \left(\sum_{i=1}^{n} \Omega_{ij}^- \left(q_{\Omega^-,i}\left(\theta^+\right)\right)^{\varepsilon_j - 1} \right)^{\frac{\alpha_j}{\varepsilon_j - 1}} = q_{\Omega^-,j}\left(\theta^+\right), \end{aligned}$$

where $q^{(n)}_{\Omega^+,j}(\theta^+)$ denote the nth term in the iteration sequence. But since the right-hand side of (6) is an increasing function of q, we have that $q^{(n+1)}_{\Omega^+,j}(\theta) \geq q^{(n)}_{\Omega^+,j}(\theta)$ for all j and n, and it follows that $q_{\Omega^+}(\theta) \geq q_{\Omega^-}(\theta)$. Since $\sum_{j=1}^{n} \beta_j q_j^{\sigma-1}$ is increasing in q, it is then immediate by (58) that $C_{\Omega^+}(\theta^+) \geq C_{\Omega^-}(\theta^+)$. □

Proposition 12. Let $\theta^*(z)$ be the efficient allocation under z and let $C(\theta, z)$ be consumption under (θ, z). Then the response of consumption after a change in productivity from z to z' is such that

$$\underbrace{C\left(\theta^*\left(z'\right), z'\right) - C\left(\theta^*(z), z\right)}_{\text{Change in consumption under the flexible network}} \geq \underbrace{C\left(\theta^*(z), z'\right) - C\left(\theta^*(z), z\right)}_{\text{Change in consumption under the fixed network}}.$$

Proof. By definition $\theta^*(z')$ maximizes welfare under z'. This implies that

$$C\left(\theta^*\left(z'\right), z'\right) \geq C\left(\theta^*(z), z'\right).$$

Subtracting $C(\theta^*(z), z)$ from both sides yields the result. □

H.5 Aggregate fluctuations

Proposition 14. If $\alpha_j = \alpha$ for all $j \in \mathcal{N}$, then the efficient network θ does not depend on aggregate productivity A.

Proof. Using Proposition 1 and Lemma 1, we can rewrite $\mathcal{P}$ as maximizing (8) over the set $\theta \in \{0,1\}^n$ where the vector q solves, for each $j \in \mathcal{N}$, (6). Let's denote this problem as

$\mathcal{P}_A$, where A refers to the aggregate productivity level. We will show that the optimal vector θ_A that solves $\mathcal{P}_A$ also solves an alternative problem $\mathcal{P}_{\tilde{A}}$ in which aggregate productivity is $\tilde{A}$ instead. Define $p_j = \left(\frac{\tilde{A}}{A}\right)^{\frac{1}{1-\alpha}} q_j$, then the objective function of $\mathcal{P}_{\tilde{A}}$ can be written as $\left(\tilde{A}/A\right)^{\frac{1}{1-\alpha}} \left(\sum_{j\in\mathcal{N}} \beta_j p_j^{\sigma-1}\right)^{\frac{1}{\sigma-1}} \left(1 - f_j \sum_{j\in\mathcal{N}} \theta_j\right) L$ and its recursive equation (6) can be written as $p_j = z_j \theta_j A \left(\sum_{i\in\mathcal{N}} \Omega_{ij} \left(p_i\right)^{\varepsilon_j - 1}\right)^{\frac{\alpha}{\varepsilon_j - 1}}$. Since the constant $\left(\tilde{A}/A\right)^{\frac{1}{1-\alpha}}$ does not affect the maximization, $\mathcal{P}_{\tilde{A}}$ is the same problem as $\mathcal{P}_A$, and, at the optimum, $p_{\tilde{A}} = q_A$ and $\theta_{\tilde{A}} = \theta_A$. The production network is therefore invariant to changes in aggregate productivity A. □

H.6 Stable equilibrium

The following lemma is useful to show that every stable equilibrium is efficient.

Lemma 5. *Let $X, Y \subset \mathbb{R}^n$. Define Problem A as*

$$\sup_{x\in X, y\in Y} f\left(x,y\right) \text{ subject to } \sum_j y_j \leq 0$$

and Problem B as

$$\sup_{x\in X, y\in Y} g\left(f\left(x,y\right)\right) - \lambda \left(\sum_j y_j\right)$$

where g is a strictly increasing function and where λ is such that $\sum_j y_j = 0$ at any solution. Suppose that for any solution to Problem A the constraint binds, then Problems A and B have the same solutions.

Proof. Take a point $\left(x^A, y^A\right)$ that solves Problem A and such that, since the constraint binds, $\sum_j y_j^A = 0$. Towards a contradiction, suppose $\left(x^A, y^A\right)$ does not solve Problem B. Then there is another point $\left(\tilde{x}, \tilde{y}\right)$ such that $\sum_j \tilde{y}_j = 0$ (by the definition of λ) and such that $g\left(f\left(\tilde{x},\tilde{y}\right)\right) - \lambda\left(\sum_j \tilde{y}_j\right) > g\left(f\left(x^A, y^A\right)\right) - \lambda\left(\sum_j y_j^A\right)$. Since g is strictly increasing this implies that $f\left(\tilde{x},\tilde{y}\right) > f\left(x^A, y^A\right)$ but, since $\left(\tilde{x},\tilde{y}\right)$ is in the feasible set of Problem A, this implies that $\left(x^A, y^A\right)$ was not a solution to Problem A, which is a contradiction. Conversely, take a point $\left(x^B, y^B\right)$ that solves Problem B. Then by the definition of λ it must be that $\sum_j y_j^B = 0$. Towards a contradiction, suppose $\left(x^B, y^B\right)$ does not solve Problem A. Then there is another point $\left(\tilde{x},\tilde{y}\right)$ such that $\sum_j \tilde{y}_j = 0$ (since the constraint in Problem A binds at the optimum) and such that $f\left(\tilde{x},\tilde{y}\right) > f\left(x^B, y^B\right)$. Since g is strictly increasing this implies that $g\left(f\left(\tilde{x},\tilde{y}\right)\right) - \lambda\left(\sum_j \tilde{y}_j\right) > g\left(f\left(x^B, y^B\right)\right) - \lambda\left(\sum_j y_j^B\right)$ so that $\left(x^B, y^B\right)$ is not a solution to Problem B, which is a contradiction. □

Proposition 13. Every stable equilibrium is efficient.

Proof. The proof proceeds by establishing restrictions that any stable equilibrium must satisfy. It then shows that any allocation that satisfies these restrictions must be efficient.

Consider a coalition made of all the firms in the economy. For the equilibrium to be stable there cannot be an alternative arrangement that would yield larger aggregate profits. Otherwise, transfers could be designed to make one firm better off while keeping the other firms at the same profit level. The arrangement $\{x_{ij}, T_{ij}\}_{i,j}$ must therefore maximize $\sum_{j\in\mathcal{N}} \pi_j$. But, by the definition of an equilibrium, this maximization is subject to the behavior of the firms. Any equilibrium allocation therefore solves

$$\max_{\{x_{ij},T_{ij}\}_{i,j}} \sum_{j\in\mathcal{N}} \left\{ \max_{\{p_j,c_j,l_j,\theta_j\}} \pi_j \left(p_j, c_j, l_j, \theta_j, \{x_{ij}\}_{ij}\right) \text{ s.t. (31) and (32)} \right\}. \tag{59}$$

It is, however, equivalent to let the coalition itself directly optimize over $\{p_j, c_j, l_j, \theta_j\}_j$. To see this, notice that, conditional on the arrangement, the inner maximization problems in (59) are all independent from each other. In other words, the decisions of a firm i have no effect on the profit of a firm j as long as the contracts specified by the arrangement are fulfilled. As a result, we can write (59) as $\max_{\{x_{ij}\}_{ij},\{c_j,l_j,\theta_j\}_j} \sum_{j\in\mathcal{N}} \pi_j$ subject to the constraints (31) and (32) for all firms. By including the household's demand curves directly in the objective function, and by using the definition of π_j, the absence of dominating deviations therefore implies that the allocation must solve

$$\max_{\{x_{ij}\}_{ij},\{c_j,l_j,\theta_j\}_j} C^{\frac{1}{\sigma}} P \sum_{j\in\mathcal{N}} \beta_j^{\frac{1}{\sigma}} c_j^{\frac{\sigma-1}{\sigma}} - w \sum_{j\in\mathcal{N}} \left(l_j + \theta_j f_j L\right)$$

subject to (31) for all $j \in \mathcal{N}$, and where C and P are taken as given. Now, by Lemma 5 this problem is equivalent to an alternative problem in which the coalition maximizes $\left(\sum_{j\in\mathcal{N}} \beta_j^{\frac{1}{\sigma}} c_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$ subject to $\sum_j l_j + \theta_j f_j L \leq L$ in addition to the other constraints.[84] This reformulated problem is identical to the problem $\mathcal{P}$ of the social planner such that any stable equilibrium must be efficient. □

[84]The corresponding function g is $g(x) = x^{\frac{\sigma}{\sigma-1}}$. The constraints (31) can be included directly in the function f in Lemma 5 by setting $f = -\infty$ for points outside the constraint set.