Department of Economics

Essays on Applied Network Theory

Mariya Teteryatnikova

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Florence
January 2010

EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

Essays on Applied Network Theory

Mariya Teteryatnikova

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Jury Members:

Prof. Franklin Allen, University of Pennsylvania
Prof. Morten Ravn, University College London (Supervisor)
Prof. Fernando Vega-Redondo, EUI
Prof. Yves Zenou, Stockholm University

Declaration of Authorship

I, MARIYA TETERYATNIKOVA, declare that this thesis titled, 'ESSAYS ON AP-PLIED NETWORK THEORY' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date:

_____

i

Introduction

Network economics is a fast growing area of study, with a lot of potential for addressing a wide variety of socio-economic phenomena. Networks literally permeate our social and economic lives. The unemployed find jobs using the information and assistance of their friends and relatives. Consumers benefit from the research of friends and family into new products. In medicine and other technical fields, professional networks shape research patterns. In all these settings, the well-being of participants depends on social, geographic, or trading relationships. The countless ways in which network structures affect our well-being make it critical to understand: (i) how network structures impact behavior, (ii) what can be done, in the way of design by policymakers, to improve systemic outcomes. This area of study, broadly called network economics, is at the heart of my research interests.

In my dissertation I focus on three specific applications of network theory. The first application concerns networks in trade, where network structure represents the organization of trade agreements between countries. The second application deals with networks in financial market, and the network is used to model the structure of interbank exposures. Lastly, for the third application, I consider networks in labor markets and migration. In this context, the network represents the structure of social relations between people. Each of these applications of network analysis is addressed by one of three chapters in the thesis.

The first chapter is devoted to studying the impact of the structure of the trade network and the effects of expansion of this network, in a process of trade liberalization, on R&D decisions and on the subsequent productivity levels of firms. Recent empirical evidence has shown that, in general, trade liberalization promotes innovation and productivity growth in individual firms. In my work, I assert that different types of trade liberalization – multilateral versus regional – may lead to different R&D and productivity levels of firms. The network of trade agreements is modelled by a set of nodes, representing countries, and a set of links, indicating trade agreements between the linked countries. In this framework, the multilateral trade agreement is represented by the *complete* network, while the overlap of regional trade agreements is represented by the *hub-and-spoke* trade system. Trade liberalization, which increases the network of trade agreements, reinforces the incentives for firms to invest in R&D through the creation of new markets (scale effect) but it may also dampen these incentives through the emergence of new competitors (competition effect). The joint action of these two effects within the multilateral and the regional trade systems gives rise to the result that, for the same number of direct trade partners, the R&D effort of a country in the multilateral agreement is lower than the R&D effort of a hub but higher than the R&D effort of a spoke. This

suggests that productivity gains of regionalism versus those of multilateralism depend heavily on the relative number of regional trade agreements signed by countries. If a country signs relatively large number of trade agreements within the regional system (core country), then its R&D and productivity are higher than R&D and productivity of a country in the multilateral system. At the same time, a country that signs a relatively small number of trade agreements within the regional system (periphery country) has lower productivity gains than a country in the multilateral system. Additionally, I find that the aggregate level of R&D activities within the multilateral trade agreement exceeds that in the *star*, the simplest representative of the hub-and-spoke trade system.

In the second chapter, I consider the network of lending and borrowing relationships between banks and study the impact of the structure of this network on systemic risk and the scale of systemic breakdown in the interbank market. I focus on the risk and potential impact of system-wide defaults in the frequently observed "tiered" banking system, where first-tier institutions, that are relatively few in number, are connected with second-tier "peripheral" banks and are also connected with each other, while the peripheral banks are almost exclusively connected with the first-tier banks. The banking network is constructed from a number of banks which are linked by interbank exposures with a certain predefined probability. In this framework, the tiered structure is represented either by a network with negative correlation in connectivity of neighboring banks, or alternatively, by a network with a scale-free distribution of connectivity across banks. The main finding of this chapter highlights the advantages of tiering within the banking system, in terms of both the resilience of the banking network to systemic shocks and the extent of necessary government intervention should a crisis evolve. Specifically, the tiered network structure, showing negative correlations in bank connectivity, is found to be less prone to systemic breakdown than other structures. Moreover, in the scale-free tiered system, the resilience of the system to shocks increases as the *level of tiering* grows. Also, the *targeted* government bail-out policy aimed at rescuing the most highly connected failing banks in the first place, is expected to be more effective and induce lower costs in a tiered system with high level of tiering.

Finally, in the third chapter, I examine the puzzling case of 19th century Italian migration to the New World, when migrants from the north and the south of Italy revealed different preferences in their location choice and willingness to migrate. I study the origins of these regional variations focusing on the potential impact of social links between people on their migration decisions. Specifically, I examine whether the *initial small difference* in the number of social connections of the Northern and Southern Italians in North America could have resulted in the subsequently *large difference* in the size of migration flows. To address this issue, I develop a model in which there is a home country and a destination country and agents in both countries share information about

job opportunities through an explicitly modelled network of social relations. Agents in the home country compare their employment opportunities at home and abroad and take once-and-for-all migration decisions via strategic interaction in a game. I study the effects of a social network on agents' equilibrium migration decisions and in particular, the effects of a small change in the number of agents' contacts in the destination country on the size of the migration flow. Using the results of analytical work and numerical simulations, I find that even a marginal increase in the number of links between countries may lead to a substantial increase in migration. For example, in the specific case when the network of social relations is complete and agents primarily care about their long-run expected income, there exists a threshold value for the number of social links to the destination country, such that no one migrates at the *maximal equilibrium* if the actual number of these links is below the threshold but everyone migrates if the number of these links is above the threshold. These results provide a support for the hypothesis that initially small variation in the number of North American contacts of the Northern and Southern Italians could have led to significant discrepancy in regional emigration.

Acknowledgements

This thesis would not have been possible without the supervision of my advisors, Morten Ravn and Fernando Vega-Redondo, as well as my former advisor, Karl Schlag. I would like to thank them for their timely advice, guidance and support in my work. I would also like to thank Pascal Courty, Omar Licandro, Elena Carletti, Yves Zenou, Francis Bloch, Matthew Jackson, Michele Ruta, Roberta Piermartini, Maarten Janssen, Juan Ruiz, and my friends at the EUI, Anna Thum, Roman Chuhay, Graham Mott, Kasia Maciejowska, Konrad Smolinski, Michal Lewandowski, Teodora Borota, Javier Rivas, Sebastian Krautheim for their advice and useful comments. I gratefully acknowledge Antony Calvó-Armengol for giving me inspiration and sharing his knowledge with me. Lastly, I am very grateful to my parents for their invaluable support.

Contents

List of Figures

List of Tables

Chapter 1

R&D in the Network of International Trade: Multilateral versus Regional Trade Agreements

1.1   Introduction

In the era of unprecedented proliferation of regional trade agreements and simultaneous developments in the WTO, assessments of the relative economic benefits of multilateralism versus regionalism take on special significance. Numerous studies investigate the difference in welfare benefits, trade volumes, GDP levels and GDP growth rates across multilateral and regional trade arrangements.[1] However, the existing literature has not examined the issue of possible variations of the impact of different types of trade liberalization on countries' *productivity*.

The latter is surprising for at least two reasons. First, productivity is a key determinant of aggregate output, which is found to vary across different types of trade arrangements. Secondly, recent empirical evidence has shown that, in general, trade liberalization has a *significant effect* on the productivity level of the country. Bustos (2007) finds that during the period of trade liberalization between Argentina and Brazil, companies in sectors benefiting from a comparatively higher reduction in Brazil's tariffs increased

---

[1]For an extensive research of theoretical models on this subject see Panagariya (2000). The empirical works are summarized in De la Torre and Kelly (1982), Srinivasan et al. (1993), and Frankel (1997). Other theoretical and empirical works include Krueger (1999), Bhagwati (1993), Kowalczyk and Wonnacott (1992), Deltas et al. (2005), Goyal and Joshi (2006), Diao et al. (2003).

1

their spending on purchases of technology goods. Likewise, Trefler (2004) observes that the U. S. tariff concessions caused a boost in labor productivity of the Canadian firms in the most impacted, export-oriented group of industries. Similar patterns are shown by Bernard et al. (2006) for the U. S., by Topolova (2004) for India, by Aw et al. (2000) for Korea and Taiwan, by Alvarez and Lopez (2005) for Chile, and by Van Biesebroeck (2005) for sub-Saharan Africa.[2] Additionally, the positive effect of trade liberalization on productivity is substantiated by extensive theoretical work.[3]

The aim of this paper is to contribute to the literature on the impact of trade liberalization on firms' productivity by studying how this impact varies across two types of trade liberalization – multilateral versus regional. I consider a model in which firms can improve their productivity by investing in costly R&D. Here R&D is viewed broadly as any activity aimed at reducing the marginal cost of production.[4]

I study the mechanisms through which trade can affect the return to innovation within different types of trade systems. The two major mechanisms are the *scale* effect due to the increased size of the market and the *competition* effect due to the increased number of competitors in the markets. The focus of this paper is on the interaction of the two effects within multilateral and regional types of trade agreements.

I model trade agreements between countries with a network. Nodes represent countries and a link between the nodes indicates the existence of a trade agreement. In every country, there is a single firm producing one good. The good is sold domestically and in markets of the trade partner countries subject to oligopolistic competition. There is, therefore, an intra-industry trade between countries which are directly linked in the network.

The advantage of modelling trade agreements with a network is that it enables distinction between various types of trade systems. In particular, it allows for concentration on such differences between trade systems as the degree of countries' trade involvement (the number of trade agreements signed) and the nature of market interaction between countries (which countries trade and compete with each other, in which markets, the number of traders in each market, etc)who trades and competes with whom, on which

---

[2]The evidence on industry-level productivity improvements is presented in Baggs et al. (2002) for Canada, in Pavcnik (2002) for Chile, in Muendler (2004) for Brazil, in Bernard et al. (2006) for the U. S., and in Del Gatto et al. (2006) for Europe.

[3]The theoretical models identify several channels through which international trade affects productivity at the industry and/or at the firm level: the improved allocation of resources through specialization (Grossman and Helpman (1991)), the knowledge spillovers effect (Rivera-Batiz and Romer (1991), Devereux and Lapham (1994)), the reallocation of economic activities from less to more productive firms (Melitz (2003), Bernard et al. (2007), Yeaple (2005)), the exploitation of economies of scale (Krugman (1980)), the pro-competitive effect of trade openness (Aghion et al. (2005), Peretto (2003), Licandro and Navas-Ruiz (2007)), and others. See World Trade Report 2008 for the survey.

[4]Examples include developing new production technology, training of employees, internal reorganization of resources and factors of production.

markets, how many traders are present in each market, etc). Given the focus of this paper on the interaction between the scale and competition effects of trade liberalization within different types of trade systems, capturing exactly these differences is key.

I constrain the analysis to two specific classes of network structures associated with the multilateral and the regional scenarios of trade liberalization. The first class of networks are symmetric, or regular, networks. It incorporates the case of a *complete* network structure – a network where any one country is directly linked to every other country. The complete network in this model represents the multilateral trade agreement. The second class of networks are asymmetric networks with two types of nodes: high and low degree nodes. This class of networks captures the basic characteristics of the so-called *hub-and-spoke* trade system, where some countries (hubs) have relatively large number of direct trade partners as compared to other countries (spokes), which are mainly involved in trade agreements with hubs. According to a number of contributions on regional trade agreements, the hub-and-spoke trade arrangement has become a typical outcome of the regional trade liberalization.[5]

The modelling approach in this paper is closely related to the common approach in the strand of literature on R&D *co-operation* between firms in oligopoly. This strand of literature is well represented by the seminal papers, D'Aspremont and Jacquemin (1988) and Goyal and Moraga (2002). They consider a framework of Cournot competition, where at a pre-competitive stage firms can exert a cost-reducing effort. The typical element of this approach is that the rationale of co-operation between firms is the existence of R&D spillovers, which creates an externality. Co-operation is intended to internalize such an externality.

Similarly to models with R&D co-operation, in this model firms compete in a Cournot fashion choosing individual R&D efforts and production levels in a two-stage non-cooperative game. However, in contrast to D'Aspremont and Jacquemin (1988), Goyal and Moraga (2002) and the related literature, I abstract from R&D collaboration and spillovers. Instead, I concentrate purely on the effects of market access and competition faced by firms within various types of trade agreements on innovation intensity of the firms. Furthermore, unlike the assumption of standard oligopolistic competition between firms in one common market imposed in D'Aspremont and Jacquemin (1988) and Goyal and Moraga (2002), the central assumption in the present framework is that firms compete not only in one, but in several separate markets and every market is accessible only to those firms which have a trade agreement with that market. Clearly, this assumption

---

[5]The concept of hub-and-spoke trade arrangement was first introduced in Lipsey (1990) and Wonnacott (1990). It was further developed in Lipsey (1991), Wonnacott (1991, 1996), Kowalczyk and Wonnacott (1991, 1992), Baldwin (2003, 2005), De Benedictis et al. (2005), and others.

results in heterogeneity between firms in terms of their market size, a feature which is absent from the previous models.

The primary result of the paper is that the impact of trade liberalization on firms' R&D efforts depends crucially on the features of the trade agreements. Basically, the sizes of the scale effect and the competition effect, due to a new trade partner, vary across the multilateral and the regional trade systems since both effects are predetermined by the structure of the trade system. For example, with regard to the scale effect, gaining access to a new market in either a multilateral or bilateral context enhances incentives for firms to innovate. Yet, the "net worth", or the effective size, of the new market depends on the number of other firms present in the market and on the competitive power of these firms – their R&D and production levels. Those are both determined by the structure of the trade system. Similarly, with respect to the competition effect, a new trade partner of a firm, both in the multilateral and regional agreements becomes an additional rival of the firm in its domestic market. However, depending on the structure of the trade system, it may also become a rival in some, all or none of the firm's foreign markets. Furthermore, the size of the market share obtained by the new rival in firm's domestic and foreign markets depends on the number and competitive strength of other firms present in these markets, as well as the competitive strength of the rival himself. Those characteristics are defined by the structure of the trade system and by the market interactions of the firms with their own trade partners.

The difference in the scale and competition effects of trade liberalization and the resulting difference in the effective sizes of markets across the multilateral and the regional trade systems leads to the variation in *levels* of R&D efforts across systems. I show that for the *same number of direct trade partners*, the R&D effort of a hub in the regional trade system is higher than that of a country in the multilateral agreement. On the other hand, the R&D effort of a spoke is lower than that of a hub and lower than the R&D effort of a country in the multilateral agreement, even if a country in the multilateral agreement has the same number of direct trade partners as a spoke.[6] Additionally, I find that the *aggregate* R&D effort within the multilateral trade agreement exceeds that in the *star* – the simplest representative of the hub-and-spoke trade system.[7]

Some other findings of the paper concern the *change* in R&D investments by firms as the network of trade agreements expands. First, consistent with the empirical evidence discussed at the outset, I find that, in both the multilateral and regional trade systems,

---

[6]The same relative effects are found for the *welfare* and the *real income* values of countries in the stylized 3-country model by Deltas et al. (2005) and by Kowalczyk and Wonnacott (1992).

[7]Formally, the *star network* is a network in which there is a central country (hub) which is directly linked to every other country (spoke), while none of the other countries have a direct link with each other. The star in the present model is essentially a set of bilaterals of a hub with spokes, where each spoke has a trade agreement only with the hub.

an increase in the number of direct trade partners enhances innovation of a firm, at least as soon as the number of other firms present in the new trade partner countries is not "too large". Secondly, in the multilateral trade agreement, the *rate* of an increase in firm's R&D effort is declining in the size of the agreement.[8] I show that this result is implied by the decreasing market-enhancement effect of the new trade partners.

The paper is organized as follows. Section 1.2 presents the model and describes the two-stage game between firms. Sections 1.3 and 1.4 describe the solution of the second and of the first stage of the game, respectively. Section 1.5 discusses the scale and the competition effects of trade liberalization on firms' innovation decisions. The joint action of these two effects within the multilateral and the hub-and-spoke trade systems is studied in Scenario 1 and Scenario 2 of trade liberalization. The scenarios are compared in Section 1.6 and their policy implications are discussed in Section 1.7. Finally, Section 1.8 concludes.

1.2 The model

Network of regional trade agreements

Consider a setting with N countries where some countries are participants of one or more trade agreements (TAs) within a certain industry. I model trade agreements between countries with a network: countries are the nodes of the network and each link indicates the existence of a trade agreement between the two linked countries. If two countries have negotiated a TA, then each offers the other a privileged access to its domestic market: the tariffs and restrictions on trade are reduced. Otherwise, for simplicity I assume that tariffs and restrictions on trade between countries which did not sign a TA are trade-prohibitive. So in fact, trade may only exist between countries which have negotiated a TA, that is, only between countries which are directly linked in the network. For any $i \in 1 : N$, I will denote by N_i the set of countries with which country i has a trade link in the network of TAs. These are *direct trade partners* of i. Let $|N_i|$ be the cardinality of set N_i.[9] Also, let N_i^2 be the set of direct trade partners of direct trade partners of i, different from i. In other words, N_i^2 is the set of *two-links-away trade partners* of i in the network of TAs. Notice that some countries may simultaneously be direct and two-links-away trade partners of i. Let $|N_i^2|$ be the cardinality of set N_i^2

This model takes the network of trade agreements as exogenously given. Besides, since the trade agreement between any two countries are reciprocal, all links in the network are undirected and no multiple links exist.

---

[8]This finding is consistent with the hump-shaped relationship between competition and innovation derived by Aghion et al. (2005).

[9]This is the degree of i in the network.

Demand and cost structure

In every country, there is a single firm producing one good. The firm in country i can sell its good in the domestic market and in the markets of those countries with which i has a trade agreement. Let the output of firm i (from country i) produced for consumption in country j be denoted by y_{ij}. The total output of firm i is given by $y_i = \sum_{j \in N_i \cup \{i\}} y_{ij}$. Each firm i exporting its good to country $j \in N_i \cup \{i\}$ faces an inverse linear demand in country j given by:

$$p_j = a - b \left(y_{ij} + \sum_{k \in N_j \cup \{j\}, k \neq i} y_{kj} \right) \tag{1.1}$$

where $a, b > 0$ and $\sum_{k \in N_j \cup \{j\}} y_{kj} \leq a/b$.

Let τ denote the trade costs faced by *every* firm per unit of exports to *any* of its direct trade partners. These costs include tariffs on unit of export, transportation costs, etc.[10] The total trade costs faced by firm i are equal to:

$$t_i(\{y_{ij}\}_{j \in N_i}) = \tau \sum_{j \in N_i} y_{ij} \tag{1.2}$$

In addition, each firm can invest in R&D. The R&D effort of the firm helps lower its marginal cost of production. The cost of production of firm i is, therefore, a function of its production, y_i, and the amount of research, x_i, that it undertakes. I assume that the cost function of each firm is linear and is given by:

$$c_i(y_i, x_i) = (\alpha - x_i) y_i \tag{1.3}$$

where $0 \leq x_i \leq \alpha \; \forall i \in 1 : N$. In the following, I will also assume that a is sufficiently large as compared to α and the costs of trade between countries. Namely, let

Assumption 1 $\quad a > \alpha \left(1 + \max_{i \in 1:N} |N_i| \right) + 2\tau$

This assumption ensures that the demand for a good is high in all markets, so that in equilibrium, all firms produce strictly positive amounts of both the physical and the technological good. The R&D effort is costly: given the level $x_i \in [0, \alpha]$ of effort, the cost of effort of firm i is

$$z_i(x_i) = \delta x_i^2, \qquad \delta > 0 \tag{1.4}$$

---

[10]The analysis carries over in a setting where $\tau = 0$. The assumption of zero trade costs is standard in the literature on the formation of the network of TAs. See, for example, Furusawa and Konishi (2007), Goyal and Joshi (2006), and Mauleon et al. (2006).

Under this specification, the cost of the *R&D* effort is an increasing function and reflects the existence of diminishing returns to R&D expenditures.[11] The parameter δ measures the curvature of this function. In the following, it is assumed that δ is sufficiently large so that the second order conditions hold and equilibria can be characterized in terms of the first-order conditions and are interior.[12]

Two-stage game

Firms' strategies consist of the level of R&D activities and a subsequent production strategy based on their R&D choice. Both strategies are chosen via interaction in a two-stage non-cooperative game. In the first stage, each firm chooses a level of its R&D effort. The R&D effort of a firm determines its marginal cost of production. In the second stage, given these costs of production, firms operate in their domestic market and in the markets of their trade partners by choosing production quantities $\{y_{ij}\}_{i \in 1:N, j \in N_i \cup \{i\}}$ for every market. Each firm chooses the profit-maximizing quantity for each market separately, using the Cournot assumption that the other firms' outputs are given.

Notice the specific nature of interaction between firms in this game. First, firms compete with each other not in one but in several separate markets. Secondly, since countries trade only with those countries with which they have a trade agreement (a direct link in the network), a firm competes only with its direct and two-links-away trade partners. Furthermore, any direct trade partner of firm i competes with i in its own market and in the market of firm i, while any two-links-away trade partner of firm i, who is not simultaneously its direct trade partner, competes with i only in the market(s) of their common direct trade partner(s). This two-links-away radius of interaction between firms does not mean however that R&D and production choices of firms are not affected by other firms. As soon as the network of TAs is connected,[13] firms that are further than two links away from firm i affect R&D and production strategies of firm i indirectly, through the impact which they have on R&D and production choices of their own trade partners and trade partners of their partners, etc.

The game is solved for the subgame perfect Nash equilibrium, using backward induction. Each stage of the game is considered in turn.

---

[11]An important justification of this assumption is that "the technological possibilities linking R&D inputs and innovative outputs do not display any economies of scale with respect to the size of the firm in which R&D is undertaken" (Dasgupta (1986), p. 523).

[12]The assumptions imposed on the parameters of the demand and cost functions are standard in the models with linear-quadratic specification of the objective function (the firm's profit function in this case). This is admittedly a special setting. Yet, even in this simple case, the analysis of the interaction between markets and R&D efforts of firms in the network is quite complicated.

[13]The network is connected if there exists a path between any pair of nodes.

1.3 Solving the second stage

In the second stage, each firm $i \in 1 : N$ chooses a vector of its production plans $\{y_{ij}\}_{j \in N_i \cup \{i\}}$ so as to maximize its profit, conditional on R&D efforts $\{x_i\}_{i \in 1:N}$. The profit of firm i is

$$
\pi_i = \sum_{j \in N_i \cup \{i\}} \left(a - b y_{ij} - b \sum_{k \in N_j \cup \{j\}, k \neq i} y_{kj} \right) y_{ij} - (\alpha - x_i) y_i - \delta x_i^2 - \tau \sum_{j \in N_i} y_{ij} =
$$

$$
= \sum_{j \in N_i \cup \{i\}} \left(-b y_{ij}^2 - b \sum_{k \in N_j \{j\}, k \neq i} y_{kj} y_{ij} \right) + (a - \alpha + x_i) y_i - \delta x_i^2 - \tau \sum_{j \in N_i} y_{ij} \quad (1.5)
$$

Notice that function π_i is additively separable and quadratic in the output levels $\{y_{ij}\}_{j \in N_i \cup \{i\}}$ of firm i. This leads to linear first-order conditions and guarantees the existence and uniqueness of the solution of each firm's maximization problem.[14] Simple algebra results in the Nash-Cournot equilibrium production levels $\{y_{ij}\}_{i \in N, j \in N_i \cup \{i\}}$ of every firm i for consumption in country j:[15]

$$
y_{ii} = \frac{1}{b(|N_i| + 2)} \left(a - \alpha + (|N_i| + 1) x_i - \sum_{j \in N_i} x_j + |N_i| \tau \right) \quad (1.6)
$$

$$
y_{ij} = \frac{1}{b(|N_j| + 2)} \left(a - \alpha + (|N_j| + 1) x_i - \sum_{k \in N_j \cup \{j\}, k \neq i} x_k - 2\tau \right), \, j \in N_i \quad (1.7)
$$

So, the equilibrium output of firm i in country $j \in N_i \cup \{i\}$ is increasing in firm's own R&D effort and it is decreasing in R&D efforts of i's rivals in market j. That is, the higher the equilibrium R&D effort of i and the lower the equilibrium effort of every $k \in N_j \cup \{j\}$, $k \neq i$, the higher the share of market j gained by i.

Additionally, notice that the presence of non-negative trade costs τ gives any firm i the competitive advantage over its rivals on the domestic market and implies at least as high production of i for the domestic market as for the markets of its direct trade partners. Indeed, as soon as $x_i = x_j$ for some $j \in N_i$, the equilibrium level of production for

---

[14]Since $b > 0$, the second order conditions hold.

[15] Notice that since $x_k \leq \alpha$,

$$
\sum_{k \in N_j \cup \{j\}} y_{kj} = \frac{|N_j| + 1}{b(|N_j| + 2)} (a - \alpha) + \frac{1}{b(|N_j| + 2)} \sum_{k \in N_j \cup \{j\}} x_k - \frac{|N_j|}{b(|N_j| + 2)} \tau \leq \frac{|N_j| + 1}{b(|N_j| + 2)} a - \frac{|N_j|}{b(|N_j| + 2)} \tau < \frac{a}{b}
$$

In addition, since $0 \leq x_k \leq \alpha$ and Assumption 1 holds,

$$
y_{kj} \geq \frac{1}{b(|N_j| + 2)} (a - \alpha - |N_j| \alpha - 2\tau) = \frac{1}{b(|N_j| + 2)} (a - (\alpha + \alpha |N_j| + 2\tau)) > 0 \quad \forall k \in 1 : N \quad \forall j \in N_k \cup \{k\}
$$

market i of firm i is at least as high as that of firm j. Similarly, if for some $j \in N_i$ $|N_i| = |N_j|$ and $\sum_{k \in N_i, k \neq j} x_k = \sum_{k' \in N_j, k' \neq i} x_{k'}$, the equilibrium level of production of firm i for the domestic market is at least as high as its production for market j.

1.4 Solving the first stage

At the first stage, firms choose R&D efforts. Plugging expressions (1.6)– (1.7) for the output levels $\{y_{kj}\}_{k \in N_j \cup \{j\}}$ of Cournot competitors in country j into the profit function (1.5) of firm i, we obtain the function of the R&D effort levels $\{x_k\}_{k \in N_j \cup \{j\}}$. After some calculations, the profit of firm i can be written as:

$$
\begin{aligned}
\pi_i &= \left[\frac{1}{b} \sum_{j \in N_i \cup \{i\}} \frac{(|N_j| + 1)^2}{(|N_j| + 2)^2} - \delta \right] x_i^2 + \frac{2}{b} \left[(a - \alpha - 2\tau) \sum_{j \in N_i} \frac{|N_j| + 1}{(|N_j| + 2)^2} + \right. \\
&+ \left. (a - \alpha + |N_i|\tau) \frac{|N_i| + 1}{(|N_i| + 2)^2} \right] x_i - \frac{2}{b} \sum_{j \in N_i} \left[\frac{|N_i| + 1}{(|N_i| + 2)^2} + \frac{|N_j| + 1}{(|N_j| + 2)^2} \right] x_i x_j - \\
&- \frac{2}{b} \sum_{j \in N_i} \sum_{k \in N_j, k \neq i} \frac{|N_j| + 1}{(|N_j| + 2)^2} x_i x_k + f(\{x_k\}_{k \in N_i \cup N_i^2})
\end{aligned}
\tag{1.8}
$$

where $f(\{x_k\}_{k \in N_i \cup N_i^2})$ is a function of R&D efforts of i's competitors in different markets which does not distort i's equilibrium effort:

$$
\begin{aligned}
f(\{x_k\}_{k \in N_i \cup N_i^2}) &= \frac{1}{b} \sum_{j \in N_i} \frac{1}{(|N_j| + 2)^2} \left(a - \alpha - 2\tau - \sum_{k \in N_j \cup \{j\}, k \neq i} x_k \right)^2 + \\
&+ \frac{1}{b} \frac{1}{(|N_i| + 2)^2} \left(a - \alpha + |N_i|\tau - \sum_{j \in N_i} x_j \right)^2
\end{aligned}
$$

The profit function (1.8) of firm i is quadratic in its own R&D effort x_i. Besides, if δ is sufficiently high, so that the R&D cost function z_i is sufficiently steep, the profit function of firm i is concave in x_i. To be more precise, for a given network of trade agreements, as soon as

$$
\delta > \frac{1}{b} \max_{i \in N} \sum_{j \in N_i \cup \{i\}} \frac{(|N_j| + 1)^2}{(|N_j| + 2)^2}
\tag{1.9}
$$

the second order conditions hold and the profit maximizing R&D efforts of all firms can be found as a solution to the system of linear first-order conditions:

$$\left[-\frac{1}{b} \sum_{j \in N_i \cup \{i\}} \frac{(|N_j|+1)^2}{(|N_j|+2)^2} + \delta \right] x_i \; + \; \frac{1}{b} \sum_{j \in N_i} \left[\frac{|N_i|+1}{(|N_i|+2)^2} + \frac{|N_j|+1}{(|N_j|+2)^2} \right] x_j +$$

$$+\frac{1}{b} \sum_{j \in N_i} \sum_{k \in N_j, k \neq i} \frac{|N_j|+1}{(|N_j|+2)^2} x_k \; = \; \frac{1}{b}(a - \alpha - 2\tau) \sum_{j \in N_i} \frac{|N_j|+1}{(|N_j|+2)^2} +$$

$$+\frac{1}{b}(a - \alpha + |N_i|\tau) \frac{|N_i|+1}{(|N_i|+2)^2} \tag{1.10}$$

for all $i \in 1 : N$. In the matrix form, this system can be written as

$$\Sigma \cdot \mathbf{x} = \mathbf{u} \tag{1.11}$$

where $\mathbf{x} \in \mathbb{R}^N$ is a vector of unknowns, $\mathbf{u} \in \mathbb{R}^N$, and Σ is $(N \times N)$ square matrix. As soon as the network of trade agreements is connected, the matrix Σ is generically nonsingular and the right-hand side vector $\mathbf{u}$ is non-zero. Then (1.11) has a unique *generic* solution in $\mathbb{R}^N$, denoted by $\mathbf{x}^*$. Below it is shown that if δ satisfies an additional restriction, stronger than condition (1.9), this solution is ensured to be such that for all $i \in 1 : N$, $0 < x_i^* \leq \alpha$.

First, note that in the ith first-order condition in (1.10), the value of the expression on the right-hand side and the coefficients multiplying all x_k, $k \in N_i \cup N_i^2 \cup \{i\}$, are positive. Therefore, the value of x_i is larger the smaller the values of x_j and x_k for all $j \in N_i$ and $k \in N_i^2$. Hence, the sufficient condition for $x_i^* > 0$ is that (1.10) evaluated at $x_j = x_k = \alpha \; \forall j \in N_i$, $k \in N_i^2$ defines the value of x_i, which is greater than zero. This condition is provided by Assumption 1. Similarly, the sufficient condition for $x_i^* \leq \alpha$ is that (1.10) evaluated at $x_j = x_k = 0 \; \forall j \in N_i$, $k \in N_i^2$ defines x_i, which is smaller than or equal to α. This condition is equivalent to

Assumption 2 $\quad \delta \geq \frac{1}{\alpha b} \max_{i \in N} \left[\sum_{j \in N_i} \frac{|N_j|+1}{(|N_j|+2)^2} \left(\alpha|N_j| + a - 2\tau \right) + \frac{|N_i|+1}{(|N_i|+2)^2} (\alpha|N_i| + a + |N_i|\tau) \right]$

Under Assumption 1, the right-hand side of inequality in Assumption 2 is strictly larger than $\frac{1}{b} \max_{i \in N} \sum_{j \in N_i \cup \{i\}} \frac{(|N_j|+1)^2}{(|N_j|+2)^2}$ from the earlier restriction on δ in (1.9). Therefore, Assumption 2 is stronger. Together, Assumptions 1 and 2 guarantee that solution $\mathbf{x}^*$ of a system of the first-order conditions (1.10)(or (1.11)) is such that $0 < x_i^* \leq \alpha$ for all $i \in 1 : N$, and the second order conditions hold. Moreover, if the inequality in Assumption 2 is strict, solution $\mathbf{x}^*$ is interior.[16]

---

[16]Intuitively, when Assumption 1 holds, the demand for a good in each market is large, which stimulates R&D investment ($x_i^* > 0$). On the other hand, by Assumption 2, the cost of R&D is high, which confines the amount of R&D expenditures ($x_i^* \leq \alpha$).

The specification of the first-order conditions (1.10) suggests that an increase in R&D efforts of firm i's direct and/or two-links-away trade partners trigger a downward shift in firm i's response. Intuitively, by exerting higher R&D efforts, firm i's rivals capture larger shares of the markets and dampen the incentive of i to invest in R&D. We say that the efforts of firm i and its direct and two-links-away trade partners are *strategic substitutes* from i's perspective.

The first-order conditions (1.10) imply that in equilibrium the profit function of firm i is given by:

$$
\pi_i = \left[-\frac{1}{b} \sum_{j \in N_i \cup \{i\}} \frac{(|N_j|+1)^2}{(|N_j|+2)^2} + \delta \right] x_i^{*2} + \tag{1.12}
$$
$$
+ \frac{1}{b} \sum_{j \in N_i} \frac{1}{(|N_j|+2)^2} \left(a - \alpha - 2\tau - \sum_{k \in N_j \cup \{j\}, k \neq i} x_k^* \right)^2 +
$$
$$
+ \frac{1}{b} \frac{1}{(|N_i|+2)^2} \left(a - \alpha + |N_i|\tau - \sum_{j \in N_i} x_j^* \right)^2
$$

The short proof of this statement is provided in Appendix B. It is easy to see that due to Assumptions 1 and 2, equilibrium profit π_i of any firm i is strictly positive.

1.5 The impact of trade liberalization on equilibrium R&D efforts

In the framework of the present model, trade liberalization can be defined as an expansion of the network of trade agreements through an increase in the number of concluded trade agreements (links or links and nodes).

First, consider an impact of trade liberalization on equilibrium R&D efforts of firms in two countries which negotiate a trade agreement with each other. There are two major mechanisms at work. On the one hand, a new trade agreement creates an additional market for each firm (*scale effect*). This amplifies the return to productivity-enhancing investment, increasing the equilibrium R&D effort of each firm. On the other hand, the new agreement opens the markets of both countries to a new competitor (*competition effect*). This has two opposite effects on R&D. The enhanced competition dampens the return to R&D through a reduction in the domestic market share of each firm (market share effect of competition); yet, it also increases the return to R&D through a depreciation of markups, which expands the domestic market (markups effect of competition).

Thus, overall trade opening between two countries has an ambiguous effect on their equilibrium R&D efforts.

In addition, trade liberalization between any two countries affects R&D decisions of firms in other countries, too. For example, when i and j negotiate a trade agreement, R&D efforts of other direct trade partners of i and j are affected because firms in these countries face higher competition in i and j. Then, the impact on R&D efforts of the direct trade partners of i and j "spreads" to a larger network: direct trade partners of the direct trade partners of i and j face different R&D efforts and hence, competitive power of their trade partners, which has an impact on their own optimal R&D, etc.

Thus, scale and competition effects of trade liberalization (in any part of the network) can reinforce or dampen the incentive for firms to invest in R&D, the sign and strength of the impact being determined by the precise network structure. In the following section, this issue is addressed in case of multilateral and regional trade liberalization, where each type of liberalization is represented by the specific type of network.

1.6 Two scenarios of trade liberalization: Multilateralism versus regionalism

Scenario 1: Symmetric network of trade agreements. Multilateral trade system

Consider a class of symmetric networks of degree $n \geq 1$. A *symmetric*, or *regular*, network of degree n is a network where every node has the same number n of direct contacts. Given the aim of the paper, we are mainly interested in one representative of this class – a *complete* network. A complete network of degree greater than one represents a multilateral trade system, where all participant countries have a trade agreement with each other and neither country has a trade agreement with a third party. In this framework, an expansion of the multilateral trade system represents a scenario of multilateral trade liberalization. When multilateral trade liberalization involves all world economies, the resulting trade system is a "global free trade".

In addition, the class of symmetric networks comprises a case of one or several *simple bilaterals*, where every country signs a trade agreement with one and only one other country.[17] This case corresponds to the symmetric network of degree one.

---

[17]Up to the early 1990s, trade agreements were, with only a few exceptions, a set of non-intersecting bilateral or "small" multilateral trade agreements (the latter are also called plurilateral RTAs). The source of this evidence is for example, Lloyd and Maclaren (2004).
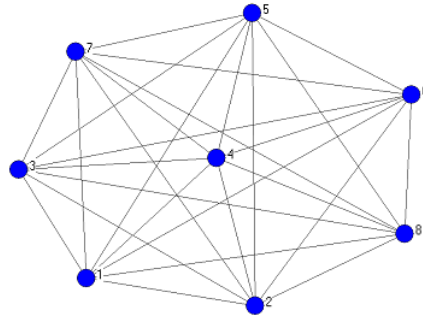
FIGURE 1.1: Complete network of degree 7 – multilateral trade agreement between 8 countries

In the symmetric network, all countries/firms are identical and hence, exert identical R&D efforts. Denote the level of this effort by x. Then the (single) first-order condition (1.10) can be written as follows:

$$\left[-\frac{(n+1)^3}{(n+2)^2} + \delta b \right] x + 2n\frac{n+1}{(n+2)^2}x + n\frac{(n+1)(n-1)}{(n+2)^2}x = \frac{(n+1)^2}{(n+2)^2}(a-\alpha) + \frac{n+1}{(n+2)^2}(-2\tau n + \tau n)$$

From this equation, the equilibrium effort x^* can be computed to be

$$x^* = \frac{a - \alpha - \frac{n}{n+1}\tau}{-1 + \delta b\left(1 + \frac{1}{n+1}\right)^2} \tag{1.13}$$

Now, using this expression for x^* and (1.12) for the equilibrium profit function, we derive the profit of any firm in the symmetric network:

$$
\begin{aligned}
\pi \;=\;& \left(-\frac{1}{b}\frac{(n+1)^3}{(n+2)^2} + \delta \right)\left(\frac{a - \alpha - \frac{n}{n+1}\tau}{-1 + \delta b(1 + \frac{1}{n+1})^2} \right)^2 + \\[2mm]
+\;& \frac{n}{b(n+2)^2}\left(a - \alpha - 2\tau - n\frac{a - \alpha - \frac{n}{n+1}\tau}{-1 + \delta b(1 + \frac{1}{n+1})^2} \right)^2 + \\[2mm]
+\;& \frac{1}{b(n+2)^2}\left(a - \alpha - n\left(\frac{a - \alpha - \frac{n}{n+1}\tau}{-1 + \delta b\left(1 + \frac{1}{n+1}\right)^2} - \tau \right) \right)^2
\end{aligned} \tag{1.14}
$$

The usual comparative statics analysis leads to the following result:

Proposition 1.1. *Suppose that Assumptions 1 and 2 hold for all $n < \bar{n}$, where $\bar{n} \geq 1$. Then for any $n < \bar{n}$, firm's equilibrium R&D effort x^* is monotonically increasing in n, while firm's profit π is monotonically decreasing in n.*

Proposition 1.1 is illustrated with Figure 1.2, where the equilibrium R&D effort and the profit of a firm in the symmetric network are drawn against the network degree n.[18]

---

[18]The simulation is done for the specific parameter values: $\alpha = 7$, $b = 1$, $\bar{n} = 10$, and $\tau = 2$; a and δ fulfill Assumptions 1 and 2.
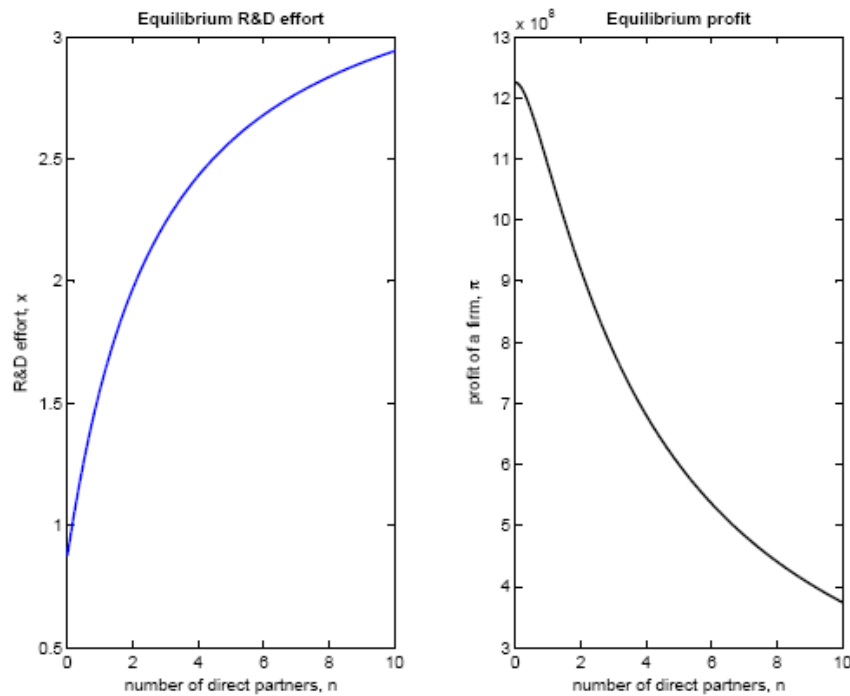
FIGURE 1.2: Equilibrium R&D effort and profit of a country in a symmetric network of degree n

Proposition 1.1 suggests that multilateral trade liberalization depreciates firms' profits. However, the incentive for firms to invest in R&D increase. The intuition for this result is easy to grasp. As a new country enters the multilateral trade agreement (or any other agreement which can be represented by the symmetric network), the reduction in the domestic and foreign market shares suffered by each firm is exactly compensated by participation in the entrant's market. That is, the negative market share effect of increased competition is exactly compensated by the positive scale effect associated with access to a new market. As a result, trade liberalization affects R&D only through the reduction in markups – the remaining component of the competition effect. Since the reduction in markups increases the aggregate market size of a firm, the optimal R&D of each firm in the multilateral agreement is increasing in the size of the agreement.

On the other hand, Figure 1.2 shows that the rates of increase in R&D and decrease in profits are both declining as the number of participant countries (size of the agreement) grows. This observation is implied by the fact that the markup-reducing effect of trade liberalization in the multilateral agreement is declining in the size of the agreement. Basically, using the definition of the demand function in (1.1), I find that the price of the good in each market is given by the decreasing and convex function of n:

$$p = \frac{\delta b(n+2)(a + \alpha(n+1) + \tau n) - a(n+1)^2}{-(n+1)^2 + \delta b(n+2)^2} \tag{1.15}$$
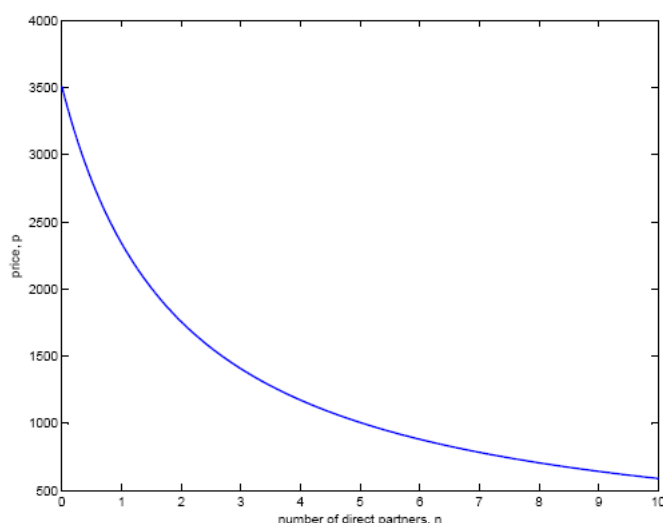
FIGURE 1.3: Price on the market of a country in a symmetric network of degree n

Furthermore, notice that Proposition 1.1 allows for the comparison of equilibrium R&D effort and an individual firm's profit in a multilateral agreement with those in a bilateral agreement or in autarky. Denote by x_a^* and π_a the equilibrium R&D effort and the profit of a firm in autarky, by x_b^* and π_b, the R&D effort and the profit of a firm in the bilateral agreement, and by $x^*(n)$ and $\pi(n)$, the R&D effort and the profit of a firm in the multilateral agreement of degree n (of size $n + 1$). Then we obtain

Corollary *For any $2 \le n < \bar{n}$, $x_a^* < x_b^* < x^*(n)$ and $\pi_a > \pi_b > \pi(n)$.*

The individual R&D investment of a firm is higher in the multilateral agreement than in the bilateral agreement or in autarky, while the profit of a firm in multilateral agreement is the lowest.

Finally, for aggregate levels of R&D, Proposition 1.1 and the Corollary imply that the aggregate level of R&D activities within the multilateral trade system is increasing in the size of the system and exceeds the aggregate R&D effort of the same number of countries where each country has negotiated one bilateral trade agreement.

Scenario 2: Asymmetric network of trade agreements. Hub-and-spoke trade system

I now examine the case of regional trade liberalization. In the process of regional trade liberalization, some countries (or groups of countries) negotiate one or several *bilateral* and/or *plurilateral* agreements with each other. Thus, in contrast to the multilateral type of liberalization considered in Scenario 1, each country may actually be a party to *several different* trade agreements where other countries do not necessarily have an agreement with each other. As a result, a complex trade system emerges where various regional

(preferential) agreements overlap. In the literature, this system is often described as a *hub-and-spoke* trade system, where some countries (hubs) have a relatively large number of direct trade partners as compared to other countries (spokes), which are mainly involved in trade agreements with hubs.

In this model, I approximate the hub-and-spoke structure by the asymmetric network with two types of nodes – nodes of high degree n (hubs) and of low degree m (spokes), $1 \leq m < n$. I assume that a *fixed positive* share of direct trade partners of hubs and spokes is represented by countries of the opposite type. For any hub, other hubs form a share $0 \leq \psi < 1$ of its direct trade partners while a share $0 < 1 - \psi \leq 1$ is represented by spokes. Similarly, for any spoke, other spokes form a share $0 \leq \varphi < 1$ of its direct trade partners and the remaining positive share $0 < 1 - \varphi \leq 1$ is represented by hubs.[19] The assumption of fixed (and identical across countries of the same type) shares ψ and φ significantly simplifies calculations and enriches the comparative statics analysis. For example, it facilitates the study of the effects of variation in the proportion of hubs to spokes among countries' direct trade partners while the number of these direct trade partners remains unchanged.[20]

Some examples of the hub-and-spoke trade system are demonstrated in Table 1.1.

In any given hub-and-spoke trade system, all hubs exert identical R&D effort (x_h) and likewise, all spokes exert identical R&D effort (x_s). Then the system (1.10) of the first-order conditions reduces to two equations:
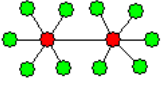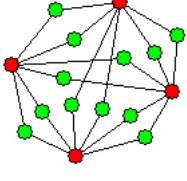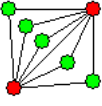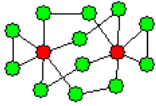
$$
\begin{aligned}
\Big(\; &- \; (1-\psi)n\frac{(m+1)^2}{(m+2)^2} - (n\psi+1)\frac{(n+1)^2}{(n+2)^2} + \delta b + \psi n\frac{2(n+1)}{(n+2)^2} + \\
&+ \; (1-\psi)n\frac{m+1}{(m+2)^2}((1-\varphi)m-1) + \psi n\frac{n+1}{(n+2)^2}(\psi n-1) \Big) \cdot x_h + \\
&+ \; (1-\psi)n\Big(\frac{n+1}{(n+2)^2} + \frac{m+1}{(m+2)^2} + \frac{n+1}{(n+2)^2}n\psi + \frac{m+1}{(m+2)^2}\varphi m \Big) \cdot x_s = \\
&= \; (a-\alpha-2\tau)\Big((1-\psi)n\frac{m+1}{(m+2)^2} + \psi n\frac{n+1}{(n+2)^2} \Big) + \frac{n+1}{(n+2)^2}(a-\alpha+n\tau)
\end{aligned}
\tag{1.16}
$$

$$
\begin{aligned}
\Big(\; &- \; (1-\varphi)m\frac{(n+1)^2}{(n+2)^2} - (\varphi m+1)\frac{(m+1)^2}{(m+2)^2} + \delta b + \varphi m\frac{2(m+1)}{(m+2)^2} + \\
&+ \; (1-\varphi)m\frac{n+1}{(n+2)^2}((1-\psi)n-1) + \varphi m\frac{m+1}{(m+2)^2}(\varphi m-1) \Big) \cdot x_s + \\
&+ \; (1-\varphi)m\Big(\frac{n+1}{(n+2)^2} + \frac{m+1}{(m+2)^2} + \frac{n+1}{(n+2)^2}n\psi + \frac{m+1}{(m+2)^2}\varphi m \Big) \cdot x_h = \\
&= \; (a-\alpha-2\tau)\Big((1-\varphi)m\frac{n+1}{(n+2)^2} + \varphi m\frac{m+1}{(m+2)^2} \Big) + \frac{m+1}{(m+2)^2}(a-\alpha+m\tau)
\end{aligned}
\tag{1.17}
$$

---

[19]Notice that in case when $\psi = 1$ ($\varphi = 1$), we obtain the complete network of degree n (m).

[20]See Proposition 1.2.

TABLE 1.1: Examples of hub-and-spoke trade system

Network characteristics	Network
Type 1: Single star (n bilaterals of a hub with spokes) $n > 1$, $m = 1$, $\psi = 0$, $\varphi = 0$	
Type 2: Stars with linked hubs $n > 1$, $m = 1$, $\psi > 0$, $\varphi = 0$	
Type 3: Stars sharing spokes $n > 1$, $m > 1$, $\psi = 0$, $\varphi = 0$	
Type 4: Stars with linked hubs, sharing spokes $n > 1$, $m > 1$, $\psi > 0$, $\varphi = 0$	
Type 5: Stars where some spokes are linked with ech other $n > 1$, $m > 1$, $\psi = 0$, $\varphi > 0$	

Remark: Red nodes stand for hubs, green nodes stand for spokes.

These equations uniquely identify equilibrium R&D efforts of a hub and a spoke. Extensive calculations lead to the closed-form solution (x_h^*, x_s^*), which has a cumbersome representation and therefore, is left for the Appendix.[21]

Consider the impact of trade liberalization in the hub-and-spoke trade system on equilibrium R&D efforts x_h^*, x_s^*. In contrast to the case with the multilateral trade system, in the asymmetric hub-and-spoke structure, the negative component of the competition effect of trade liberalization (market share effect of competition) is generally *not* compensated by the positive scale effect. Therefore, a priori, the impact of trade liberalization on R&D in the hub-and-spoke system is ambiguous. However, the comparative statics of x_h^* and x_s^* reveal some insights about the effects of an expansion/variation of the hub-and-spoke trade structure on the R&D effort of a firm.

Proposition 1.2. *Suppose that Assumptions 1 and 2 hold for all $m < n \leq \bar{n}$, where $\bar{n} > 1$. Then there exists $\Delta > 0$ such that for any $\delta \geq \Delta$ and for any $m < n < \bar{n}$, the following statements are fulfilled:*

---

[21]See the proof of Proposition 1.2 in Appendix B.

1. *the equilibrium R&D effort x_h^* of a hub is monotonically increasing in n and monotonically decreasing in m and in ψ;*

2. *the equilibrium R&D effort x_s^* of a spoke is monotonically decreasing in n and monotonically increasing in φ. Furthermore, x_s^* is monotonically increasing in m if at least one of the conditions holds:*

   (a) *the trade costs are sufficiently high: $\tau \geq \frac{1-\varphi}{3-2\varphi}(a - \alpha)$*

   (b) *the share of other spokes among direct trade partners is at least $1/3$: $\varphi \geq \frac{1}{3}$*

   (c) *the gap between n and m is relatively small: $n \leq m^2$, that is, $1 < \frac{n}{m} \leq m$*

Proposition 1.2 states that as soon as the specified parameter restrictions hold (including conditions (a) – (c)), the equilibrium R&D effort of a hub (spoke) is increasing in the number of its direct trade partners but is decreasing in the number of direct trade partners of spokes (hubs). In addition, for both a hub and a spoke, the higher the share of hubs among their direct trade partners, the lower the optimal R&D effort. Thus, the larger the number of directly accessible markets and the lower the number of competitors in these markets, the higher the incentive for firms to innovate. This observation suggests that, in the hub-and-spoke trade system, the competition effect of trade liberalization on firm's R&D is negative but the positive scale effect of any direct trade partner dominates its negative competition effect.

The negative impact of competition on equilibrium R&D decisions of spokes sheds light on conditions (a) – (c), which guarantee that an increase in m enhances spokes' R&D investments. Recall that the specification of a hub-and-spoke trade system in this model is such that an increase in the number of a spoke's direct trade partners m is associated with an increase in the number of *both* types of partners – hubs and spokes.[22] Since the market of a hub is "small" – smaller than the market of a spoke, an increase in the spoke's foreign market share may actually be smaller than a decrease in the share of the domestic market. As a result, the positive scale effect of an increase in m on R&D investment of a spoke may be dominated by the negative competition effect. Conditions (a) – (c) ensure that this would not be the case if: (a) the trade costs of firms are sufficiently high to restrict the amount of exports from new trade partners, (b) hubs represent only a minor share of direct trade partners of a spoke, or alternatively, (c) the number m of competitors in the spoke's market is comparable to n, so that the loss in the domestic market share of the spoke is not larger than the gain in the market of a new hub market.

---

[22]The proportion of spokes to hubs among the new trade partners is determined by φ: the lower φ, the higher the relative number of hubs.

The results of Proposition 1.2 are illustrated with Figures 1.5 and 1.6 in Appendix C.[23]

Comparison of multilateral and regional trade systems

In this section, I examine how the impacts of different types of trade liberalization compare in terms of R&D investments of firms. The comparison is made in two steps. First, I investigate the basic relationship between the R&D level of a firm in the multilateral trade agreement and a firm in the regional, hub-and-spoke trade system. After that I distinguish between different types of the hub-and-spoke system and study the *ranking* of R&D efforts of firms across various types of the hub-and-spoke system and the multilateral system. Consider each step in turn.

Step 1 To gain some insights about the sources of variation in R&D efforts of a firm across the multilateral and the regional, hub-and-spoke trade systems, let us first assume that the demand/price for the good is the same across markets, and that all firms operating in the market of a country obtain the same share of the market. Then, given the fixed number of direct trade partners, it is purely the *number* of other firms/competitors present in each trade partner country that determines the aggregate market size of a firm. The fewer competitors, the larger the aggregate market size of a firm and the higher the return to R&D investment.

In this simplified framework, for any number n of direct trade partners, the aggregate market size of a hub in *any* hub-and-spoke trade system is larger than that of a firm in the multilateral agreement. The opposite is true for spokes. For any number m of direct trade partners, the total market size of a spoke is smaller than that of a country in the multilateral agreement. To clarify the first statement, observe that while in the multilateral agreement (of degree n), the number of competitors of a firm is n in *each* of its n foreign markets, in the hub-and-spoke trade system, the number of competitors of a hub is n only in $\psi \cdot n$ of its foreign markets and it is less than n in the remaining markets. A similar argument is applicable for spokes.

Recall that these conclusions are derived under the assumption of equal demand and equal market shares of firms in every market. But in fact, they hold without this assumption, too. Formally, the result is an immediate implication of Proposition 1.2 and the short proof is provided in Appendix B:

Proposition 1.3. *For any $0 \leq \psi, \varphi < 1$ and for any $n, m > 1$ such that $n > m$,*

$$x_h^* > x^*(n) > x^*(m) > x_s^*$$

---

[23]Both figures are produced using the same parameter values as for Figure 1.2 in Scenario 1. In addition, for Figure 1.5, I set $\psi = \varphi = 0$ and for Figure 1.6, $n = 6$, $m = 2$.

Moreover, the same inequalities hold when a hub and a spoke belong not just to one, but to any different types of the hub-and-spoke structure.[24]

Step 2   Now, I compare equilibrium R&D efforts of firms across various types of the hub-and-spoke trade structure. To that end, I restrict attention to the specific types of the hub-and-spoke structure presented in Table 1.1. Notice that by Proposition 1.3, it only remains to compare separately R&D efforts of hubs and R&D efforts of spokes, since R&D of a hub is always higher than R&D of a spoke both in one and in different types of the hub-and-spoke structure.

As before, assume that the demand for the good is the same in all markets and that all firms (hubs and spokes) share each market equally. Consider the differences in market sizes of hubs and spokes across various hub-and-spoke structures. With regard to hubs, observe that for any number, n, of a hub's direct trade partners, a hub in the star (Type 1 system) enjoys the lowest competition in any of its foreign markets as compared to hubs in the other systems. Therefore, a hub in the star has the largest total market size. As a number of firms (competitors) in markets of a hub's direct trade partners grows, the aggregate market size of the hub decreases. This is the case when either the number of a spoke's direct trade partners, m, grows (Type 3 system), the share of hubs among direct trade partners, ψ, increases (Type 2 system) or when both changes in m and ψ happen simultaneously (Type 4 system). Furthermore, the larger the increase in m and/or ψ, the smaller the size of a hub's aggregate market.

For spokes the situation is symmetric. Given any number, n, of a hub's direct trade partners, a spoke in the star (Type 1 system) has access to a single foreign market ($m = 1$). Therefore, a spoke's market in the star is smaller than a spoke's market in any other hub-and-spoke trade system.[25] As the number of direct trade partners of a spoke, m, increases (Type 3 system), the market of a spoke expands. It expands even further if the share of spokes among direct trade partners, φ, grows (Type 5 system). Moreover, the larger the increase in m and/or φ, the larger the aggregate market size of a spoke.

As in Step 1, the insights gained on the assumption of equal demand and equal market shares of firms in every market prove to be valid when the assumption is relaxed. This leads to Proposition 1.4, which is formally derived in Appendix B. To state the proposition, I denote by x_{hi}^* the equilibrium R&D effort of a hub and by x_{si}^*, the equilibrium R&D effort of a spoke in the hub-and-spoke trade system of Type i, $i \in 1 : 5$.

---

[24]Recall from Scenario 1 that $x^*(n)$ denotes the equilibrium R&D effort of a firm in the multilateral agreement of degree n.

[25]In fact, on the assumption of equal demand and equal market shares of firms in every market, the market size of a spoke in Type 2 system is the same as in the star.

Proposition 1.4. *Consider Types 1–5 of the hub-and-spoke trade structure. Suppose that (i) n is the same across all types, (ii) m is the same across all types where $m > 1$ (Types 3, 4 and 5), and (iii) ψ is the same across all types where $\psi > 0$ (Types 2 and 4). Let $x^*(n)$ and $x^*(m)$ be defined for n and $m > 1$, identical to those in Types 1 – 5 of the hub-and-spoke structure. Then firms' equilibrium R&D efforts in Types 1–5 of the hub-and-spoke structure and in the multilateral agreement rank as follows:*

$$x_{h1}^* > x_{h3}^* > x_{h4}^* > x^*(n) > x^*(m) > x_{s5}^* > x_{s3}^* > x_{s1}^*$$

With respect to the equilibrium R&D effort x_{h2}^* of a hub and x_{s2}^* of a spoke in Type 2 system, the following inequalities hold:

$$x_{h1}^* > x_{h2}^* > x_{h4}^* \qquad and \qquad x_{s4}^* > x_{s2}^*$$

Proposition 1.4 is illustrated with Figure 1.4 and Figure 1.7.

Thus, the R&D efforts of firms in the multilateral system and in various types of the hub-and-spoke trade systems vary substantially. The highest R&D incentives exist for a hub, especially for a hub in the star (Type 1 system), whereas for a spoke in the star the incentives are the lowest. As the number of direct trade partners of a spoke and/or the share of spokes (hubs) among direct trade partners of each spoke (hub) increase, the levels of R&D investment of hubs and spokes converge. They coincide at the level of R&D investment of a firm in the multilateral agreement, which therefore, takes an average position: it is lower than R&D of a hub but higher than R&D of a spoke.

Further to comparing individual R&D investments by firms, I compare the aggregate levels of R&D activities of the same total number of countries in the multilateral trade agreement and in the star – bilateral agreements of one country (a hub) with the others (spokes). I find that although the individual R&D effort of a hub in the star is much higher than the R&D effort of a single country in the multilateral agreement, the aggregate R&D in the star is *lower* than in the multilateral agreement. This observation is demonstrated by Figure 1.8 in Appendix C.
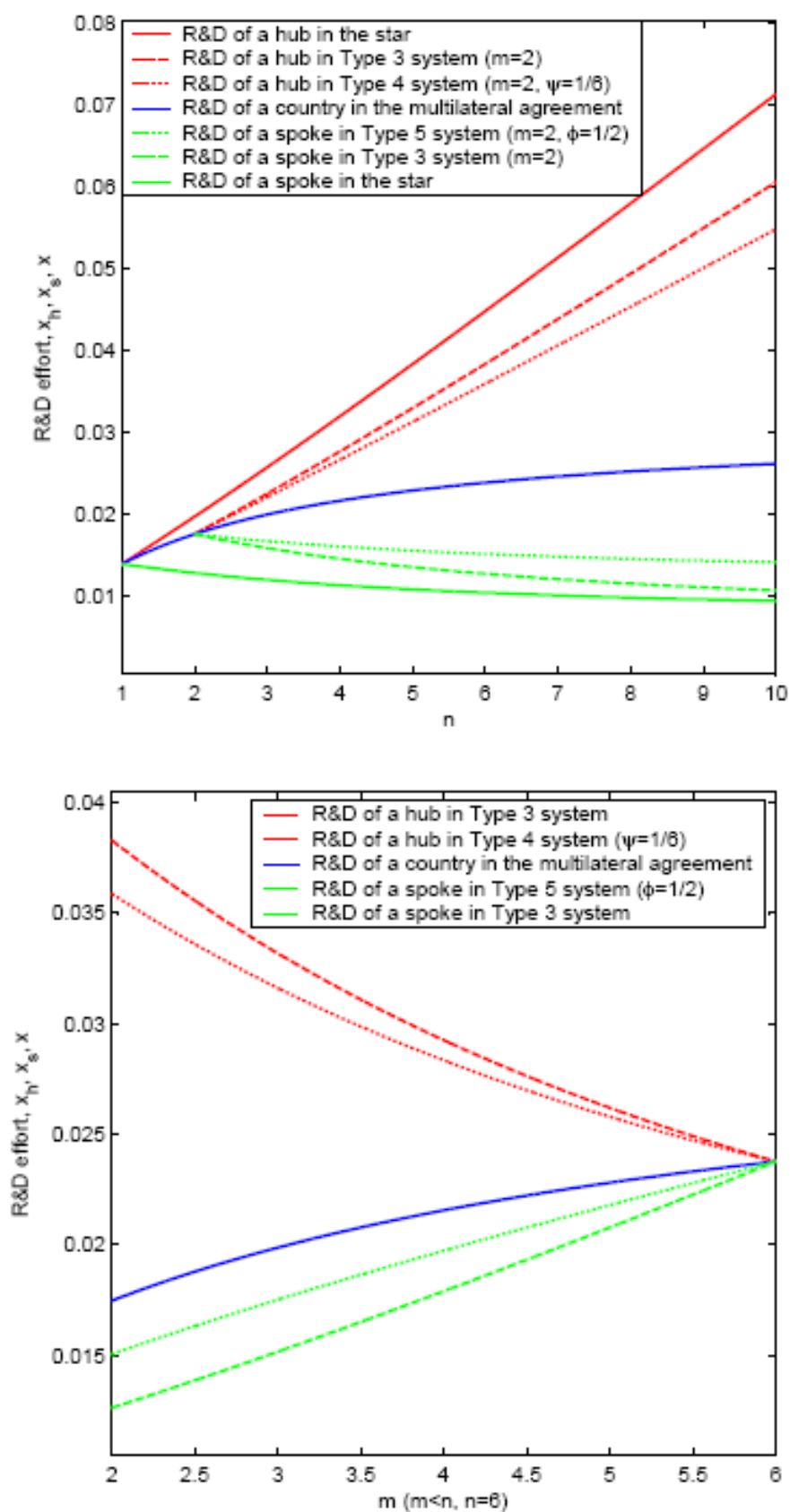
FIGURE 1.4: Equilibrium R&D efforts in the multilateral and in the hub-and-spoke trade systems as a function of n (the upper sub-figure) and as a function of m (the lower sub-figure).

1.7 Policy implications

The previous analysis suggests that the structure of the network of trade agreements and the position of a country in this network are key for understanding the differences in R&D and productivity levels of firms across the multilateral and the regional types of trade systems. This feeds into the ongoing debate on gains and losses of multilateralism versus regionalism, especially with respect to the intensive proliferation of regional trade agreements among the WTO member countries. The paper suggests that productivity gains of regionalism versus those of multilateralism depend heavily on the relative number of regional trade agreements signed by countries. If a country signs relatively large number of trade agreements within the regional system (core country), then its R&D and productivity are higher than R&D and productivity of a country in the multilateral system. At the same time, a country that signs relatively small number of trade agreements within the regional system (periphery country) has lower productivity gains than a country in the multilateral system. This observation suggests that from the perspective of a small economy, which normally becomes a spoke/periphery in the regional trade system, the prospects for productivity improvements within the multilateral trade system are generally better than within the regional system.[26]

In addition, the finding of the unambiguously positive impact on R&D of the multilteral trade liberalization indicates that the expansion of the WTO as well as the consolidation of several plurilateral blocks or their accession to the WTO enhance R&D in every country.[27]

Recall that for the specific classes of networks considered in Scenario 1 and in Scenario 2, the equilibrium R&D effort of a firm is increasing in the number of its direct trade partners (scale and competition effects of trade liberalization) but decreasing in the number of its two-links-away trade partners (competition effect only). In Appendix A, I further investigate the issue of the impact which direct and two-links-away trade partners have on R&D of a firm in a *generic* network under the assumption of *small external effects*. Consistent with the results of the previous analysis, I find that new direct trade partners of a firm (mostly) lead to an increase in the firm's R&D investment and the smaller the number of competitors in the new markets, the larger the increase.

---

[26]The finding of a substantially lower R&D and productivity levels in a spoke economy as compared to those in a hub and in a country within the multilateral system, supports the argument of earlier studies about the disadvantageous position of spokes. For instance, Baldwin (2003), Kowalczyk and Wonnacott (1992), Deltas et al. (2005), Lloyd and Maclaren (2004), and De Benedictis et al. (2005) find that welfare and income levels are lower for spokes than for hubs and than for countries in the complete network.

[27]According to Fiorentino et al. (2007), the number of merging regional trade agreements is currently increasing. Examples include EC-GCC, SACU-MERCOSUR, among others.

1.8 Conclusion

This paper develops a model of international trade with firm-level productivity improvements via R&D. Firms in different countries sell their product and compete in a Cournot fashion with other firms in the domestic market and in the markets of their trade partners. The trade partners of any country/firm are defined by the network of trade agreements: countries which are linked in the network are (direct) trade partners of each other.

I focus on two specific types of networks: the complete and the hub-and-spoke network. In the model, these networks represent trade arrangements which arise as a result of multilateral or regional trade liberalization, respectively. I study how the structure of the trade arrangement and the position of a country in this structure affect R&D investments by firms. In this manner, I address the issue of the difference in the impact of multilateral and regional types of trade liberalization on firms' R&D and productivity.

I show that the R&D response of firms to trade liberalization is the net outcome of two different effects: one, stimulating R&D through the creation of new markets (scale effect), and the other, deterring or improving R&D through the emergence of new competitors (competition effect). I find that the sizes of both effects vary across the multilateral and the regional/hub-and-spoke trade systems. Basically, a new country entering the multilateral or the hub-and-spoke trade system represents different "value added" for firms in every system since the effective size of the new market and the competitive impact of the new firm depend on the structure of the trade system.

The variations in the scale effect and the competition effect of trade across structures lead to variations in firms' aggregate market sizes. In turn, the difference in firms' market sizes explains the difference in *levels* of firms' R&D investments. For the *same number of direct trade partners*, the R&D effort of a hub in the hub-and-spoke trade system is higher than the R&D effort of a country in the multilateral agreement. On the other hand, R&D of a spoke is lower than R&D of a hub and lower than R&D of a country in the multilateral agreement, even if a country in the multilateral agreement has the same number of direct trade partners as a spoke.

In addition, consistent with the empirical evidence, I find that a new market increases R&D of a firm in the multilateral system and R&D of a hub in the regional system. It also increases R&D of a spoke, at least as soon as the level of competition in the new

spoke's trade partner is not too high.[28] Furthermore, for the aggregate levels of R&D activities, I find that the aggregate R&D effort within the multilateral trade agreement exceeds that in the *star* – the simplest representative of the hub-and-spoke trade system.

The paper suggests some policy implications. For example, with regard to benefits and losses of regionalism versus multilateralism, the paper indicates that the regional trade liberalization is likely to be more beneficial than the multilateral trade liberalization for R&D and productivity level of hubs, countries which sign sufficiently many regional trade agreements. At the same time, R&D and productivity level of spokes are higher if the countries choose the multilateralist alternative.

Lastly, it is important to emphasize that the results of this paper are driven purely by the scale effect and the competition effect of trade liberalization, which are in turn determined by the structural characteristics of trade agreements. In order to precisely isolate the impact of the structure of trade agreements, the model abstracts from other channels through which trade liberalization may affect R&D. For example, the R&D spillover effect and the firms selection effect of trade are excluded.[29] Additionally, for the sake of simplicity, the model disregards important inequalities in terms of geographical size and initial income/resources across countries. These limitations of the paper suggest directions for further research. In particular, empirical testing of the model could provide more insights.

---

[28]Recall that in the regional hub-and-spoke trade system, R&D of a spoke *declines* in response to a new market opening if the negative competition effect on R&D of the spoke's new trade partners outweighs the positive scale effect. However, by simulating the model for the star network under various parameter assumptions, I find that initiating trade with the hub decreases R&D of a spoke when the number of competitors in the hub's market is "unrealistically high" (more than 100).

[29]As suggested by Melitz (2003), the firms selection effect results from the initial difference in productivity of *multiple* domestic firms in each country and from the existence of the fixed costs of production and exporting.

1.9   Appendix

Appendix A: Equilibrium R&D efforts in arbitrary network. The case of small external effects

Consider the system of first-order optimality conditions (1.10). Below I study the properties of the solution to this system when the magnitude of local effects – effects of interaction between firms in the network – is arbitrarily small. I seek the ranking of optimal R&D decisions of firms in accordance with simple characteristics of firms' positions in the network, such as the nodal degrees and the sum of neighbors' degrees. To derive this ranking, I employ the asymptotic approach suggested by Bloch and Quérou (2008).[30]

Notice that the system of linear first-order conditions (1.10) can be written as:

$$\delta x_i \quad - \quad \frac{1}{b}\left[\sum_{j \in N_i \cup \{i\}} \frac{(|N_j|+1)^2}{(|N_j|+2)^2}x_i - \sum_{j \in N_i}\left(\frac{|N_i|+1}{(|N_i|+2)^2} + \frac{|N_j|+1}{(|N_j|+2)^2}\right)x_j - \right.$$

$$\left. - \sum_{j \in N_i}\sum_{k \in N_j, k \neq i} \frac{|N_j|+1}{(|N_j|+2)^2}x_k\right] =$$

$$= \quad \frac{1}{b}(a - \alpha - 2\tau)\sum_{j \in N_i}\frac{|N_j|+1}{(|N_j|+2)^2} + \frac{1}{b}(a - \alpha + |N_i|\tau)\frac{|N_i|+1}{(|N_i|+2)^2}, \qquad i \in 1:N$$

In the matrix form this has a simple representation:

$$\left(\delta \mathbf{I} - \frac{1}{b}\mathbf{B}\right) \cdot \mathbf{x} = \frac{1}{b}\tilde{\mathbf{u}}$$

Alternatively:

$$(\mathbf{I} - \lambda\mathbf{B}) \cdot \mathbf{x} = \lambda\tilde{\mathbf{u}} \tag{1.18}$$

where $\lambda = \frac{1}{b\delta}$. In this system, matrix $\lambda\mathbf{B}$ is the matrix of local effects. Below, I investigate the solution to (1.18) when the norm of matrix $\lambda\mathbf{B}$ capturing the magnitude of local effects is small.

First, following Bloch and Quérou (2008), I define a vector sequence $\mathbf{f} = (\mathbf{c}^1, \mathbf{c}^2, \ldots, \mathbf{c}^m, \ldots)$, where each vector $\mathbf{c}^m$ is given by:

$$\mathbf{c}^m = \lambda^m \tilde{\mathbf{u}}\mathbf{B}^{m-1}$$

---

[30]As emphasized in Bloch and Quérou (2008), at least two arguments can defend the usefulness of studying network effects whose magnitude is small. First, when the matrix of interactions is complex, this may be the only way to evaluate the equilibrium R&D decisions for an *arbitrary* network structure. Secondly, by continuity, the insights obtained for small external effects continue to hold as the magnitude of externalities increases.

The first terms of this sequence are

$$
\begin{aligned}
\mathbf{c}^1 &= \lambda\tilde{\mathbf{u}}, \\
\mathbf{c}^2 &= \lambda^2\tilde{\mathbf{u}}\mathbf{B}, \\
\mathbf{c}^3 &= \lambda^3\tilde{\mathbf{u}}\mathbf{B}^2
\end{aligned}
$$

Using the sequence $\mathbf{f}$, I can now state the approximation result, which provides an equivalence between the ranking of the components of the solution $\mathbf{x}^*$ to (1.18) and the lexicographic ordering of the components of $\mathbf{f}$ when the magnitude of $\|\lambda\mathbf{B}\|$ is close to zero.

Proposition 1.5. *Consider a system of linear equations* (1.18). *Suppose that $\|\lambda B\|$ is sufficiently small:*[31] *for a given $0 < \bar{\varepsilon} < 1$, $\|\lambda B\| \leq \frac{\bar{\varepsilon}}{N}$. Then there exists a unique solution $\mathbf{x}^*$ and $K > 1$ such that for any $i, j \in 1 : N$, $i \neq j$,*

$$
|x_i^* - x_j^* - (c_i^M - c_j^M)| \leq \lambda \cdot \frac{\bar{\varepsilon}^{K+1}}{1 - \bar{\varepsilon}} \cdot 2\|\tilde{\mathbf{u}}\|
$$

where $(c_M)_i$ and $(c_M)_j$ are the first unequal elements of the sequences $\mathbf{f}_i = (\mathbf{c}_i^1, \mathbf{c}_i^2, \ldots, \mathbf{c}_i^m, \ldots)$ and $\mathbf{f}_j = (\mathbf{c}_j^1, \mathbf{c}_j^2, \ldots, \mathbf{c}_j^m, \ldots)$: $\mathbf{c}_i^M \neq \mathbf{c}_j^M$ and $\mathbf{c}_i^m = \mathbf{c}_j^m$ for all $m < M$.

Thus, if the upper bound for the magnitude of local effects is close to zero,

$$
x_i^* > x_j^* \Leftrightarrow \mathbf{f}_i \succ \mathbf{f}_j
$$

where $\mathbf{f}_i \succ \mathbf{f}_j$ stands for lexicographic dominance of $\mathbf{f}_i$ over $\mathbf{f}_j$. This means that in order to compare equilibrium R&D efforts of different firms, one can restrict attention to the first order term $\mathbf{c}^1$, or if the first order terms are equal, to the second order term $\mathbf{c}^2$, etc. As a result, the ranking of optimal R&D choices of firms reduces to the ranking of characteristics of firms' positions in the network.

Consider a pair of firms (i, i'), $i, i' \in 1 : N$, such that $\tilde{\mathbf{u}}_i \neq \tilde{\mathbf{u}}_{i'}$. Then by Proposition 1.5, if

$$
\|\lambda B\| \leq \frac{\bar{\varepsilon}}{N}
$$

for some $0 < \bar{\varepsilon} < 1$, then the difference between x_i^* and $x_{i'}^*$ can be approximated by the difference between $\tilde{\mathbf{u}}_i$ and $\tilde{\mathbf{u}}_{i'}$ such that the measurement error does not exceed $\lambda \cdot \frac{\bar{\varepsilon}^{K+1}}{1-\bar{\varepsilon}} \cdot 2\|\tilde{\mathbf{u}}\|$, where

$$
\|\tilde{\mathbf{u}}\| = \max_i (a - \alpha - 2\tau) \sum_{j \in N_i} \frac{|N_j| + 1}{(|N_j| + 2)^2} + (a - \alpha + |N_i|\tau)\frac{|N_i| + 1}{(|N_i| + 2)^2}
$$

---

[31]As in Bloch and Quérou (2008), I use the l_∞ vector norm defined by $\|\mathbf{A}\| = \max_{i,j}|a_{ij}|$

So, when the local effects are small, the R&D effort chosen by firm i is at least as high as the R&D effort of firm i' if and only if

$$(a - \alpha - 2\tau) \sum_{j \in N_i} \frac{|N_j| + 1}{(|N_j| + 2)^2} + (a - \alpha + |N_i|\tau) \frac{|N_i| + 1}{(|N_i| + 2)^2} \geq \qquad (1.19)$$

$$\geq (a - \alpha - 2\tau) \sum_{j \in N_{i'}} \frac{|N_j| + 1}{(|N_j| + 2)^2} + (a - \alpha + |N_{i'}|\tau) \frac{|N_{i'}| + 1}{(|N_{i'}| + 2)^2}$$

The inequality (1.19) suggests that at the first order the R&D effort of firm i is decreasing in the number $|N_j|$, $j \in N_i$, of i's two-links-away trade partners. In addition, the R&D effort of firm i is increasing in the number $|N_i|$ of i's direct trade partners as soon as the new trade partner j' is such that

$$\frac{|N_i| + 2}{(|N_i| + 3)^2}(a - \alpha + (|N_i| + 1)\tau) - \frac{|N_i| + 1}{(|N_i| + 2)^2}(a - \alpha + |N_i|\tau) + (a - \alpha - 2\tau)\frac{|N_{j'}| + 1}{(|N_{j'}| + 2)^2} > 0$$

Alternatively, this can be written as:

$$(a - \alpha - 2\tau)\frac{|N_{j'}| + 1}{(|N_{j'}| + 2)^2} > \frac{|N_i|^2 + 3|N_i| + 1}{(|N_i| + 2)^2(|N_i| + 3)^2}(a - \alpha + |N_i|\tau) - \frac{|N_i| + 2}{(|N_i| + 3)^2}\tau \quad (1.20)$$

It is easy to see that under Assumption 1 the left-hand side of inequality (1.20) is decreasing in $|N_{j'}|$. This means that the additional direct trade partner j' of i increases i's incentives to innovate as soon as the number of competitors $|N_{j'}|$ of i in market j' is sufficiently low. Thus, in accordance with the earlier discussion in the paper, opening trade with an additional trade partner increases the R&D investment of firm i if the actual market size of the new trade partner is large enough.

The finding of a positive effect of direct trade partners and a negative effect of two-links-away trade partners of i on i's equilibrium R&D effort, together with the conditions which guarantee these effects, are consistent with the findings of Scenarios 1 and 2 discussed in Section 1.6.

Appendix B: Proofs

Derivation of the profit function in (1.12)

The profit function in (1.8) can be written as:

$$
\begin{aligned}
\pi_i &= 2x_i^* \left[\frac{1}{b}(a - \alpha - 2\tau) \sum_{j \in N_i} \frac{|N_j| + 1}{(|N_j| + 2)^2} + \frac{1}{b}(a - \alpha + |N_i|\tau) \frac{|N_i| + 1}{(|N_i| + 2)^2} - \right. \\
&\quad - \frac{1}{b} \sum_{j \in N_i} \left[\frac{|N_i| + 1}{(|N_i| + 2)^2} + \frac{|N_j| + 1}{(|N_j| + 2)^2} \right] x_j^* - \frac{1}{b} \sum_{j \in N_i} \sum_{k \in N_j, k \neq i} \frac{|N_j| + 1}{(|N_j| + 2)^2} x_k^* \left. \right] - \\
&\quad - \left[-\frac{1}{b} \sum_{j \in N_i \cup \{i\}} \frac{(|N_j| + 1)^2}{(|N_j| + 2)^2} + \delta \right] x_i^{*2} + f(\{x_k\}_{k \in N_i \cup N_i^2})
\end{aligned}
$$

By the first-order conditions (1.10), this reduces to:

$$
\begin{aligned}
\pi_i &= 2 \left[-\frac{1}{b} \sum_{j \in N_i \cup \{i\}} \frac{(|N_j| + 1)^2}{(|N_j| + 2)^2} + \delta \right] x_i^{*2} - \left[-\frac{1}{b} \sum_{j \in N_i \cup \{i\}} \frac{(|N_j| + 1)^2}{(|N_j| + 2)^2} + \delta \right] x_i^{*2} + \\
&\quad + f(\{x_k\}_{k \in N_i \cup N_i^2}) = \left[-\frac{1}{b} \sum_{j \in N_i \cup \{i\}} \frac{(|N_j| + 1)^2}{(|N_j| + 2)^2} + \delta \right] x_i^{*2} + \\
&\quad + \frac{1}{b} \sum_{j \in N_i} \frac{1}{(|N_j| + 2)^2} \left(a - \alpha - 2\tau - \sum_{k \in N_j \cup \{j\}, k \neq i} x_k^* \right)^2 + \\
&\quad + \frac{1}{b} \frac{1}{(|N_i| + 2)^2} \left(a - \alpha + |N_i|\tau - \sum_{j \in N_i} x_j^* \right)^2
\end{aligned}
$$

Proof of Proposition 1.1

First, notice that in case of a symmetric network of degree n, the right-hand side of inequality in Assumption 1 is an increasing function of n and also the right-hand side of inequality in Assumption 2 is an increasing function of n, *provided* that Assumption 1 holds. Therefore, for Assumptions 1 and 2 to be fulfilled for all $n < \bar{n}$, it is enough to ensure that these assumptions hold for $n = \bar{n}$. The resulting restrictions are

$$
a > \alpha(1 + \bar{n}) + 2\tau, \quad \text{and} \tag{1.21}
$$

$$
\delta \geq \frac{1}{\alpha b} \frac{(\bar{n} + 1)}{(\bar{n} + 2)^2} ((\alpha \bar{n} + a)(\bar{n} + 1) - \tau \bar{n}) \tag{1.22}
$$

The proof of Proposition 1.1 is established in two steps.

1. R&D effort x^* is monotonically increasing in n

   Taking a derivative of x^* in (1.13) with respect to n, we obtain:

   $$\frac{\partial x^*}{\partial n} = \frac{-\frac{\tau}{(n+1)^2}\left(-1 + \delta b\left(1 + \frac{1}{n+1}\right)^2\right) + 2\delta b\left(1 + \frac{1}{n+1}\right)\frac{1}{(n+1)^2}\left(a - \alpha - \frac{n}{n+1}\tau\right)}{\left(-1 + \delta b\left(1 + \frac{1}{n+1}\right)^2\right)^2} =$$
   $$= \frac{\frac{1}{(n+1)^2}\left(\tau + \delta b\left(1 + \frac{1}{n+1}\right)\left(-\tau\left(1 + \frac{1}{n+1}\right) + 2\left(a - \alpha - \frac{n}{n+1}\tau\right)\right)\right)}{\left(-1 + \delta b(1 + \frac{1}{n+1})^2\right)^2}$$

   The sign of this derivative is positive as soon as

   $$2\left(a - \alpha - \frac{n}{n+1}\tau\right) > \tau\left(1 + \frac{1}{n+1}\right)$$

   One can readily see that this inequality holds due to the restriction on a in (1.21).

2. Profit, π, is monotonically decreasing in n

   Due to the computational complexity, I present only a schematic proof of this statement.

   Taking the derivative of π in (1.14) with respect to n, we obtain the expression represented by the product of the ratio $\frac{1}{b(2n-4b\delta+n^2-4bn\delta-bn^2\delta+1)^3}$ and the quadratic polynomial of τ. The ratio is negative for any $n \geq 1$ due to the restriction on δ in (1.22). On the other hand, the value of the polynomial is positive for any $n \geq 1$ as soon as parameters satisfy the restrictions (1.21) and (1.22). The latter is established via two steps.

   - First, I find that due to the restriction (1.22) the coefficient of the polynomial at the quadratic term τ^2 is negative for any given $n \geq 1$. Besides, the constant term is positive. Hence, the graph of the quadratic function is a parabola with downward-directed branches and two real roots – one positive and one negative.

   - Since the unit trade cost τ is positive and by the restriction (1.21), it does not exceed $\frac{1}{2}(a - \alpha)$, to establish that the value of the polynomial is positive for all $\tau \in (0, \frac{1}{2}(a-\alpha))$, it suffices to show that the value of the polynomial is positive at $\tau := \frac{1}{2}(a - \alpha)$. One can find that this is indeed the case, provided that (1.21) and (1.22) hold.

   Thus, for all $n \geq 1$ and any parameter values satisfying the conditions (1.21) and (1.22), the derivative of π with respect to n is negative, so that the profit function is decreasing in n.

   ∎

Sketch of the proof of Proposition 1.2

Notice that to ensure that Assumptions 1 and 2 hold for all $m < n \leq \bar{n}$, it is enough to impose the restrictions:

$$a > \alpha(1 + \bar{n}) + 2\tau \quad \text{and} \tag{1.23}$$

$$\delta \geq \frac{1}{\alpha b}\left[\frac{\bar{n} + 1}{(\bar{n} + 2)^2}\bar{n}(\alpha\bar{n} + a - 2\tau) + \frac{2}{9}((\alpha\bar{n} + a)(\bar{n} + 1) - \tau\bar{n})\right] \tag{1.24}$$

The equilibrium R&D effort of a hub and a spoke are given by the solution to the system of equations (1.16) − (1.17):

$$x_s^* = \frac{\begin{pmatrix}(a - \alpha - 2\tau)(1 - \varphi)m(n + 1)(m + 2)^2 + \\ + (m + 1)(n + 2)^2\big[(a - \alpha - 2\tau)\varphi m + a - \alpha + m\tau\big]\end{pmatrix}\cdot \\ \cdot\begin{pmatrix}(n + 2)^2\big[b\delta(m + 2)^2 - n(1 - \psi)(m + 1)(2 + \varphi m)\big] - \\ - (n\psi + 1)(n + 1)(n(1 - \psi) + 1)(m + 2)^2\end{pmatrix} - \\ - (1 - \varphi)m\begin{pmatrix}(n\psi + 1)(n + 1)(m + 2)^2 + (\varphi m + 1)(m + 1)(n + 2)^2\end{pmatrix}\cdot \\ \cdot\begin{pmatrix}(a - \alpha - 2\tau)\big[n\psi(n + 1)(m + 2)^2 + (n - n\psi)(m + 1)(n + 2)^2\big] + \\ + (n + 1)(m + 2)^2(a - \alpha + n\tau)\end{pmatrix}}{\begin{pmatrix}(n + 2)^2\big[b\delta(m + 2)^2 - n(1 - \psi)(m + 1)(2 + \varphi m)\big] - \\ - (n\psi + 1)(n + 1)(n(1 - \psi) + 1)(m + 2)^2\end{pmatrix}\cdot \\ \cdot\begin{pmatrix}-(1 - \varphi)m(m + 2)^2(n\psi + 2)(n + 1) + (n + 2)^2\delta b(m + 2)^2 - \\ - (n + 2)^2\big[(\varphi m + 1)(m + 1)^2 - m\varphi(m + 1)(1 + \varphi m)\big]\end{pmatrix} - \\ - (1 - \varphi)(1 - \psi)mn\begin{pmatrix}(n + 1)(m + 2)^2(1 + n\psi) + (m + 1)(n + 2)^2(1 + \varphi m)\end{pmatrix}^2}$$

$$x_h^* = \frac{\begin{aligned}&(a - \alpha - 2\tau)\big(n\psi(n + 1)(m + 2)^2 + (n - n\psi)(m + 1)(n + 2)^2\big) + \\ &+ (n + 1)(m + 2)^2(a - \alpha + n\tau) - x_s^*(n - n\psi)\big((m + 1)(n + 2)^2 + \\ &+ (n + 1)(m + 2)^2 + \varphi m(m + 1)(n + 2)^2 + n\psi(n + 1)(m + 2)^2\big)\end{aligned}}{\begin{aligned}&b\delta(n + 2)^2(m + 2)^2 - (n\psi + 1)(n + 1)^2(m + 2)^2 - \\ &- (n - n\psi)(m + 1)^2(n + 2)^2 + n\psi(2n + 2)(m + 2)^2 + \\ &+ ((1 - \varphi)m - 1)(n - n\psi)(m + 1)(n + 2)^2 + n\psi(n\psi - 1)(n + 1)(m + 2)^2\end{aligned}}$$

Taking a derivative of x_h^* and x_s^* with respect to each of the parameters m, n, φ and ψ, we obtain a ratio, where the denominator is unambiguously positive while the sign of the numerator is determined by the sign of a cubic polynomial in δ. As soon as δ is sufficiently large – greater than the largest real root of the polynomial, the sign of the polynomial is defined by the sign of the coefficient at the highest degree.

Thus, to simplify calculations, I assume that δ is large enough ($\delta > \Delta$) and focus on the sign of the polynomial's coefficient at δ^3. I obtain that under the parameter restriction (1.23), partial derivatives $\frac{\partial x_s^*}{\partial n}$, $\frac{\partial x_h^*}{\partial m}$, and $\frac{\partial x_h^*}{\partial \psi}$ are negative and the derivatives $\frac{\partial x_h^*}{\partial n}$ and $\frac{\partial x_s^*}{\partial \varphi}$ are positive. As regarding the derivative $\frac{\partial x_s^*}{\partial m}$, this derivative is positive if and only if the following inequality holds:

$$(a-\alpha-2\tau)(1-\varphi)\cdot A+(a-\alpha-2\tau)\cdot B+\tau\cdot C > (a-\alpha-2\tau)(1-\varphi)\cdot D-(a-\alpha-2\tau)\varphi\cdot E$$

$$(1.25)$$

where

$$
\begin{aligned}
A \;=\; & m^6n^4 + 7m^6n^3 + 18m^6n^2 + 20m^6n + 8m^6 + 12m^5n^4 + 84m^5n^3 + 216m^5n^2 + \\
& + 240m^5n + 96m^5, \\
B \;=\; & -30m^4n^4\varphi + 50m^4n^4 - 300m^4n^3\varphi + 380m^4n^3 - 840m^4n^2\varphi + 1000m^4n^2 - \\
& -960m^4n\varphi + 1120m^4n - 384m^4\varphi + 448m^4 + 40m^3n^4\varphi + 100m^3n^4 - \\
& -320m^3n^3\varphi + 880m^3n^3 - 1280m^3n^2\varphi + 2400m^3n^2 - 1600m^3n\varphi + 2720m^3n - \\
& -640m^3\varphi + 1088m^3 + 240m^2n^4\varphi + 120m^2n^4 + 240m^2n^3\varphi + 1200m^2n^3 - \\
& -480m^2n^2\varphi + 3360m^2n^2 - 960m^2n\varphi + 3840m^2n - 384m^2\varphi + 1536m^2 + 288mn^4\varphi \\
& +112mn^4 + 576mn^3\varphi + 1024mn^3 + 384mn^2\varphi + 2816mn^2 + 3200mn + 1280m + \\
& +16n^5\varphi + 96n^4\varphi + 64n^4 + 192n^3\varphi + 448n^3 + 128n^2\varphi + 1152n^2 + 1280n + 512, \\
C \;=\; & 160n^4 + 1024m + 1280n + 768m^2 + 256m^3 + 32m^4 + 1280n^2 + 640n^3 + 512 + \\
& +16n^5 + 1929m^2n^2 + 960m^2n^3 + 640m^3n^2 + 320m^3n^3 + 80m^4n^2 + 24m^2n^5 + \\
& +80m^3n^4 + 40m^4n^3 + 1286mn^3 + 8m^3n^5 + 10m^4n^4 + m^4n^5 + 2560mn + \\
& +2560mn^2 + 1920m^2n + 640m^3n + 80m^4n + 32mn^5 + 320mn^4 + 240m^2n^4, \\
D \;=\; & m^4n^5 + 6m^3n^5 + 12m^2n^5 + 8mn^5, \\
E \;=\; & 2m^4n^5 + 14m^3n^5 + 36m^2n^5 + 40mn^5
\end{aligned}
$$

Notice that A, B, C, D, and E are all positive, so that the left-hand side of (1.25) is positive, while the sign of the right-hand side is determined by the relative values of $(1-\varphi)\cdot D$ and $\varphi\cdot E$. It is easy to see that $2D < E$. Hence, for $\varphi \geq 1/3$, $(1-\varphi)\cdot D < \varphi\cdot E$ and the right-hand side of (1.25) are negative. This establishes condition (b) of the proposition.

Observe also that $C > D$. Then as soon as $\tau \geq (a - \alpha - 2\tau)(1 - \varphi)$, inequality (1.25) holds. This justifies condition (a).

Finally, condition (c) follows from the series of inequalities. First, when $n \leq m^2$,

$$A > m^4 n^5 + 12 m^3 n^5 + 7 m^2 n^5 + 84 m n^5 \tag{1.26}$$

Secondly, since $m > n$,

$$m^4 n^5 + 12 m^3 n^5 + 7 m^2 n^5 + 84 m n^5 > m^4 n^5 + 6 m^3 n^5 + 13 m^2 n^5 + 84 m n^5 > D \tag{1.27}$$

Combining (1.26) and (1.27), we obtain that $A > D$, so that inequality (1.25) is satisfied.

■

Proof of Proposition 1.3

First, notice that a complete network of degree n (m) can be regarded as a hub-and-spoke network "composed only of hubs", that is, where $\psi = 1$ (composed only of spokes where $\varphi = 1$). Then inequality $x_h^* > x^*(n)$ follows from part 1 of Proposition 1.2, stating that x_h^* is decreasing in ψ. Similarly, $x^*(m) > x_s^*$ is implied by the result that x_s^* is increasing in φ. Lastly, the inequality $x^*(n) > x^*(m)$ follows from Proposition 1.1.

■

Proof of Proposition 1.4

Consider the first series of inequalities in Proposition 1.4:

$$x_{h1}^* > x_{h3}^* > x_{h4}^* > x^*(n) > x^*(m) > x_{s5}^* > x_{s3}^* > x_{s1}^*$$

There, the first three inequalities follow from part 1 of Proposition 1.2: $x_{h1}^* > x_{h3}^*$ since x_h^* is decreasing in m, while $x_{h3}^* > x_{h4}^* > x^*(n)$ since x_h^* is decreasing in ψ. Similarly, the last three inequalities are implied by part 2 of Proposition 1.2: $x^*(m) > x_{s5}^* > x_{s3}^*$ since x_s^* is increasing in φ, while $x_{s3}^* > x_{s1}^*$ since x_s^* is increasing in m. The intermediate inequality $x^*(n) > x^*(m)$ is a result of Proposition 1.1.

Likewise, with regard to the equilibrium R&D efforts x_{h2}^* and x_{s2}^* in Type 2 system, the inequality $x_{h1}^* > x_{h2}^*$ follows from the fact that x_h^* is decreasing in ψ, while $x_{h2}^* > x_{h4}^*$ and $x_{s4}^* > x_{s2}^*$ are the result of x_h^* and x_s^* being decreasing and increasing in m, respectively.

■

Proof of Proposition 1.5

The proof is suggested by the proof of Lemma 2.3 and Lemma 7.1 in Bloch and Quérou (2008).

Consider the system of linear equations (1.18). Since $\|\lambda\mathbf{B}\| \leq \frac{\bar{\varepsilon}}{N} < \frac{1}{N}$, Lemma 7.1 in Bloch and Quérou (2008) states that (1.18) possesses a unique solution and

$$\|\mathbf{x}^* - \lambda\tilde{\mathbf{u}} \cdot \sum_{k=0}^{K} \lambda^k \mathbf{B}^k\| \leq \frac{N^{K+1}\|\lambda\mathbf{B}\|^{K+1}\lambda\|\tilde{\mathbf{u}}\|}{1 - N\|\lambda\mathbf{B}\|} \leq \frac{\lambda\bar{\varepsilon}^{K+1}\|\tilde{\mathbf{u}}\|}{1 - \bar{\varepsilon}}$$

Observe that $\mathbf{c}^m$ is defined so that

$$\lambda\tilde{\mathbf{u}} \sum_{k=0}^{K} \lambda^k \mathbf{B}^k = \sum_{m=1}^{K+1} \mathbf{c}^m$$

So,

$$\|\mathbf{x}^* - \sum_{m=1}^{K+1} \mathbf{c}^m\| \leq \frac{\lambda\bar{\varepsilon}^{K+1}\|\tilde{\mathbf{u}}\|}{1 - \bar{\varepsilon}}$$

By definition of the l_∞ vector norm, this means that $\forall i \in 1 : N$

$$|\mathbf{x}_i^* - \sum_{m=1}^{K+1} \mathbf{c}_i^m| \leq \frac{\lambda\bar{\varepsilon}^{K+1}\|\tilde{\mathbf{u}}\|}{1 - \bar{\varepsilon}} \tag{1.28}$$

Consider a pair (i, j) of players and let M be the first element of the sequences $\mathbf{f}_i$, $\mathbf{f}_j$ such that $\mathbf{c}_i^M \neq \mathbf{c}_j^M$. Applying (1.28) to i and j, we obtain:

$$|\mathbf{x}_i^* - \mathbf{x}_j^* - (\mathbf{c}_i^M - \mathbf{c}_j^M)| \leq 2 \cdot \frac{\lambda\bar{\varepsilon}^{K+1}\|\tilde{\mathbf{u}}\|}{1 - \bar{\varepsilon}}$$

This concludes the proof.

■

Appendix C: Figures
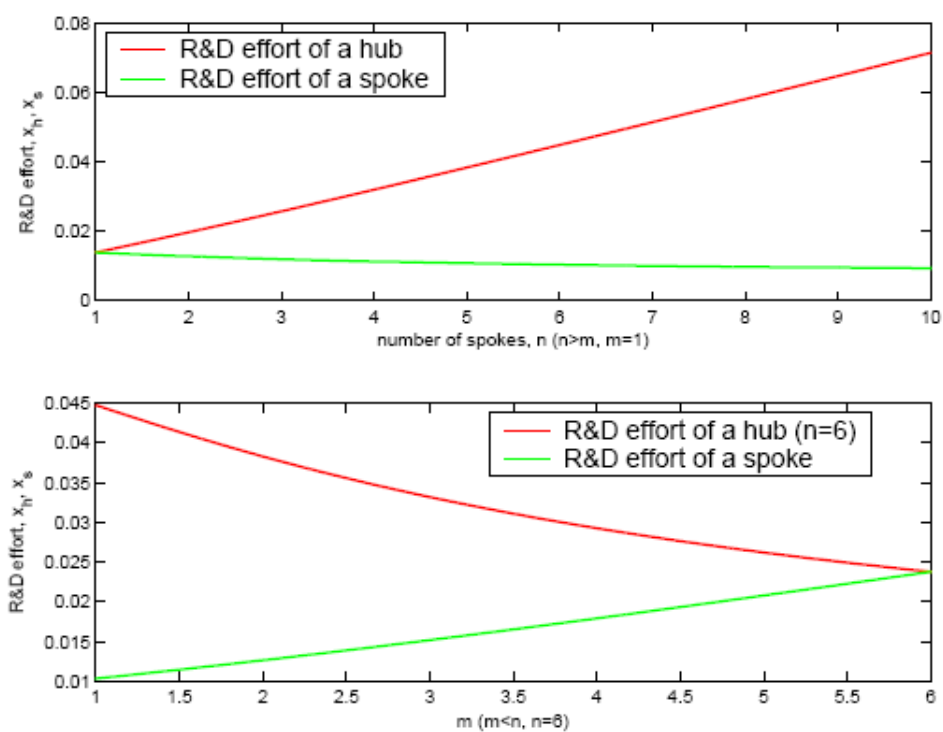
FIGURE 1.5: Equilibrium R&D efforts in the hub-and-spoke trade system as a function of n (the upper sub-figure) and as a function of m (the lower sub-figure).
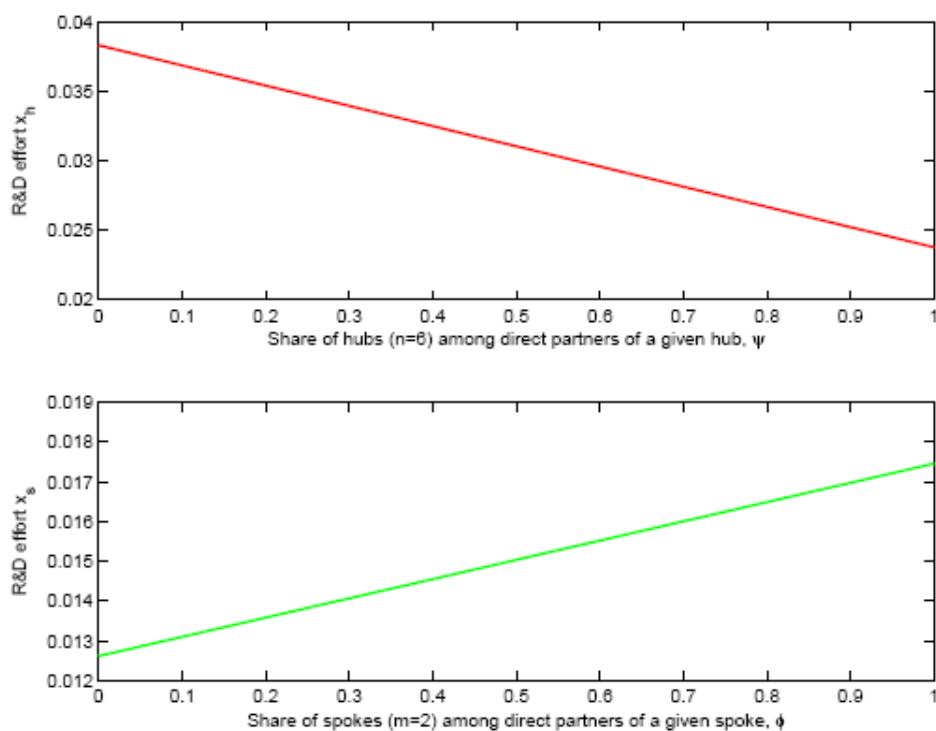
FIGURE 1.6: Equilibrium R&D efforts in the hub-and-spoke trade system as a function of ψ (the upper sub-figure) and as a function of φ (the lower sub-figure).
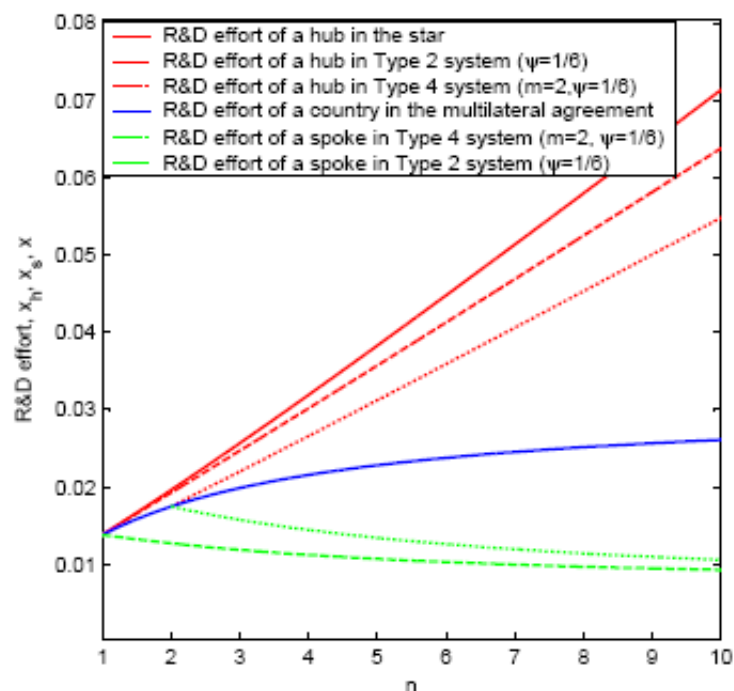
FIGURE 1.7: Equilibrium R&D efforts in Type 2 system as compared to R&D efforts in other hub-and-spoke systems and to R&D of a country in the multilateral agreement.
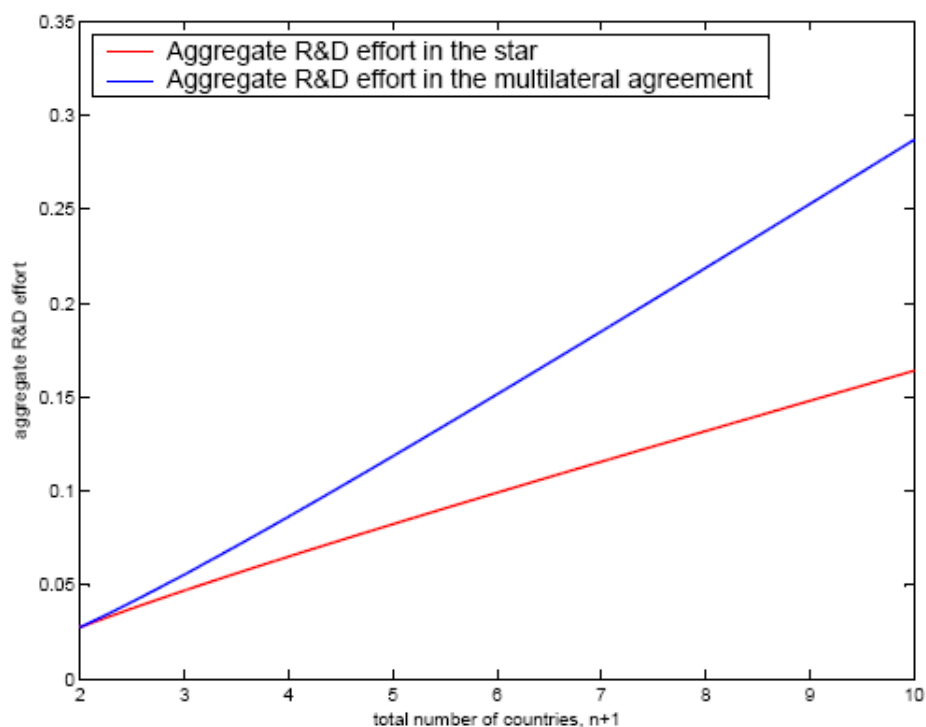
FIGURE 1.8: Aggregate equilibrium R&D efforts of n countries in the star and in the multilateral agreement.

1.10   Bibliography

Aghion P., Bloom, N., Blundell, R., Griffith, R. and P. Howitt (2005) *"Competition and Innovation: An Inverted-U relationship,"* The Quatrly Journal of Economics, 120(2): 701-728.

Alvarez R. and R. Lopez (2005) *"Exporting and Performance: Evidence from Chilean Plants,"* Canadian Journal of Economics, 38(4): 1384-1400.

Aw, B. Y., Chung, S. and M. J. Roberts (2000) *"Producivity and Turnover in the Export Market: Micro-Level Evidence from the Republic of Korea and Taiwan,"* World Bank Economic Review, 14(1): 65-90.

Baggs, J., Head, K. and J. Ries (2002) *"Free Trade, Firm Heterogeneity and Canadian Productivity,"* Paper presented at the conference "Canada-U.S. Economic and Monetary Integration", Kennedy School of Government, Harvard University, 9-11 May

Baldwin, R. E. (2003) *"The Spoke Trap: Hub-and-Spoke Regionalism in East Asia,"* KIEP Discussion Paper 04-02, Seoul: Korea Institute for International Economic Policy.

Baldwin, R. E. (2005) *"Stepping Stones or Building Blocks? Regional and Multilateral Integration,"* in Regional Economic Integration in a Global Framework (eds. J. McKay, M. O. Armengol, and G. Pineau), European Central Bank, Frankfurt.

Ballester, C., Calvo-Armengol A. and Y. Zenou (2006) *"Who's Who in Networks. Wanted: The Key Player,"* Econometrica, Econometric Society, 74(5): 1403-1417, 09.

Bernard, A. B., Eaton J., Jensen, J. B. and S. Kortum (2003) *"Plants and Productivity in International Trade,"* American Economic Review, 93(4): 1268-1290.

Bernard, A. B., Jensen, J. B. and P.K. Schott (2006) *"Falling Trade Costs, Heterogeneous Firms, and Industry Dynamics,"* Journal of Monetary Economics, 53: 917-937.

Bernard, A. B., Redding, S. J. and P.K. Schott (2007) *"Comparative Advantage and Heterogeneous Firms,"* Review of Economic Staudies, 74(1): 31-66.

Bhagwati, J. (1993) *"Regionalism and Multilateralism: an Overview,"* in Melo and Panagariya, eds.

Bloch F. and N. Quérou (2008) *"Optimal Pricing and Strategic Decisions in Networks,"* Unpublished manuscript.

Bustos, P. (2007) *"Multilateral Trade Liberalization, Exports and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms,"* Mimeo, Barcelona, Universitat Pompeu Fabra.

Dasgupta, P. (1986) *"The Theory of Technological Competition,"* in New Developments in the Analysis of Market Structure (ed. J. E. Stiglitz and F. Matthewson), London: Macmillan and Cambridge, Mass: MIT Press. Reprinted in K. Binmore and P. Dasgupta, 1986.

Dasgupta, P. and J. Stiglitz (1980) *"Industrial Structure and the Nature of Innovative Activity,"* Economic Journal, 90: 266-93.

D'Aspremont, C. and A. Jacquemin (1988) *"Cooperative and Noncooperative R&D in Duopoly with Spillovers,"* American Economic Review, 78: 1133-1137.

De la Torre, A. and M. R. Kelly (1992) *"The Regional Trading Arrangements,"* Occasional Paper, 93, IMF, Washington, DC.

De Benedictis, L., De Santis, R. and C. Vicarelli (2005) *"Hub-and-Spoke or Else? Free Trade Agreements in the 'Enlarged' European Union,"* ENERPI and ISAE.

Del Gatto, M., Mion, G. and G. I. P., Ottaviano (2006) *"Trade Integration, Firm Selection and the Costs of Non-Europe,"* Mimeo, University of Bologna.

Deltas, G., Desmet, K. and G. Facchini (2005) *"Hub-and-Spoke Free Trade Areas,"* Unpublished manuscript.

Devereux, M. and B., Lapham (1994) *"The Stability of Economic Integration and Endofenous Growth,"* The Quaterly Journal of Economics, 109: 299-305.

Diao, X., Diaz-Bonilla, E. and S. Robinson (2003) *"Scenarios for Trade Integration in the Americas,"* Economie Internationale, 94-95: 33-52.

Fiorentino, R. V., Verdeja, L. and C. Toqueboeuf (2007) *"The Changing Landscape of Regional Trade Agreements: 2006 Update,"* WTO Discussion Paper 12.

Frankel, J. A. (1997) *"Regional Trading Blocs in the World Trading System,"* Washington, DC: Institute for International Econ.

Furusawa, T. and H., Konishi (2007) *"Free Trade Networks,"* Journal of International Economics, 72: 310-335.

Goyal, S. and J. Moraga (2001) *"R&D Networks,"* RAND Journal of Economics, 32(4): 686-707.

Goyal, S. and S. Joshi (2006) *"Bilateralism and Free Trade,"* International Economic Review, 47 (3): 749-778.

Grossman, G. and E. Helpman (1991) *"Innovation and Growth in the Global Economy,"* The MIT Press.

Krugman, P. R. (1980) *"Scale Economies, Product Differentiation, and the Pattern of Trade,"* American Economic Review, 70 (5): 950-959.

Kowalczyk, C. and R. Wonnacott (1991) *"Substitute and Complement Trading Clubs,"* Dartmouth College, Department of Economics, working paper 91-16.

Kowalczyk, C. and R. Wonnacott (1992) *"Hubs and Spokes, and Free Trade in the Americas,"* NBER Working Paper 4198.

Krueger, A. O. (1999) *"Are Preferential Trading Agreements Trade-Liberalizing or Protectionist?"* Journal of Economic Perspectives, 13(4): 105-24.

Licandro, O. and A. Navas-Ruiz (2007) *"Trade Liberalization, Competition, and Growth"* Mimeo, EUI.

Lipsey, R. G. (1990) *"Canada and the U. S. - Mexico Free Trade Dance: Wallflower or Partner?,"* C. D. Howe Institute Commentary 20.

Lipsey, R. G. (1991) *"The Case for Trilateralism,"* in Continental Accord: North American Economic Integration (ed. S. Globerman), Vancouver: The Fraser Institute.

Lloyd, P. J. and D. Maclaren (2004) *"Gains and Losses from Regional Trading Agreements: A Survey,"* The Economic Record, 80(251): 445-467.)

Mauleon, A., Song, H. and V. Vannetelbosch (2006) *"Networks for Free Trade Agreements among Heterogeneous Countries,"* Working Paper 2006029, Université catholique de Louvain, Département des Sciences Economiques.

Melitz, M. J. (2003) *"The Impact of Trade on Intra-Industry Reallocations and Aggreate Industry Productivity,"* Econometrica, 71: 1695-1725.

Muendler, M. A. (2004) *"Trade, Technlogy, and Productivity: A Study of Brazilian Manufacturers, 1986-1998,"* CESifo Working Paper 1148, Munich: CESifo GmbH.

Navas-Ruiz, A. and D. Sala (2007) *"Technology Adoption and the Selection Effect of Trade,"* Working Paper 07-67, Universidad Carlos III de Madrid

Panagariya, A. (2000) *"Preferential Trade Liberalization: The Traditional Theory and New Developments,"* Journal of Economic Literature, 38 (2): 287-331.

Pavcnik, N. (2002) *"Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants,"* Review of Economic Studies, 69 (1): 245-276.

Peretto, P. (2003) *"Endogenous Market Structure and the Growth and Welfare Effects of Economic Integration,"* Journal of International Economics, 60: 177-201.

Rivera-Batiz, L. and P. Romer (1991) *"Economic Integration and Endogenous Growth,"* The Quaterly Journal of Economics, 106: 531-556.

Srinivasan, T. N., Whalley, J. and I. Wooton (1993)*"Measuring the Effects of Regionalism on Trade and Welfare,"* in Regional Integration and the Global Trading System (eds. K. Anderson and R. Blackhurst), NY: St. Martin's Press, 52-79.

Topalova, P. (2004) *"Trade Liberalization and Firm Productivity: The Case of India,"* IMF Working Paper WP/04/28

Trefler, D. (2004) *"The Long and Short of the Canada-U.S. Free Trade Agreement,"* American Economic Review, 94(4): 870-895.

Van Biesebroeck, J. (2005) *"Exporting Raises Productivity in sub-Saharan African Manufacturing Firms,"* Journal of International Economics, 67(2): 373-391.

Wonnacott, R. (1990) *"U. S. Hub-and-Spoke Bilaterals and the Multilateral Trading System,"* C. D. Howe Institute Commentary 23.

Wonnacott, R. (1991) *"The Economics of Overlapping Free Trade Areas and the Mexican Challenge,"* C. D. Howe Institute and National Planning Association.

Wonnacott, R. (1996) *"Trade and Investment in a Hub-and-Spoke System versus a Free Trade Area,"* World Economy, 19: 237-252.

World Trade Report 2008 *"Trade in a Globalizing World"*.

Yeaple, S. R. (2005) *"A Simple Model of Firm Heterogeneity, International Trade, and Wages,"* Journal of International Economics, 65: 1-20.

Chapter 2

Resilience of the Interbank Network to Shocks and Optimal Bail-Out Strategy: Advantages of "Tiered" Banking Systems

2.1 Introduction

One of the major concerns in recent policy debates over financial stability is how "tiering" in the banking system may impact on so-called systemic risk, the large-scale breakdown of financial intermediation due to the domino effect of insolvency.[1] The tiered banking system is commonly defined as an organization of lending-borrowing relations/linkages between banks, where relatively few first-tier or "head" institutions have a large number of interbank linkages, whereas many second-tier or "peripheral" banks have only few links. First-tier banks are connected to second-tier banks and are also connected with each other, whereas second-tier banks are almost exclusively connected to first-tier banks. Interbank linkages may act as a device for co-insurance against uncertain liquidity shocks (Bhattachrya and Gale (1987)) and/or improve market discipline by providing incentives for peer-monitoring (Flannery (1996) and Rochet and Tirole (1996)) but can also serve as a channel through which problems in one bank spread to another.

Tiered banking systems are found in a range of countries but the empirical evidence of contagion risk in these systems is mixed. In their 2003 Financial System Stability assessment of the United Kingdom, the International Monetary Fund (IMF) highlights

---

[1] The domino effect of insolvency occurs when the non-repayment of interbank obligations by the failing bank jeopardizes the ability of its creditor banks to meet their obligations to interbank creditors.

the potential contagion risk arising from the highly tiered structure of the U.K. large-value payment systems. However, several subsequent studies including Wells (2004), Harrison et al. (2005), Lasaosa and Tudela (2008) report relatively limited scope for contagion among U.K. banks. Boss et al. (2004b) and Degryse and Nguyen (2005) find that tiered banking systems in Austria and Belgium are stable and systemic crises are unlikely. By contrast, Upper and Worms (2004) suggests that in the structurally similar banking system in Germany, "the effects of the breakdown of a single bank could potentially be very strong" (p. 847) and system-wide bank failures are possible. In addition, somewhat differently from other studies, Elsinger et al. (2006) and Mistruilli (2007) find that while contagious failures in tiered Austrian and Italian banking systems are relatively rare, large parts of the system are affected in the worst-case scenarios.

The controversy in the empirical literature leaves the question of benefits and risks of tiering open for further investigation. In this paper, I aim to shed more light on this issue by proposing an analytical and simulations-based approach. I develop a simple theoretical model to study how structural characteristics of a tiered banking system may affect its susceptibility to systemic breakdown and the scope of the breakdown, particularly in comparison with other types of systems.[2] This enables an analysis of the scale of bank bail-outs that might be needed to sustain the stability of tiered banking system and allows comparing the costs of bail-outs across systems.

The banking system is modelled with a *random* network where nodes represent banks and links are interbank exposures that connect any two banks with a certain predefined probability. Links in the network are directed, reflecting the fact that interbank exposures comprise assets as well as liabilities. An important assumption about the structure of links is that with probability 1 the number of incoming and outgoing links of each bank is the same. Furthermore, interbank assets of a bank are assumed to be evenly distributed over its incoming links and the total amount of these assets is set equal across banks.[3] Admittedly, this is a restrictive framework. However it allows deriving some results of the paper analytically and brings attention to effects of structural features of banking systems on contagious defaults in otherwise symmetric setting.

I examine the impact of tiering in the banking system on the stability of the system by studying the consequences of negative correlation in connectivity (known as *degree*) between neighboring banks/nodes and of the highly right-skewed distribution of connectivity across banks/nodes, such as for example, in the case of a power-law, or scale-free, distribution.

---

[2]In fact, the susceptibility to systemic breakdown of every system is only evaluated in *relative* terms. This allows *comparative* assessment of risk across various types of banking systems.

[3]The assumptions of even distribution of interbank assets over incoming links and identical interbank asset positions of all banks in the system are the same as in Gai and Kapadia (2009).

First, I model a tiered banking system with a network displaying negative degree correlations (a *disassortative* network),[4] and compare this network with other types of structures, showing either positive degree correlations (an *assortative* network) or no correlations (a *neutral* network). For these three types of banking networks, I employ a numerical analysis to investigate (i) the resilience of the networks to systemic failure, (ii) the scale of the failure if systemic crisis occurs, and (iii) the optimal bail-out strategy of the government or of the central bank to guarantee the financial stability of the system at minimal costs.[5] In studying the optimal bail-out strategy of the government, I restrict attention to a common practice of "targeted" bail-outs, that is, the policy to first rescue the most connected failing banks. While the cost of each bail-out is not modelled explicitly, the optimal bail-out strategy minimizes the *aggregate* cost of bail-outs since by construction, it minimizes the total number of targeted bank rescues.

I find that both, the risk of systemic crisis and the scope of the crisis in the banking system are minimized when the network is disassortative/tiered. Intuitively, in a disassortative banking network, high-degree banks are broadly distributed over the network and therefore, presumably form links on many paths between other banks. As a result, with high probability an initial shock hitting a random bank reaches a high-degree bank in a small number of "steps", gets absorbed by that bank and does not spread any further. This implies that disassortative networks are relatively resilient to shocks. For the same reason, targeted bail-outs are less costly in a disassortative network: while many bank rescues may be required in assortative and neutral banking networks, in a disassortative network, bail-outs may not be needed at all.

Further, I consider a tiered banking system as represented by the *scale-free network* and study both analytically and numerically the impact of tiering in such a system on its susceptibility to shocks. I focus on the effects of variation in the *level of tiering* in scale-free networks, where the level of tiering is represented by the inverse of a parameter of the scale-free distribution.[6] In the scale-free network, the degree distribution is such that an absolute majority of banks have very low degree whereas a few banks have high degree. High-degree banks are mainly connected to low-degree ones and those low-degree banks

---

[4]In the next section, it is argued that even with negative degree correlation between neighboring nodes, first-tier highly connected banks are still likely to be connected with each other, while the probability of connectedness between second-tier low-degree banks is very small.

[5]The recent events triggered by the sub-prime crisis of August 2007 highlighted the criticality of these questions. For example, recent rescue of some institutions, such as American International Group (AIG), remains a highly disputable issue. As a rule, the main argument of policymakers in favour of these rescues is that many yet unaffected banks (across the national or international financial system) might be exposed to the defaulting institutions. But in fact, no rigorous assessment is suggested of how far contagion could have spread had AIG been allowed to fail.

[6]The parameter of the scale-free degree distribution governs the rate at which probability decays with connectivity. Therefore, for smaller values of this parameter, the fraction of highly-connected banks in the network is larger and so is the probability that poorly-connected banks are linked with highly-connected banks rather than with each other.

rarely have links anywhere else. Some empirical studies find close parallels to scale-free networks in real-world tiered banking systems. For example, Boss et al. (2004) confirms that the Austrian interbank market has a scale-free structure. An advantage of using the scale-free distribution to study tiered network structures is its analytical tractability. In the simplified framework of absent degree correlations, the scale-free distribution allows obtaining the exact closed-form solution for the threshold at which systemic failure can occur and government intervention is needed.

I find that the resilience of scale-free banking networks to systemic breakdown is *increasing* and the maximal expected number of required bail-outs is *decreasing* in the level of tiering. When the system is more tiered, links emanating from and arriving at highly-connected banks span a larger part of the banking system. This implies that initial shocks at any part of the system reach highly-connected banks "quickly" and with large probability subside at these banks. Therefore, the resilience of such a system to global breakdown is higher and the expected number of necessary targeted bail-outs is lower than in a system with lower tiering.

Thus, the findings of this paper demonstrate the advantages of tiering in the banking network. This highlights some specific regulatory issues. For example, with regard to stability-improvement of the interbank market, the results suggest that regulation should promote the tiered structure of interbank relations since this can reduce systemic risk. Furthermore, when designing the optimal bail-out policy, regulators should consider specific features of the degree distribution underlying the pattern of connections in the banking network. This may help to determine the limits of government regulation when confronted with systemic shock, so as to guarantee the global stability of the system at minimal cost.

To solve the model, I apply techniques from the literature on complex networks (Strogatz (2001), Newman et al. (2001), Vega-Redondo (2007)) to a financial system setting. I use these techniques to model contagion stemming from unexpected shock to a single institution in complex banking networks. The banking system and transmission of shocks are modelled as in Gai and Kapadia (2009). However, I focus on the impact of tiering in the banking system on the risk and potential scope of defaults, while Gai and Kapadia (2009) examine the susceptibility to shocks of a *uniform* (Poisson) banking network, where each possible directed link is present with *independent* and *identical* probability. I show that when the degree distribution is not highly skewed, disregarding degree correlations between banks may substantially change the predictions for the risk and scale of systemic defaults in the tiered banking network.

Within the complex network literature, I exploit results from the standard epidemic/information diffusion and percolation literature. Specifically, the framework builds on the

generating-function techniques used in the SI(R) models[7] of Watts (2002), Callaway et al. (2000), Cohen et al. (2001), and Newman (2002b), and in the structured network model of Newman (2002a). This literature describes the behavior of connected groups of nodes in a random network, with or without internode degree correlations, and characterizes *phase transition*, the point at which extensive contagion outbreaks occur, as well as the size of a susceptible cluster beyond that point.

Unlike the generic, undirected network models of Watts (2002) and Newman (2002a), the model in this paper provides an explicit characterization of balance sheets, which specifies the direction of links connecting banks in the financial system. The distinction between incoming links (claims) and outgoing links (obligations) implies that in contrast to epidemiological and percolation models, greater connectivity does not only create more channels through which contagion can spread but also improves counteracting risk-sharing benefits, since exposures are diversified across a larger set of banks. Moreover, whereas in most epidemiological models the susceptibility of a node to contagion is determined solely by the total number of its infected neighbors, in the present setup the *share* of neighbors that default determines the contagion risk.

An alternative approach to studying contagion in financial networks is represented by the seminal contribution of Allen and Gale (2000) and by more recent literature on endogenous network formation (Leitner (2005), Castiglionesi and Navarro (2007)). These studies are based on small networks with rigid structures and provide key insights into the mechanism through which the pattern of interconnectedness between banks affects the spread of defaults. However, being restricted to the analysis of simple networks, this literature is limited in addressing the case of real-world contagion in large and complex banking systems. For the same reason the representation of tiered systems in this approach is confined to a very limited set of cases.[8] In contrast, the complex network approach allows for a wide range of network structures with arbitrarily large number of banks. Furthermore, given the scarce information that policymakers have about the true interlinkages involved, the connections between banks are, perhaps, best captured by a random network.

At the same time a natural critique of the approach in this paper is that it assumes that interbank connections are formed randomly and exogenously and are static in nature.

---

[7]SI(R) (*susceptible-infected (-recovered)*) models are canonical epidemiological models, where the life history of each node passes from being susceptible (S), to becoming infected (I) and, in the SIR setting, to finally being recovered (the definition based on Vega-Redondo (2007, p.75)). The primary theoretical approach used in the SI(R) context is the generating-function analysis.

[8]For example, one representation of a tiered structure in the deterministic network setting is the "money center bank" structure proposed by Freixas et al. (2000). They consider a model with three banks, where the banks on the periphery are linked to the bank at the center but not to each other. In this framework, they find that for some parameter values the failure of a bank on the periphery will not trigger the breakdown of other institutions while the failure of the money center bank would.

This leads to modelling contagion in a relatively mechanical manner, keeping balance sheets and the structure of interbank linkages fixed as default propagates through the system. Yet, as suggested by Gai and Kapadia (2008), this framework still yields a useful and realistic benchmark for the analysis. Arguably, in crises contagion spreads rapidly through the system and banks have little time to change their behavior before they are affected. Moreover, banks have no choice over whether to default, which precludes strategic behavior on networks of the type assumed in Morris (2000), Jackson and Yariv (2007) and Galeotti and Goyal (2007).

The remainder of the paper is organized as follows. Section 2.2 introduces the model and describes the structure of the banking network, the transmission process for contagion, and the generating-function approach to measure the extent of contagion and phase transition. Section 2.3 uses the introduced techniques to study the effects of a failure of an individual institution on the risk and extent of system-wide contagion in tiered and other banking systems. Section 2.4 summarizes the findings and suggests potential policy implications.

2.2 The model

2.2.1 Interbank network

Consider a banking system represented by a network where each node is a bank and each link represents a directional lending relationship between two banks. Two crucial properties of this network are that it is (i) *directed* and (ii) *random.* The first property implies that links in the network are directed, so that every node has two degrees: an *in-degree*, the number of links that point into the node, and an *out-degree*, the number of links that point out. Incoming links to a bank reflect the interbank assets of the bank, that is, funds owed to the bank by a counterparty. In contrast, outgoing links from a bank indicate its interbank liabilities. Randomness of the network means that any two banks are linked with a certain predefined probability, so that the connectivity, or degree, of each bank is not deterministic but random. In case of a directed random network, the connectivity of nodes is described by the *joint probability distribution of in- and out-degrees*. I assume a specific form of the joint distribution such that with probability 1 in- and out-degrees of each node are the same:

Assumption 1 $\qquad p_{jk} = p(k)\delta_{jk}$

where j denotes in-degree of a bank, k denotes out-degree, and δ_{jk} is the Kronecker delta, i.e. $\delta_{jk} = 1$ if $j = k$ but 0 otherwise. By definition, $p(k)$ is the probability

that a randomly chosen node has in- and out-degrees each equal to k. The assumption of almost surely equal number of lenders and borrowers of each bank in the interbank market is imposed for two reasons. First, it allows for the notion of highly-connected and "peripheral" banks in an intuitive manner: both in- and out-degree of highly-connected banks are large and both in- and out-degree of peripheral banks are low.[9] Secondly, it simplifies the discussion and the derivation of the results thereafter.[10]

Also in this framework, the network of interbank linkages is assumed to be formed exogenously: issues related to endogenous network formation, optimal network structures and network efficiency are left to one side.

2.2.2 Tiered network structure

A tiered banking network is defined as a network in which relatively few high-degree (first-tier) banks\nodes are connected with low-degree (second-tier) banks\nodes and also connected with each other, whereas low-degree banks are almost exclusively connected with high-degree banks. This property of tiering can be modelled by assuming that the banking network shows *disassortative mixing* on its degrees, that is, degrees of neighboring banks in the network are negatively correlated. It should be noticed that even with negative degree correlations, first-tier highly connected banks are still likely to be linked with each other. This is explained by the fact that in *any* random network, the probability of a given source node being connected to a target node is an increasing function of the degree of the target node. Disassortativity implies that the probability of connectedness between second-tier low-degree banks is decreased and that the probability of connectedness between low-degree and high-degree banks is increased relative to the situation when degree correlations between neighboring nodes are either absent or positive. In fact, if the network displays zero or positive internode degree correlations, low-degree banks are likely to be connected with each other, especially if the proportion of these low-degree banks in the network is sufficiently large.

Alternatively, as suggested by Boss et al. (2004) and Nier et al. (2007), the same property of tiering in banking networks can be captured by the assumption that nodes' degree distribution is highly right-skewed. One example of such distribution is a power-law, or scale-free, distribution. In *scale-free networks*, the vast majority of nodes have small degree and nodes with high degree, although relatively few, display connectivity

---

[9]It also seems likely (although this question could be for future research) that in the real-world interbank market, banks with a large number of outgoing links also have a large number of incoming ones and peripheral banks have a low number of both incoming and outgoing links.

[10]For example, it makes the definition of *internode* degree correlation and joint probability distribution of the neighboring nodes more intuitive.

that greatly exceeds the average. As a result, high-degree nodes attach mainly to low-degree ones and those low-degree nodes rarely have links to other nodes, including the nodes of low-degree.

2.2.3   Bank's solvency condition

An individual bank's assets consist of external assets (investors' borrowing), denoted by A_i^E, and interbank assets (other banks' borrowing), denoted by A_i^{IB}. As in Gai and Kapadia (2009), the total interbank asset position of each bank is assumed to be evenly distributed over its incoming links and independent of the number of links the bank has[11]:

Assumption 2 $A_i^{IB} = j_i w_i,$ and $A_i^{IB} = A_k^{IB} = A^{IB}$ $\forall i \neq k$

where j_i is the number of incoming links of bank i and w_i is the interbank assets held by i against any of its debtor banks. These assumptions provide a useful benchmark for studying the effects of tiering in the banking network on the potential for contagion spread. By setting the size of interbank assets equal for all banks, they isolate the effects of risk sharing and risk spreading as predefined by the structure of interbank links. In particular, they allow me to show that widespread contagious defaults may still occur even if risk sharing between banks in the system is maximized.

A bank's liabilities are composed of interbank liabilities, denoted by L_i^{IB}, and customer deposits, denoted by D_i. Since every interbank liability of a bank is another bank's asset, interbank liabilities are determined endogenously. Customer deposits are instead exogenous.

For any bank i the condition to be solvent can be written as[12]:

$$(1 - \varphi)A^{IB} + A_i^E - L_i^{IB} - D_i \geq 0 \tag{2.1}$$

where φ is the fraction of banks with obligations to bank i that have defaulted. Here I assume that when a linked bank defaults, bank i loses *all* of its interbank assets held

---

[11]Without loss of generality, each bank is assumed to have at least one incoming link, so that interbank assets of any bank are strictly positive.

[12]The price of external assets is fixed at 1 so that liquidity effects associated with the knock-on defaults are ruled out. In principle, Acharya and Yorulmazer (2007) show that the asset price may be depressed (become less than 1) when failed banks' assets need to be sold but financial markets have limited overall liquidity to absorb the assets.

against that bank.[13] Alternatively, the solvency condition can be stated as:

$$\varphi \leq \frac{K_i}{A^{IB}} \qquad (2.2)$$

where $K_i = A^{IB} + A_i^E - L_i^{IB} - D_i$ is the capital buffer of bank i, or the net worth of a bank, equal to the difference between the book value of bank's assets and liabilities.

2.2.4 The transmission of shocks and bank's vulnerability

I examine the consequences of an unexpected idiosyncratic shock hitting one of the banks in the system. This shock can be thought of as resulting from operational risk (fraud) or credit risk.[14] Although for credit risk in particular, aggregate or correlated shocks affecting all banks at the same time may be more relevant in practice, idiosyncratic shocks are a clearer starting point for studying contagious defaults due to interbank exposures. Moreover, a single bank failure may actually result from an aggregate shock which has particularly adverse consequences for one institution.[15]

The failure of one bank reduces the interbank asset positions of all creditor banks and may give rise to a wave of contagious defaults in the network. In fact, the larger the number of *outgoing* links of a defaulting institution, the larger the potential for the spread of shocks. On the other hand, the impact of a bank default on the solvency of its creditors is determined by the degree of diversification of creditors' asset portfolio. The larger the number of *incoming* links of a creditor bank, the smaller the losses it incurs due to a default of a borrowing neighbor and the lower the risk of contagious failures in the system.

To demonstrate the latter inference, let us denote by j_i the number of incoming links of bank i. Since creditor banks each lose a fraction $1/j_i$ of their interbank assets when a single counterparty defaults, inequality (2.2) suggests that a default can only spread if there is at least one creditor bank for which

$$\frac{1}{j_i} > \frac{K_i}{A^{IB}}, \qquad \text{or} \qquad j_i < \frac{A^{IB}}{K_i} \qquad (2.3)$$

Following Gai and Kapadia (2009), bank i is called *vulnerable* if its in-degree fulfils (2.3), that is, if i is exposed to the default of a single neighbor. The other banks are called

---

[13]The same "zero recovery" assumption is made in Gai and Kapadia (2009). The authors argue that this assumption is realistic in the midst of a crisis: in the immediate aftermath of a default, the rate and timing of recovery are highly uncertain and banks' funders are likely to assume the worst-case scenario.

[14]See for example, the case of a failure of Barings in the U.K., Drexel Burnham Lambert in the U.S. or Société Générale in France.

[15]In this model, the aggregate shocks can be captured through a simultaneous reduction in the capital stocks of all banks, with a major loss for one particular bank.

safe. According to the figures for developed countries reported by Upper (2007) and to the calibration results in Gai and Kapadia (2009) based on data for published accounts of large, international financial institutions, the ratio $\frac{A^{IB}}{K_i}$ lies in the range between 5 and 6.[16] In section 2.3 I will use this finding to demonstrate some of the model's results.

For simplicity, in the following I assume that capital buffers of all banks are identical:

Assumption 3 $\qquad K_i = K_j = K \quad \forall i \neq j$

This allows denoting the right-hand side of (2.3) with a constant identical for all banks. Let $\overline{K} = \frac{A^{IB}}{K}$. Then from condition (2.3) it is clear that given $\overline{K}$, the vulnerability of a bank is fully determined by its in-degree j_i.

Thus, incoming and outgoing links of each bank play opposite roles in determining the extent of contagion. Their joint effect is predetermined by the specific features of the network structure as specified by the degree distribution. In the next section, I introduce various types of network structures and later on study the impact of structural characteristics on the spread of defaults in more detail.

2.2.5 The reach of contagious defaults

To compare the extent of contagious defaults and the risk of system-wide contagion across tiered and other types of banking structures, I introduce the notion of internode degree correlation and define three types of networks according to the pattern of this correlation.

Assortative, disassortative and neutral networks

Recall that the joint distribution of in- and out-degrees in the banking network is such that with probability 1 each node has identical in- and out-degree. Let $\xi(k, k')$ be the probability that a randomly selected link goes from a node with in- and out-degree k (node (k, k)) to a node with in- and out-degree k' (node (k', k')). This quantity obeys the sum rules:

$$\sum_{k,k'=1}^{\infty} \xi(k, k') \;=\; 1, \tag{2.4}$$

$$\sum_{k'=1}^{\infty} \xi(k, k') \;=\; \zeta_d(k), \quad k = 1, 2, \dots \tag{2.5}$$

$$\sum_{k=1}^{\infty} \xi(k, k') \;=\; \zeta_c(k'), \quad k' = 1, 2, \dots \tag{2.6}$$

---

[16]Further details are available from the author.

where $\zeta_d(k)$ represents the marginal frequency of neighboring *debtor* banks with in- and out-degree k and $\zeta_c(k')$ is the marginal frequency of neighboring *creditor* banks with in- and out-degree k'. Notice however that every debtor with in- and out-degree k is simultaneously a creditor with the same degree. This implies that the frequency of debtor and creditor banks with degree k is the same: $\zeta_d(k) = \zeta_c(k)$. I denote this quantity $\zeta(k)$ and call it the marginal frequency of neighboring nodes with in- and out-degree k. Notice that $\zeta(k)$ is different from $p(k)$, the probability that a *randomly* chosen node on a network has in- and out-degree k. Instead it is biased in favor of nodes of high degree since more edges end (and start) at a high-degree node than at a low-degree one. Therefore, the degree distribution of a neighboring node is proportional to $kp(k)$ and the correctly normalized distribution is given by:

$$\zeta(k) = \frac{kp(k)}{\sum_{k=1}^{\infty} kp(k)} \tag{2.7}$$

If internode degree correlations are absent, the network is called *neutral* and $\xi(k, k')$ takes the value $\zeta(k) \cdot \zeta(k')$. Otherwise, $\xi(k, k')$ is different from this value and the network is either *assortative* or *disassortative*. The network is called assortative if high-degree nodes show tendency to be connected to other high-degree nodes. Conversely, the disassortative network is one where high-degree nodes tend to attach to low-degree ones. The amount of assortative/disassortative mixing can be quantified by the function which represents the degree correlation between neighboring nodes:

$$\langle kk' \rangle - \langle k \rangle \langle k' \rangle = \sum_{k,k'=1}^{\infty} kk'(\xi(k, k') - \zeta(k)\zeta(k')) \tag{2.8}$$

where $\langle \ldots \rangle$ indicates an average over links. This correlation is equal to zero for neutral networks and positive or negative for assortative or disassortative networks, respectively. For convenience of comparing different networks, the function in (2.8) can be normalized by dividing by its maximal value; the value it achieves on a perfectly assortative network. In the perfectly assortative network, $\xi(k, k') = \zeta(k)\delta_{kk'}$ and the degree correlation between neighboring nodes is equal to the variance of the distribution $\zeta(k)$:

$$\sigma_\zeta^2 = \sum_{k=1}^{\infty} k^2 \zeta(k) - \left[\sum_{k=1}^{\infty} k\zeta(k) \right]^2 \tag{2.9}$$

So, the normalized correlation function is

$$r = \frac{1}{\sigma_\zeta^2} \sum_{k,k'=1}^{\infty} kk'(\xi(k, k') - \zeta(k)\zeta(k')) \tag{2.10}$$

This is the standard Pearson correlation coefficient of the degrees at either end of a link and $-1 \leq r \leq 1$.

For the purpose of further analysis, I also use the joint degree distribution of the linked nodes, $\xi(k, k')$, to obtain the collection of conditional probability distributions $p(k'|k)$, $k = 1, 2, ..$, where $p(k'|k)$ denotes the conditional probability that there is a link going from a node with in- and out-degree k to a node with in- and out-degree k'. This probability is given by:

$$p(k'|k) = \frac{\xi(k, k')}{\sum_{k''=1}^{\infty} \xi(k, k'')} = \frac{\xi(k, k')}{\zeta(k)}, \quad k = 1, 2, \ldots \tag{2.11}$$

Notice that if internode degree correlations are absent,

$$p(k'|k) = \zeta(k') \tag{2.12}$$

Throughout the remainder of section 2.2.5, I study the extent of contagion and thresholds for the system-wide contagion, conditional on a general form of the joint probability distribution for the degrees of the neighboring nodes, $\xi(k, k')$. Then in section 2.3.1, I return to the question of variations in the pattern of degree correlations and study the effect of these variations on the spread of contagious defaults.

Generating functions

To evaluate the extent of contagion in banking networks, I use the approach based on probability generating functions.[17] First, assume that the components outgoing from the defaulting node are tree-like in structure, that is, contain no cycles, or closed loops.[18] This assumption is likely to hold if the number of nodes/banks in the financial network, n, is sufficiently large since the chances of a component containing a closed loop decrease as n^{-1}. The tree-like structure of the components outgoing from the defaulting node guarantees that any bank in the system can be exposed to a default of no more than one borrowing neighbor and no second round of contagion can occur. This implies that safe banks, resistant to a default of a single debtor, never default. As a result, contagion in the banking system propagates only through vulnerable banks: from one vulnerable bank to another.

Given this feature of the model, characterizing the spread of contagious defaults in the banking network comes to deriving the distribution of the sizes of components,

---

[17]A detailed description of the key properties of the probability generating functions can be found in Newman et al. (2001) and Vega-Redondo (2007).

[18]The assumption of absent closed loops in the financial network is supported by empirical research which implies that clustering coefficients in real-world banking networks are relatively low (for example, Boss et al. (2004)).

or clusters, of vulnerable banks that can be reached after an initial default. First, consider the clusters of vulnerable banks that can be reached by following a *randomly chosen directed link*, after an initial default. Let $H_1(y|k)$ be the function generating this probability distribution conditional on the fact that a randomly chosen link via which the vulnerable components are accessed is outgoing from the bank with in- and out-degree k. The pattern of shock transmission can take many different forms. A random link emanating from the node/bank (k,k) may lead to a component consisting of a single bank, either safe or vulnerable, or it may lead to a vulnerable bank with one, two, or more other tree-like vulnerable clusters emanating from it via single outgoing edges. Clusters are reached *independently* for each of the outgoing edges. Then, assuming that all clusters are almost surely of finite size, $H_1(y|k)$ must satisfy a self-consistency condition of the form:

$$H_1(y|k) = Pr[\text{reach safe bank}|(k,k)] + y \sum_{j'=1}^{\overline{K}} \sum_{k'=1}^{\infty} p[(j',k')|(k,k)] \big(H_1(y|k')\big)^{k'}, \quad k = 1, 2, \ldots$$

(2.13)

where $p[(j',k')|(k,k)]$ is the conditional probability that a node (j',k') is reached via the outgoing link from a node (k,k). By definition, $p[(k',k')|(k,k)] = p(k'|k)$, where $p(k'|k)$ is given by (2.11). The leading factor y in (2.13) accounts for the first node encountered along the initial edge and I have used the fact that a generating function of the sum of m i.i.d. random variables is equal to the mth power of the generating function of a single random variable. Notice that the probability of reaching a safe bank from the initial defaulting bank (k,k), $Pr[\text{reach safe bank}|(k,k)]$, is equal to 1 minus the probability of the complementary event that a vulnerable bank is reached:

$$Pr[\text{reach safe bank}|(k,k)] = 1 - \sum_{j'=1}^{\overline{K}} \sum_{k'=1}^{\infty} p[(j',k')|(k,k)]$$

(2.14)

Using (2.14) and recalling that with probability 1 all nodes in the network have identical in- and out-degree, we can write (2.13) as:

$$H_1(y|k) = \left(1 - \sum_{k'=1}^{\overline{K}} \frac{\xi(k,k')}{\zeta(k)}\right) + y \sum_{k'=1}^{\overline{K}} \frac{\xi(k,k')}{\zeta(k)} \big(H_1(y|k')\big)^{k'}, \quad k = 1, 2, \ldots$$

(2.15)

Based on functions $H_1(y|k)$, $k = 1, 2, \ldots$, it is now possible to characterize the distribution of sizes of the outgoing vulnerable components reachable from a *randomly selected bank*, that is, the distribution of sizes of the vulnerable components to which a randomly

selected bank belongs. This distribution is generated by:

$$
\begin{aligned}
H_0(y) &= Pr[\text{bank is safe}] + y \sum_{j=1}^{\overline{K}} \sum_{k=1}^{\infty} p_{j,k} \big(H_1(y|k)\big)^k \\
&= \left(1 - \sum_{k=1}^{\overline{K}} p(k)\right) + y \sum_{k=1}^{\overline{K}} p(k)\big(H_1(y|k)\big)^k \quad k = 1, 2, \ldots
\end{aligned}
\tag{2.16}
$$

Two possibilities are contemplated in (2.16): either a randomly chosen bank is safe or it is vulnerable with in- and out-degree k, $k = 1, 2, \ldots$, and each of its k outgoing links leads to a vulnerable cluster with size distribution generated by $H_1(y|k)$.

In principle, one can now solve equations (2.15) for the collection of $H_1(y|k)$, $k = 1, 2, \ldots$, and substitute into (2.16) to get $H_0(y)$. In practice, this is usually impossible: (2.15) is a complicated and often transcendental equation, which rarely has a known solution.[19] One way to tackle this problem is to find a numerical approximation of $H_1(y|k)$ for any $k = 1, 2, \ldots$, using iterations of (2.15) with initial set of values $H_1 = 1$. However, for the purposes of this paper, it is enough to obtain the first moment of the distribution of the vulnerable cluster sizes, the expression for the *average* size of the vulnerable cluster.

Average size of the vulnerable component and phase transition

I now derive the expression for the average size of the vulnerable cluster in the banking network and study the conditions for the emergence of the *giant vulnerable component*, when the size of the vulnerable cluster diverges. Formally, a giant component is a unique component whose relative size remains bounded *above* zero as the number of nodes in the network increases indefinitely. In the framework of this model, the formation of the giant vulnerable cluster can be interpreted as a threshold condition for the possibility of system-wide contagion, or "global" banking crisis: with positive probability, a random initial default of one bank can cause failure of a substantial fraction of vulnerable institutions in the banking system. Then the relative size of the giant vulnerable component can be regarded as a scope of the crisis should systemic failure occur.

Due to a basic property of generating functions, the average size of the vulnerable cluster, S, can be computed as:

$$
S = H_0'(1)
\tag{2.17}
$$

Taking a derivative of $H_0'(y)$ in (2.16) and evaluating it at $y = 1$, we obtain that

$$
S = \sum_{k=1}^{\overline{K}} p(k)\big(H_1(1|k)\big)^k + \sum_{k=1}^{\overline{K}} kp(k)\big(H_1(1|k)\big)^{k-1} H_1'(1|k)
\tag{2.18}
$$

---

[19]For reference, see for example, Newman et al. (2001).

$H_1(y|k)$ is a standard generating function so that $H_1(1|k) = 1$ for all k. Also, $kp(k) = \langle k \rangle \zeta(k)$ as follows from the definition of $\zeta(k)$ in (2.7). Therefore, (2.18) becomes:

$$S = \sum_{k=1}^{\overline{K}} p(k) + \langle k \rangle \sum_{k=1}^{\overline{K}} \zeta(k) H_1'(1|k) \qquad (2.19)$$

Differentiating (2.15) and solving for the set of values $H_1'(1|k)$, $k = 1, 2, \ldots$, the average size of the vulnerable cluster can be written in a form of the following matrix expression:

$$S = \mathbf{1}'\mathbf{p} + \langle k \rangle \zeta' \mathbf{M^{-1}} \zeta \qquad (2.20)$$

where $\mathbf{1}$ is the vector of 1's, $\mathbf{p} = \{p(k)\}_{k=1}^{\overline{K}}$ and $\zeta = \{\zeta(k)\}_{k=1}^{\overline{K}}$ are the vectors of degree frequencies for a randomly chosen and neighboring node, respectively, and $\mathbf{M}$ is the matrix of elements $\{m_{k,k'}\}_{k,k'=1}^{\overline{K}}$ such that

$$m_{k,k'} = k\xi(k, k') - \zeta(k')\delta_{k,k'}, \quad k, k' = 1, 2, \ldots \qquad (2.21)$$

Expression (2.20) diverges when $\det(\mathbf{M}) = 0$. This condition marks the *phase transition* at which a giant vulnerable component first appears in the network. By studying the behavior of (2.20) close to the transition, where S must be large and positive in the absence of a giant component, one can show that a giant vulnerable component exists in the network when $\det(\mathbf{M}) > 0$.[20]

So, in the framework of a given model, when $\det(\mathbf{M}) < 0$, the vulnerable clusters in the banking system are small and contagion stops quickly. However, if $\det(\mathbf{M}) \geq 0$, a giant vulnerable cluster arises and occupies a finite fraction of the network. The next section presents the analytical base for computing the relative, or *fractional* size of the giant vulnerable component. I employ this analysis later, in section 2.3.1, to compare the fractional size of the giant vulnerable component across tiered and other banking networks, so as to compare the susceptibility to risk and the scope of global crises in various types of banking structures.

Relative size of the giant vulnerable component

When a giant vulnerable component exists in the financial network, it is straightforward to assess its relative size, ω, – the fraction of banks in the network that belong to the giant vulnerable component. To find ω I first define $1 - \widehat{\omega}_k$ – the probability that a randomly selected link outgoing from a node (k, k) does *not* lead to the giant vulnerable component. This means that the endnode of this randomly selected link is either a safe bank or a vulnerable bank that does not belong to the giant vulnerable component. By

---

[20]See Newman (2002a) and Molloy and Reed (1995) for more details.

consistency, in the latter case none of the outgoing links of the endnode bank leads to the giant vulnerable component. This results in the following set of equations:[21]

$$1 - \widehat{\omega}_k = \left(1 - \sum_{k'=1}^{\overline{K}} p(k'|k)\right) + \sum_{k'=1}^{\overline{K}} p(k'|k)\big(1 - \widehat{\omega}_{k'}\big)^{k'}, \quad k = 1, 2, \ldots \tag{2.22}$$

Using the definition of $p(k'|k)$ in (2.11), (2.22) can be written as

$$1 - \widehat{\omega}_k = \left(1 - \frac{1}{\zeta(k)} \sum_{k'=1}^{\overline{K}} \xi(k, k')\right) + \frac{1}{\zeta(k)} \sum_{k'=1}^{\overline{K}} \xi(k, k')\big(1 - \widehat{\omega}_{k'}\big)^{k'}, \quad k = 1, 2, \ldots \tag{2.23}$$

Having obtained $\{\widehat{\omega}_k\}_{k=1}^{\overline{K}}$ from (2.23), let us now determine the probability ω that a randomly selected bank belongs to the giant vulnerable component, that is, the fractional size of the component. The probability $(1 - \omega)$ that a randomly selected node does *not* belong to the giant vulnerable component reflects the possibility of two events. First, a randomly selected bank may be safe and hence does not span even a vulnerable singleton. Alternatively, the bank may be vulnerable but such that none of its outgoing links leads to the giant vulnerable component. The two possibilities sum up to:

$$1 - \omega = \left(1 - \sum_{k=1}^{\overline{K}} p(k)\right) + \sum_{k=1}^{\overline{K}} p(k)\big(1 - \widehat{\omega}_k\big)^k, \quad k = 1, 2, \ldots \tag{2.24}$$

As with equations (2.15), (2.16), it is usually not possible to solve for ω in closed form, but the approximate solution can be found by numerical iteration from a suitable set of starting values for $\{1 - \widehat{\omega}_k\}_{k=1}^{\overline{K}}$. Section 2.3.1 presents the results of the numerical analysis to compare the phase transition thresholds and the sizes of the giant vulnerable components in assortative, disassortative and neutral networks.

2.3 Results

2.3.1 The giant vulnerable component and optimal bail-out strategy in assortative, disassortative and neutral banking networks

The numerical analysis is conducted following Newman (2002a). Let us consider the symmetric binomial form of the joint probability distribution for the degrees of the

---

[21]The first term in brackets on the right-hand side of (2.22) is the probability that a randomly chosen edge emanating from a node (k, k) leads to a safe bank.

neighboring nodes:

$$\xi(k, k') = Ne^{-(k+k')/\nu} \left[\binom{k + k'}{k} \alpha^k \beta^{k'} + \binom{k + k'}{k'} \alpha^{k'} \beta^k \right] \qquad (2.25)$$

where $\alpha + \beta = 1$, $\nu > 0$, and $N = \frac{1}{2}(1 - e^{-1/\nu})$ is a normalizing constant. The choice of this distribution function is explained by analytical tractability. Moreover, as argued by Newman (2002a), the behavior of this distribution is also quite natural: the distribution of the sum of the degrees, $k + k'$, decreases as a simple exponential, while that sum is distributed between the two neighboring nodes binomially. The parameter α controls the assortative, disassortative or neutral mixing in the network. From equation (2.10), the correlation coefficient of the degrees of the neighboring nodes, r, is equal to

$$r = \frac{8\alpha\beta - 1}{2e^{1/\nu} - 1 + 2(\alpha - \beta)^2} \qquad (2.26)$$

The value of r can be positive or negative, passing through zero when $\alpha = \frac{1}{2} \mp \frac{1}{4}\sqrt{2} \approx 0.1464$ or 0.8536.

Using the specification (2.25) of the joint degree distribution of the neighboring nodes, and equations (2.23), (2.24), I calculate numerically the fractional size of the giant vulnerable component, ω, for $\alpha = 0.05$, where the network is disassortative, $\alpha = 0.1464$, where it is neutral, and $\alpha = 0.5$, where it is assortative. Figure 2.1 demonstrates the results, showing the size of the giant vulnerable component as a function of the degree cutoff parameter $\overline{K}$. The three panels of the figure represent cases of different values of the degree scale parameter ν, where $1/\nu$ measures the rate at which the probability of the aggregate degree of neighboring nodes, $k + k'$, declines as $k + k'$ becomes larger.

When $\overline{K}$ increases, more and more banks in the network are characterized as vulnerable and at some point, the giant vulnerable component is formed and a risk of global financial breakdown is realized. Two important outcomes are suggested by Figure 2.1. First, the phase transition, the instance at which a giant vulnerable component arises, shifts to the left as the network becomes more assortative. That is, the giant vulnerable component arises more easily and the risk of "global" crisis in the banking system is higher if poorly-connected banks preferentially associate with other poorly-connected banks. Conversely, the disassortative/tiered banking systems show more resilience to global defaults. For example, when $\nu = 10$ or $\nu = 15$ and $5 \leq \overline{K} \leq 6$, as it is suggested by the empirical evidence and by the calibration results in Gai and Kapadia (2009), no risk of system-wide contagion exists in the disassortative/tiered banking network but in the assortative network, the risk of global default is positive. Second, for the same value of degree cutoff $\overline{K}$, at least in the plausible range of values, the size of the giant vulnerable component as a fraction of the whole banking network, is larger in the assortatively
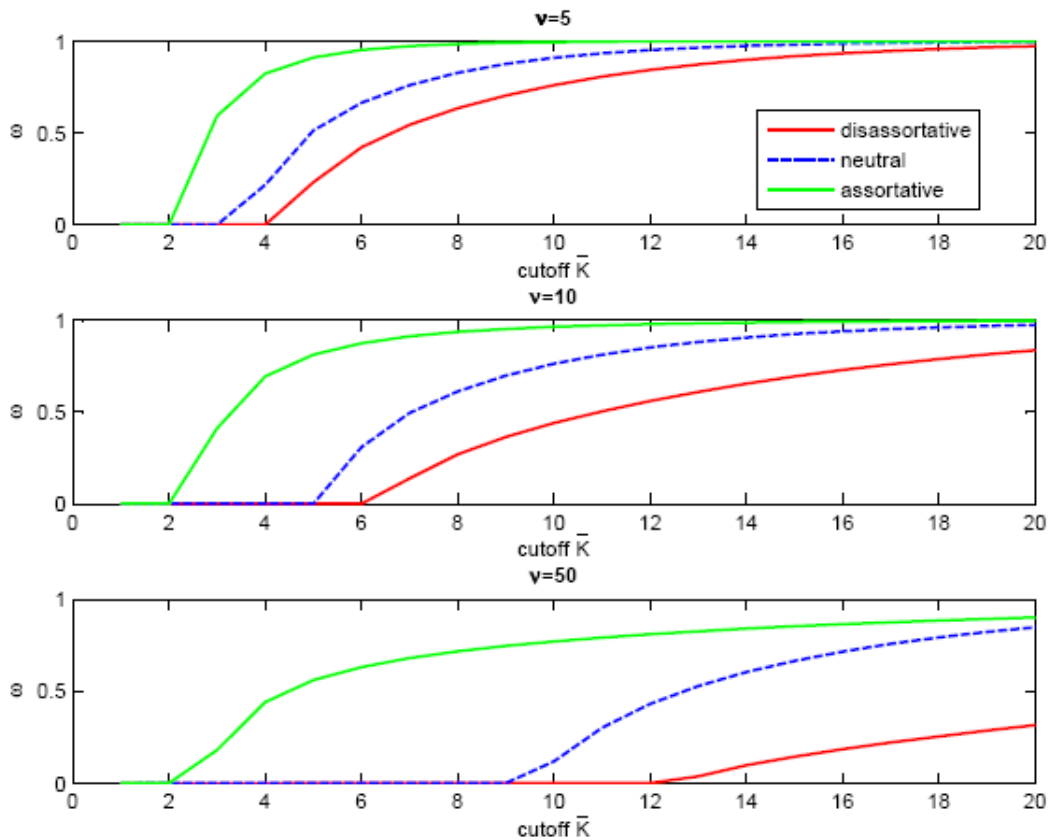
FIGURE 2.1: Fractional size of the giant vulnerable component as a function of $\overline{K}$ for the joint degree distribution of the neighboring nodes given in (2.25).

mixed network. This means that as soon as contagious defaults become system-wide (the giant vulnerable component emerges), the fraction of defaulting institutions in the banking network, or the scope of the crisis, is larger if the network is assortative. As a result, in the disassortative/tiered banking structure, both the risk of systemic crisis and the scope of the crisis are lower than in the other types of structures.

These results are intuitively reasonable. In a disassortative banking network, high-degree banks are broadly distributed over the network and therefore, presumably form links on many paths between other banks. This implies that with high probability an initial shock hitting a random bank reaches a high-degree bank in a small number of "steps". Since high-degree banks are relatively resilient to neighbors' defaults, the shock gets absorbed at that bank and does not spread any further. Therefore, both the risk of systemic defaults and the scale of these defaults in the disassortative network are lower than in the other structures. In fact, this simple intuition suggests that the finding of relative resilience to shocks of disassortative networks, derived here for the specific distribution in (2.25), can be generalized to a wider range of degree distributions, that is, to a larger class of assortative, disassortative and neutral networks. However, the formal investigation of this question is beyond the scope of this paper.

Furthermore, the comparison of the susceptibility to crises of assortative, disassortative and neutral banking networks allows for the evaluation of the relative effectiveness and costs of the optimal government bail-out strategy in these networks. Here I focus on *targeted* bail-outs, when the government first rescues the defaulting institutions that have a particularly large number of interbank connections and therefore, represent the highest risk in terms of the spread of shocks through the network.[22] Let $K_{\max}$ be the threshold degree which satisfies the condition that bailing-out all defaulting banks with in- and out-degree higher than this threshold guarantees that no global crisis will emerge (with probability 1 the giant vulnerable component will not arise). Then the banks with in- and out-degree lower or equal to $K_{\max}$ either do not default or default but do not need to be rescued for the global stability of the system to be preserved. In this sense, $K_{\max}$ represents the highest in- and out-degree among defaulting institutions whose failure does not endanger the stability of the banking system.

The threshold $K_{\max}$ determines *a priori optimal bail-out strategy* of the government in the following sense. Given only the degree distribution in the banking network, it minimizes the number of targeted bail-outs (with in- and out-degree higher than $K_{\max}$) subject to the constraint that the system stability is preserved. Any higher threshold degree would not guarantee the stability of the system, whereas a lower threshold would induce unnecessary bail-out costs.

The definition of $K_{\max}$ implies that it is nothing but the point on the scale of possible values of $\overline{K}$ at which the giant vulnerable component first forms. It can, therefore, be found from the conditions which determine the phase transition. Furthermore, for any degree cutoff $\overline{K}$ in the banking system, the value $K_{\max} \geq \overline{K}$ implies that no bail-outs are required since by definition, all banks with in- and out-degree larger or equal to $\overline{K}$ are safe. Conversely, if $1 \leq K_{\max} < \overline{K}$, banks with degrees in the range between $K_{\max}$ and $\overline{K}$ are vulnerable and need to be rescued upon default since their failure endangers the stability of the system.

In view of the aforesaid, Figure 2.1 suggests that in the assortative and neutral banking networks targeted bail-outs are needed for lower values of $K_{\max}$ than in the disassortative network. For example, when $\nu = 5$, the threshold degree $K_{\max} = 2$ if the network is assortative, 3 if it is neutral and 4 if it is disassortative. Moreover, when $\nu = 10$ or $\nu = 50$ and $5 \leq \overline{K} \leq 6$, no bail-outs are needed to preserve the stability of the disassortative banking network, while the assortative network requires rescues of all failing banks with in- and out-degree exceeding 2.

---

[22]Notice that in the framework of this model, bailing-out defaulting banks simply means "transforming" these initially vulnerable institutions into safe institutions. After having been rescued by the government, they become "immune" from defaults and do not transmit shocks to their neighbors in the banking network.

2.3.2 Resilience to systemic breakdown and targeted bail-outs in tiered banking system: scale-free networks

Let us now consider the alternative way of representing the tiered banking network. Suppose that there are *no degree correlations* between neighboring nodes but the degree distribution is specified to capture (i) the tendency of highly-connected nodes to be connected to poorly-connected ones and vice versa, and (ii) the numerical superiority of the low-degree nodes over the high-degree ones. Both characteristics are inherent in the power-law, or scale-free, degree distribution defined by:

$$p(k) = (\lambda - 1)k^{-\lambda}, \quad k = 1, 2, \ldots \quad \text{and} \quad 2 < \lambda < 3 \tag{2.27}$$

where, as before, $p(k)$ is the probability that a randomly chosen node has in- and out-degrees each equal to k, λ governs the rate at which the probability decays with connectivity and the assumption $2 < \lambda < 3$ means that the degree distribution is so broad that it displays unbounded second moments but still possesses a well-defined average degree. Arguably, it is more reasonable to think of a banking network as being generated by the scale-free degree distribution with such low values of λ, since this implies that $p(k)$ does not decline "too fast" with k and the substantial share of banks in the network have more than one in- and out-going connection. This assumption is supported by the finding of Boss et al. (2004) which suggests that the power-law degree distribution of the Austrian interbank market has exponent parameter $\lambda = 2.01$.

In the framework of this model, λ has an important interpretation as the inverse of a tiering level in the scale-free network. Indeed, as the value of λ declines, the fraction of highly-connected banks in the network increases and so does the probability that poorly-connected banks link with highly-connected ones rather than with each other. This implies that by varying the value of λ, one can study and compare the risks of systemic crisis and optimal bail-out strategy of the government in scale-free banking networks with different levels of tiering. Below, I consider these questions in detail. Specifically, I find the closed-form expression for the threshold degree $K_{\max}$ and study the functional dependence of $K_{\max}$ on parameter λ. Hereafter, for simplicity, I regard the degree of each node, k, as a continuous variable and the probability $p(k)$ in (2.27) as a continuous density function.

Generating functions
Given the assumption of absent degree correlations between neighboring nodes and the continuous framework of the degree distribution, the generating functions H_1 and H_0

in (2.15) and (2.16) can be written as follows:

$$
\begin{aligned}
H_1(y) &= \left(1 - \int_1^{\overline{K}} \zeta(k)dk\right) + y \int_1^{\overline{K}} \zeta(k)\big(H_1(y)\big)^k dk \\
&= \left(1 - \frac{1}{\langle k \rangle} \int_1^{\overline{K}} p(k)k\,dk\right) + y\frac{1}{\langle k \rangle} \int_1^{\overline{K}} p(k)k\big(H_1(y)\big)^k dk, \qquad (2.28)
\end{aligned}
$$

$$
H_0(y) = \left(1 - \int_1^{\overline{K}} p(k)dk\right) + y \int_1^{\overline{K}} p(k)\big(H_1(y)\big)^k dk \qquad (2.29)
$$

where the second equality in (2.28) uses the definition of $\zeta(k)$ in (2.7). Both expressions can be formulated more compactly by means of the other two generating functions: $G_0(y)$, for the number of links leaving a randomly chosen vulnerable bank, and $G_1(y)$, for the number of links leaving a vulnerable bank reached by following a randomly chosen incoming link. Since the degree distribution of a randomly selected bank is given by $p(k)$, and the degree distribution of a *neighboring* bank reached by following a randomly chosen link is $\zeta(k) = kp(k)/\langle k \rangle$, $G_0(y)$ and $G_1(y)$ can be defined as:

$$
G_0(y) = \int_1^{\overline{K}} p(k)y^k dk, \qquad (2.30)
$$

$$
G_1(y) = \frac{1}{\langle k \rangle} \int_1^{\overline{K}} kp(k)y^k dk \qquad (2.31)
$$

Functions $G_0(y)$ and $G_1(y)$ allow $H_0(y)$ and $H_1(y)$ to be rewritten in the following form:

$$
H_1(y) = (1 - G_1(1)) + yG_1(H_1(y)), \qquad (2.32)
$$

$$
H_0(y) = (1 - G_0(1)) + yG_0(H_1(y)), \qquad (2.33)
$$

Now, as in (2.17), the average size of the vulnerable cluster, S, can be computed as:

$$
S = H_0'(1) = G_0(H_1(1)) + G_0'(H_1(1))H_1'(1) = G_0(1) + G_0'(1)H_1'(1) \qquad (2.34)
$$

where $H_1(1) = 1$ due to the basic property of generating functions. From (2.32) we know that

$$
H_1'(1) = \frac{G_1(1)}{1 - G_1'(1)} \qquad (2.35)
$$

Hence, the average size of the vulnerable cluster becomes:

$$
S = G_0(1) + \frac{G_0'(1)G_1(1)}{1 - G_1'(1)} \qquad (2.36)
$$

Phase transition and threshold degree $K_{\max}$

The term $G_1'(1)$ on the right-hand side of (2.36) is the average out-degree of a vulnerable

first neighbor. If this quantity is less than one, all vulnerable clusters in the network are small and contagion dies out quickly. But as $G'_1(1) \nearrow 1$, the average size of the vulnerable cluster, S, increases unboundedly. It diverges when $G'_1(1) = 1$. It is at this point that the giant vulnerable component, whose size scales linearly with the size of the whole network, first forms. The threshold degree $K_{\max}$ for vulnerable banks, above which systemic risk is realized and massive contagion is likely, can be determined from this equation. Using the definition of $G_1(y)$ in (2.31) and the definition of $p(k)$ in (2.27), we obtain:

$$G'_1(1) = \frac{1}{\langle k \rangle} \int_1^{K_{\max}} k^2 p(k) dk = \frac{\lambda - 1}{\langle k \rangle} \int_1^{K_{\max}} k^{2-\lambda} dk = \frac{\lambda - 1}{\langle k \rangle} \left(K_{\max}^{3-\lambda} - 1 \right) \quad (2.37)$$

Since $\langle k \rangle = (\lambda - 1)/(\lambda - 2)$, condition $G'_1(1) = 1$ leads to the closed-form expression for $K_{\max}$.

Proposition 2.1. *In the scale-free network, the threshold degree $K_{\max}$ at which the giant vulnerable component first forms is given by*

$$K_{\max} = \left[\frac{1}{\lambda - 2} \right]^{\frac{1}{3-\lambda}} \quad (2.38)$$

$K_{\max}$ is monotonically decreasing in λ for all $2 < \lambda < 3$

So, the critical level for the formation of the giant vulnerable component and the targeted bail-out strategy of the government are determined by the value of λ, the level of tiering in the banking system. Furthermore, $K_{\max}$ is decreasing in λ, that is, it is increasing in the level of tiering.[23] This functional dependence of $K_{\max}$ on λ is illustrated with Figure 2.2.

For λ sufficiently close to 2, the threshold degree $K_{\max}$ is large but it declines sharply with λ. For example, when λ is 2.1, $K_{\max}$ is approximately 13 but it is only around 4 for λ equal to 2.4 and around 3 for λ equal to 2.7. This finding suggests that the resilience to global crisis of the scale-free banking system with high level of tiering (λ is close to 2) is very high but declines strongly when the level of tiering decreases (λ becomes large). Specifically, since all banks in the network with in- and out-degree greater or equal to $\overline{K}$ are safe and $\overline{K}$ is estimated to be in the range between 5 and 6 for real-world banking systems, no system-wide contagion occurs and hence, no bail-outs are needed in the highly-tiered scale-free network with λ close to 2.[24] In contrast, for the scale-free network with lower level of tiering, when λ is around 2.4 or larger, the

---

[23]This simple comparative statics result is derived in the Appendix.

[24]This result agrees with the conclusion of the empirical studies by Boss et al. (2004, 2004b). For the scale-free Austrian banking system with $\lambda = 2.01$, they report that "the banking system is very stable and default events that could be classified as a "systemic crisis" are unlikely" (Boss et al. (2004b)).
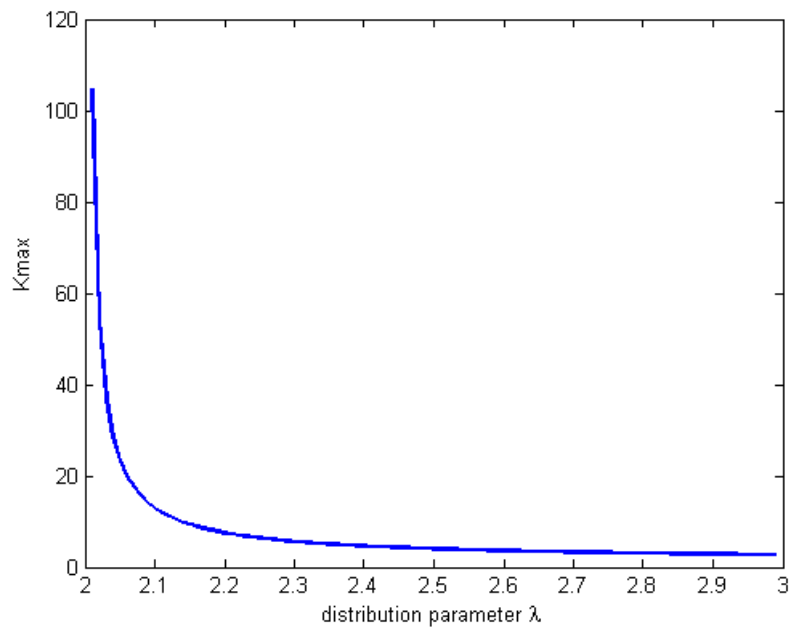
FIGURE 2.2: Threshold degree $K_{\max}$ as a function of λ.

risk of global contagion exists unless the government rescues all defaulting banks with in- and out-degree exceeding $K_{\max}$. This means that the required bail-outs comprise all defaulting institutions with in- and out-degree higher than 4 if λ is 2.4, higher than 3 if λ is 2.7 and higher than 2 if λ is arbitrarily close to 3.[25]

These conclusions are confirmed by Figure 2.3, which shows the threshold $K_{\max}$ at which the giant vulnerable component first forms, and the size of the giant vulnerable component beyond this threshold, ω, for three values of λ.[26]

Intuitively, a decline in resilience of the scale-free network to shocks as the level of tiering lowers (parameter λ grows) can be explained as follows. When the level of tiering is high (λ is small), the fraction of highly-connected banks in the network is relatively large and so is the number of paths between other banks in the network which "pass" through the

---

[25]Notice, that all these values of $K_{\max}$ in the scale-free network with λ equal to 2.4, 2.7 and λ arbitrarily close to 3 exceed the mean degree, $\langle k \rangle$, of the scale-free distribution with the corresponding parameter λ.

[26]The size of the giant vulnerable component is computed numerically from the equations analogous to (2.23) and (2.24). Under the assumption of absent degree correlations and continuity of the degree distribution, (2.23) and (2.24) become:

$$1 - \widehat{\omega} \quad = \quad \left(1 - \frac{1}{\langle k \rangle} \int_1^{\overline{K}} p(k)k\,dk\right) + \frac{1}{\langle k \rangle} \int_1^{\overline{K}} p(k)k\big(1 - \widehat{\omega}\big)^k dk, \tag{2.39}$$

$$1 - \omega \quad = \quad \left(1 - \int_1^{\overline{K}} p(k)\,dk\right) + \int_1^{\overline{K}} p(k)\big(p(k)(1 - \widehat{\omega})\big)^k dk \tag{2.40}$$

As before, $\widehat{\omega}$ represents the fraction of links leading to the giant vulnerable component and ω is the fraction of nodes in the giant vulnerable component.
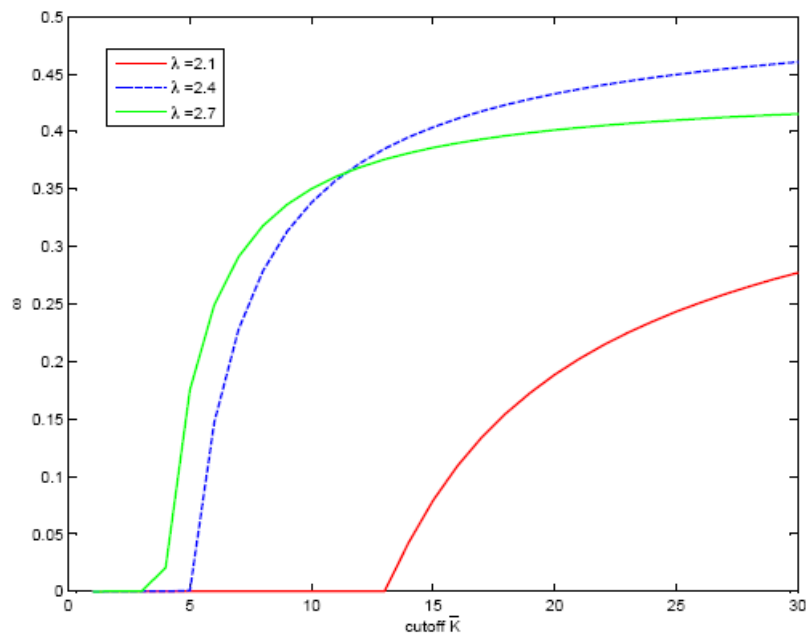
FIGURE 2.3: Fractional size of the giant vulnerable component as a function of $\overline{K}$, for $\lambda = 2.1$, $\lambda = 2.4$, and $\lambda = 2.7$.

highly-connected banks. As a result, in a scale-free network with higher level of tiering, initial shocks in any part of the network reach highly-connected banks and are absorbed by these banks "faster" than in a less tiered network. Therefore, more tiered systems are more resilient to massive defaults.

Figures 2.2 and 2.3 shed some light on the empirical evidence about the susceptibility of tiered banking networks to systemic shocks. They suggest that at least in the tiered banking systems represented by the scale-free networks, the extent of resilience to shocks is determined by the level of tiering. While a highly-tiered scale-free network is extremely resilient to systemic defaults, less tiered structures are fragile and financial contagion propagates through such structures easily.

Finally, given the gap between $K_{\max}$ and $\overline{K}$, I calculate the *maximal expected share* of banks, ρ, which *may need to be rescued* to preserve the stability of the scale-free banking network. This maximal expected share of rescued banks, or the upper bound for the percentage of bail-outs in the system, corresponds to the situation when *all* vulnerable banks with in- and out-degree exceeding $K_{\max}$ actually default.

Proposition 2.2. *In the scale-free network, the maximal expected share of banks which may need to be rescued to preserve the stability of the system is equal to*

$$\rho = \int_{K_{\max}}^{\overline{K}} p(k)dk = \left[\frac{1}{K_{\max}}\right]^{\lambda-1} - \left[\frac{1}{\overline{K}}\right]^{\lambda-1} \tag{2.41}$$

ρ is monotonically increasing in λ for all $2 < \lambda < 3$

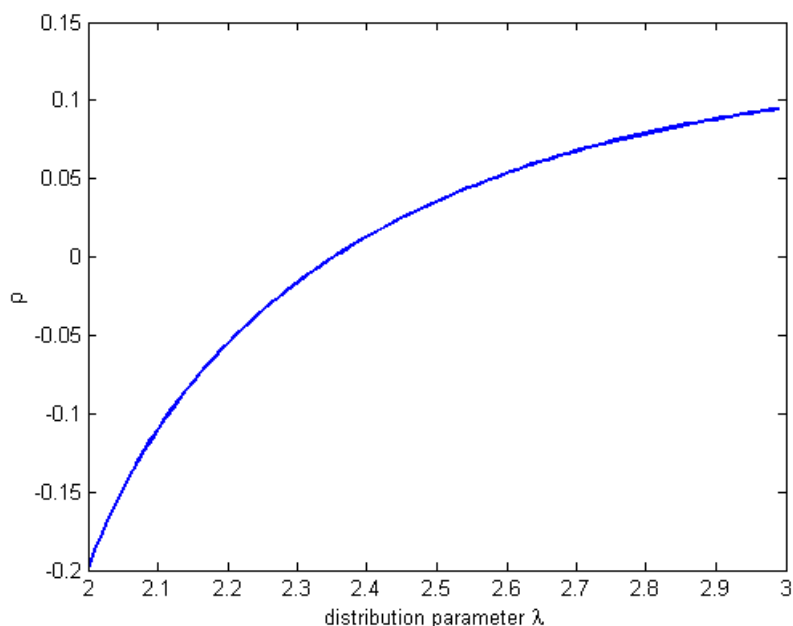So, the upper bound for the percentage of bail-outs in the scale-free networks is increasing in λ for any $2 < \lambda < 3$, or equivalently, it is decreasing in the level of tiering.[27]

FIGURE 2.4: The upper bound for the percentage of bail-outs in scale-free networks as a function of λ, for $\overline{K} = 5$. Negative values of ρ correspond to $K_{max} < \overline{K}$.

For example, when $\overline{K}$ is 5, ρ is equal to 1.3% of all banks if λ is 2.4, and 6.8% of all banks if λ is 2.7. Thus, the total number of bail-outs needed can be as high as 6.8% of the system for λ equal to 2.7 and at the other extreme, bail-outs may not be needed at all if λ is sufficiently close to 2. As before, when the system is more tiered (λ is smaller), highly connected banks form a larger fraction of the system and the number of targeted bail-outs sufficient to stop the massive contagion is smaller.

2.4 Policy implications and conclusions

This paper develops a model of contagion in a banking system and studies the effects of tiering in the system on the risk and potential impact of system-wide defaults. High policy relevance of this issue and controversial empirical findings provide strong motivation for this research.

---

[27]The short proof of this result is provided in the Appendix.

The banking system is constructed by a random directed network, where nodes represent banks and links represent claims and obligations of banks to each other. In this framework, the tiered banking system is modelled in two ways: first, by a disassortative network, displaying negative degree correlations between neighboring nodes, and then by a scale-free network. Using the first modelling approach, I compare the resilience to systemic defaults and optimal bail-out strategy of the government in the tiered disassortative network with those in other types of networks: the assortative network, displaying positive degree correlations, and the neutral network, displaying no correlations. Subsequently, using the second approach, I concentrate on the tiered banking systems displaying various levels of tiering. I argue that the level of tiering in the scale-free network can be approximated by the inverse of the exponent parameter. By changing this parameter, I evaluate the effects of the level of tiering on the susceptibility of the banking system to shocks and on the maximal expected number of targeted bail-outs.

The key feature of the model, which highlights the importance of network structure in determining the spread of contagious defaults, is the counteracting effects of bank connectivity. While greater connectivity increases the spread of contagion in the banking network, it also improves risk sharing among neighboring banks and thereby reduces the susceptibility of banks to defaults. These opposing effects of risk sharing and risk spreading interact differently in varying structures. As a result, the resilience of a banking system to shocks and the optimal number of bank rescues depend on the features of the underlying degree distribution.

Specifically, for the degree distribution which generates assortative, disassortative, and neutral networks, I find that the disassortative/tiered banking network is more resilient to shocks and in the event of a crisis, shows a lower frequency of failures than other types of structures. Consequently, the threshold degree above which the defaulting banks endanger the stability of the system and require government assistance is higher in the disassortative network, so that the overall number of bail-outs and the associated bail-out costs are expected to be lower.

Further, for the scale-free banking system, I find that the higher level of tiering implies higher resilience to shocks and, in the worst-case scenario, lower number of bail-outs. In particular, when the level of tiering is relatively low, the required number of bail-outs may be as high as 6.8% of the system, whereas if the level of tiering is higher, bail-outs may not be needed at all.

These findings suggest not only the advantages of tiering in the banking network over other types of network organization, but also the advantages of relatively high levels of tiering regardless.

These insights provide the basis for specific policy recommendations. First, the relatively high resilience of the tiered system to contagion and the minor extent of government intervention needed in order to reduce systemic risk imply that the formation of a more highly tiered system of interbank relations should be among the priorities of the central bank or the government. Secondly, the *optimal* level of targeted bail-outs which guarantees the stability of the system at minimal cost should be chosen by explicitly taking into account the features of the underlying banking structure. Specifically, if the degree distribution generating the network of interbank connections is scale-free, optimal targeted bail-outs are predetermined by the exponent parameter, or the level of tiering in the scale-free network. The number of these bail-outs declines as the level of tiering increases. Finally, insights about the role of bank connectivity as a shock absorber or a shock amplifier suggest that the close monitoring by the regulator is required for banks whose connectivity is "risky": the smaller the number of incoming links (low level of risk diversification) and the larger the number of outgoing links (high number of obligations to other banks), the higher the potential of a bank to be a key transmitter of shocks in the system.[28]

The model and results presented in this paper suggest some directions for future research. An interesting extension of the paper would be to simulate the model for a large banking system, using real balance sheets for all banks and calibrating the joint degree distribution to match the observed data. Alternatively, one could think of endogenizing the formation of the banking network so as to demonstrate and explain the incentives of banks in the real world to settle within the tiered banking systems. The framework of the network formation model could loosen the assumprions of the present setting, allowing for distinctions in interbank asset positions of banks and for the possibility of clusters/cycles in the banking network. In addition, to reflect the difference in conditions for borrowing and lending in the interbank market, it could also differentiate between the costs of formation of incoming and outgoing links. These strands of research would add realism to the model and provide new, potentially valuable, insights.

2.5   Appendix

Proof of Proposition 2.1

---

[28]If the number of incoming and outgoing links of a bank is the same, as it is assumed in the model, then the highest risk of contagion transmission is posed by the institutions with somewhat average degree: high-degree banks absorb shocks and therefore, are "immune" against contagion, whereas low-degree banks may default but have only a limited scope for spreading the shock further.

K_{max} is monotonically decreasing in λ as soon as its derivative with respect to λ is strictly negative for all $2 < \lambda < 3$. To calculate the derivative of K_{max}, notice that from (2.38):

$$\ln(K_{max}) = \frac{1}{3 - \lambda} \ln\left(\frac{1}{\lambda - 2}\right)$$

so that

$$\frac{1}{K_{max}} \cdot \frac{dK_{max}}{d\lambda} = \frac{1}{(3 - \lambda)^2} \ln\left(\frac{1}{\lambda - 2}\right) - \frac{1}{3 - \lambda} \cdot \frac{1}{\lambda - 2}$$

From this expression it follows that

$$\frac{dK_{max}}{d\lambda} = K_{max} \cdot \frac{1}{3 - \lambda} \cdot \frac{-(\lambda - 2)\ln(\lambda - 2) - (3 - \lambda)}{(3 - \lambda)(\lambda - 2)} \tag{2.42}$$

Since $2 < \lambda < 3$, this derivative is negative as soon as $f(\lambda) = -(\lambda - 2)\ln(\lambda - 2) - (3 - \lambda) < 0$. Notice that $f(\lambda)$ is monotonically increasing in λ for any $2 < \lambda < 3$. Indeed:

$$\frac{df(\lambda)}{d\lambda} = -\ln(\lambda - 2) > 0 \quad \forall 2 < \lambda < 3$$

Hence, for all $2 < \lambda < 3$, $f(\lambda) < f(3) = 0$. Therefore, $dK_{max}/d\lambda$ in (2.42) is strictly negative for any $2 < \lambda < 3$.

∎

Proof of Proposition 2.2

ρ is monotonically increasing in λ if its derivative with respect to λ is strictly positive for all $2 < \lambda < 3$. I differentiate with respect to λ both ratios on the right-hand side of (2.41). Denote the first ratio by y:

$$y = \left[\frac{1}{K_{max}}\right]^{\lambda - 1} = (\lambda - 2)^{\frac{\lambda - 1}{3 - \lambda}}$$

Then

$$\ln(y) = \frac{\lambda - 1}{3 - \lambda} \ln(\lambda - 2)$$

and

$$\frac{dy}{d\lambda} = (\lambda - 2)^{\frac{\lambda - 1}{3 - \lambda}} \cdot \left(\frac{2}{(3 - \lambda)^2} \ln(\lambda - 2) + \frac{\lambda - 1}{(3 - \lambda)(\lambda - 2)}\right)$$

Therefore,

$$\begin{aligned}
\frac{d\rho}{d\lambda} &= (\lambda - 2)^{\frac{\lambda - 1}{3 - \lambda}} \cdot \left(\frac{2}{(3 - \lambda)^2} \ln(\lambda - 2) + \frac{\lambda - 1}{(3 - \lambda)(\lambda - 2)}\right) - \left(\frac{1}{\overline{K}}\right)^{\lambda - 1} \ln\left(\frac{1}{\overline{K}}\right) = \\
&= (\lambda - 2)^{\frac{\lambda - 1}{3 - \lambda}} \cdot \frac{1}{3 - \lambda} \cdot \frac{2(\lambda - 2)\ln(\lambda - 2) + (\lambda - 1)(3 - \lambda)}{(3 - \lambda)(\lambda - 2)} - \left(\frac{1}{\overline{K}}\right)^{\lambda - 1} \ln\left(\frac{1}{\overline{K}}\right)
\end{aligned}$$

Since $2 < \lambda < 3$ and $\overline{K} \geq 1$, this derivative is positive as soon as $f(\lambda) = 2(\lambda - 2)\ln(\lambda - 2) + (\lambda - 1)(3 - \lambda) > 0$. Notice that $f(\lambda)$ is monotonically decreasing in λ for any $2 < \lambda < 3$. Indeed:

$$\frac{df(\lambda)}{d\lambda} = 2(\ln(\lambda - 2) - \lambda + 3) < 0 \quad \forall 2 < \lambda < 3$$

where $2(\ln(\lambda - 2) - \lambda + 3) < 0$ due to the fact that $2(\ln(\lambda - 2) - \lambda + 3)$ is monotonically increasing in $\lambda \ \forall 2 < \lambda < 3$ and its maximal value at $\lambda = 3$ is equal to 0.

So, $f(\lambda)$ is decreasing in λ for all $2 < \lambda < 3$. Therefore, for $2 < \lambda < 3$, $f(\lambda) > f(3) = 0$. This implies that $d\rho/d\lambda$ is strictly positive for any $2 < \lambda < 3$.

∎

2.6 Bibliography

Allen F. and A. Babus (2008) *"Networks in Finance,"* forthcoming in Network-Based Strategies and Competencies Kleindorfer P. and J. Wind (eds.), Wharton school publishing.

Allen F. and D. Gale (2000) *"Financial Contagion,"* Journal of Political Economy, 108: 1-33.

Acharya V. and T. Yorulmazer (2007) *"Cash-in-the-Market Pricing and Optimal Bank Bailout Policy,"* forthcoming in Review of Financial Studies.

Bhattachrya S. and D. Gale (1987) *"Preference Shocks, Liquidity, and Central Bank Policy,"* New Approaches to Monetary Economics: Proceedings of the Second International Symposium in Economic Theory and Econometrics. Cambridge University Press, Cambridge, New York, and Melbourne: 69-88.

Boss M., Elsinger H., Summer M., and S. Thurner (2004) *"The Network Topology of the Interbank Market,"* Quantitative Finance 4: 1-8.

Boss M., Elsinger H., Summer M., and S. Thurner (2004a) *"An Empirical Analysis of the Network Structure of the Austrian Interbank Market,"* Oesterreichische Nationalbank Financial Stability Report, June: 77-87.

Boss M., Summer M., and S. Thurner (2004b) *"Contagion Flow Through Banking Network,"* Lecture Notes in Computer Science 3038 (eds. M. Bubak, G. D. v Albada, P. M. A. Sloot, and J. J. Dongarra (Berlin: Springer)): 10701077

Brusco S. and F. Castiglionesi (2007) *"Liquidity Coinsurance, Moral Hazard, and Financial Contagion,"* Journal of Finance, 62: 2275-2302.

Callaway D. S., Newman M. E. J., Strogatz S. H., and D. J. Watts (2000) *"Network Robustness and Fragility: Percolation on Random Graphs,"* Physical Review Letters, 85: 5468-5471.

Castiglionesi F. and N. Navarro (2007) *"Optimal fragile financial networks,* Tilburg University Discussion Paper 2007-100.

Cohen R., Erez K., Ben-Avraham D., and S. Havlin (2001) *"Breakdown of the Internet Under Intentional Attack,"* Physical Review Letters, 86: 3682-3685.

Degryse H. A. and G. Nguyen (2006) *"Interbank Exposures: An Empirical Examination of Contagion Risk in the Belgian Banking System,"* Discussion Paper 2006-016, Tilburg University, Tilburg Law and Economic Center.

Elsinger H., Lehar A. and M. Summer (2006) *"Risk Assessment for Banking Systems,"* Management Science, 52: 1301-1314.

Flannery M. (1996) *"Financial Crises, Payments System Problems and Discount Window Lending,"* Journal of Money, Credit and Banking, 28: 804824.

Furfine C.H. (1999) *"Interbank Exposures: Quantifying the Risk of Contagion,"* BIS Working Paper No. 70, Basel.

Freixas X., Parigi B. and J.-C. Rochet (2000) *"Systemic Risk, Interbank Relations and Liquidity Provision by the Central Bank,"* Journal of Money, Credit and Banking, 32: 611-638.

Gai P. and S. Kapadia (2008) *"A Network Model of Super-Systemic Crises,"* Bank of England, forthcoming.

Gai P. and S. Kapadia (2009) *"Contagion in Financial Networks,"* Bank of England, forthcoming.

Galeotti A. and S. Goyal (2007) *"A theory of strategic diffusion,"* University of Essex Department of Economics Discussion Paper 635.

Haldane A. G. (2009) *"Rethinking the Financial Network,"* Bank of England, Speech delivered at the Financial Student Association, Amsterdam.

Harrison S., Lasaosa A., and M. Tudela (2005) *"Tiering in UK Payment Systems-Credit Risk Implications,"* Bank of England Financial Stability Review, December: 6372.

Jackson M. and L. Yariv (2007), *"Diffusion of behavior and equilibrium properties in network games,"* American Economic Review (97): 92-98.

Leitner Y. (2005), *"Financial networks: contagion, commitment and private sector bailouts,"* Journal of Finance (6): 2925-53.

Mistrulli P. (2007) *"Assessing Financial Contagion in the Interbank Market: Maximum Entropy versus Observed Interbank Lending Patterns,"* Bank of Italy Working Paper 641.

Molloy M. and B. Reed (1995) *"A Critical Point for Random Graphs With a Given Degree Sequence,"* Random Structures and Algorithms, 6: 161179.

Morris, S (2000) *"Contagion,"* Review of Economic Studies (67): 57-78.

Newman M. E. J., Strogatz S. H. and D. J. Watts (2001) *"Random Graphs with Arbitrary Degree Distributions and Their Applications,"* Physical Review E, 64, 026118.

Newman M. E. J. (2002a) *"Assortative Mixing in Networks,"* Physical Review Letters, 89, 208701.

Newman M. E. J. (2002b) *"The Spread of Epidemic Disease on Networks,"* Physical Review E, 66, 016128.

Nier E., Yang J., Yorulmazer T., and A. Alentorn (2007) *"Network Models and Financial Stability,"* Journal of Economic Dynamics and Control, 31: 2033-2060.

Rochet J.-C. and J. Tirole (1996) *"Interbank Lending and Systemic Risk,"* Journal of Money, Credit and Banking, 28 (4): 733762.

Sheldon G. and M. Maurer (1998) *"Interbank Lending and Systemic Risk: An Empirical Analysis of Switzerland,"* Swiss Journal of Economics and Statistics, 134 (4.2): 685704.

Strogatz S. H. (2001) *"Exploring Complex Networks,"* Nature, 410: 268-276.

Upper C. and A. Worms (2004) *"Estimating Bilateral Exposures in the German Interbank Market: Is There a Danger of Contagion?"* European Economic Review, 48: 827849.

Vega-Redondo F. (2007) *"Complex Social Networks,"* Cambridge University Press.

Watts D. (2002) *"A Simple Model of Global Cascades on Random Networks,"* Proceedings of the National Academy of Sciences, 99: 5766-5771.

Wells S. (2002) *"UK Interbank Exposures: Systemic Risk Implications,"* Bank of England Financial Stability Review, December: 175182.

Chapter 3

The Effects of Social Networks on Migration Decisions: Destination Choice and Local Variations in 19th Century Italian Emigration

3.1   Introduction

One of the puzzling cases in the history of international migration is the destination choice of Italian emigrants in the second half of the 19th century. Originally emigrants from Italy revealed strong preferences for countries of South America, especially Argentina and Brazil. As emigration from Italy grew, the prominence of the United States as a destination increased. Most of this shift occurred due to emigration from the southern part of Italy where emigration traditions to South America were weakest. However, despite the substantial wage gaps favoring North America over South America and despite the higher level of urbanization in North America, urban Northern Italians continued migrating to rural South America. Over time this resulted in an essential difference in Italian regional emigration; the Southern Italians migrated in large numbers to the United States, while the Northern Italians predominantly moved to South America.[1]

This case of 19th century Italian emigration hints at the potential importance of social connections between Italians and New World residents (natives and/or prior migrants)

---

[1]This apparent anomaly has attracted the attention of historians and migration economists. See, for example, Baily (1983), Klein (1983), Hatton and Williamson (1998).

in determining migration choices. In this paper I aim to shed more light on this issue. I examine whether the puzzling lack of migration from the north of Italy to the United States, despite the evident economic benefits of that move, and simultaneous massive migration from the south of Italy could have resulted from the *initial small difference* in the number of social contacts in the United States of the Northern and Southern Italians. While social links between the U. S. and the north of Italy were largely absent, the connections of the Southern Italians were formed by few pioneer migrants from the south of Italy.

Traditional migration literature, valuable in providing insights into major determinants of migration, lacks for an explanation of regional variations in emigration from Italy. The key determinants of international migration suggested by this literature are differentials in wages and employment conditions between countries as well as migration costs (Lewis (1954), Ranis and Fei (1961), Sjaastad (1962), Harris and Todaro (1970), Todaro (1969, 1976, 1989), Todaro and Maruszko (1987)). These factors seem to have carried a lot of weight in determining migration decisions of the Southern Italians. However, they were clearly of minor importance for destination choices of the Northern Italians.

Some other studies focus specifically on explaining regional differences in emigration. They identify local variations in degrees of economic development, urbanization and patterns of land tenure among the main drivers of such differences (Walker (1964), Lowell (1987), Hatton and Williamson (1998)). However, if differences in Italian regional emigration had been driven by these factors, the observed patterns would have actually been reversed: the urban Northern Italians would have favored the United States over rural South America, while the Southern Italians would have rather migrated to South America.

Another stream of migration literature, so-called *chain migration* literature, emphasizes the positive impact of the *stock of previous migrants* and the *lagged emigration rate* on continuation of migration (Böcker (1994), Boyd (1989), Gurak and Caces (1992), Massey et al. (1993), Faist (2000), Dahya (1973)).[2] This literature provides empirical evaluation of migration flows, using regressions with the size of migration flow or current emigration rate as the dependent variable and the stock of previous migrants or last-period emigration rate as an explanatory variable. The estimated positive and statistically significant impact of these explanatory variables is then interpreted as information provision and assistance, especially with regard to finding work and housing, which already settled

---

[2]Hatton and Williamson (1998) report that as many as 90 percent of immigrant arrivals to the US at the turn of the century were travelling to meet a friend or relative who had previously emigrated. Also their estimation results suggest that for each thousand of previous emigrants a further 20 were "pulled" abroad each year.

migrants can pass on to new migrants.[3] So, from the perspective of chain migration studies, regional variations in Italian emigration can be explained by variations in the stocks of previous migrants from the south and from the north of Italy. However, the *source* of this variation as well as the mechanism through which the migration process started when no stocks of migrants had yet been accumulated remain unclear.

In this paper, I extend the approach of chain migration literature so as to address the *origins* of large differences in migration to North America from the south and from the north of Italy. Using the key insight of chain migration studies on the importance of "friends and relatives effect" for the persistence in emigration, I do assume the existence of initial difference in the number of social contacts which Southern and Northern Italians had in the U. S. However, in contrast to the research on chain migration, I assume that this initial difference was only marginal and that the number of social contacts itself of either Southern or Northern Italians could be as small as 1. Moreover, in the model, the North American acquaintances of Italians are assumed to be *any* North American residents and not necessarily previous migrants from Italy.

The social connections between people are modelled with the network, where nodes represent individuals/agents and links indicate the existence of social relation between the linked individuals. Links serve for transmission of information about job opportunities. They predetermine employment prospects and expected earnings of each agent. Agents are residents of either *home* or *destination* country/region. In the context of Italian emigration, home denotes a region or a social group within a region in Italy (north or south) and destination denotes a region in North America where some of the residents had relation with Italians. I study the impact of a small change in the number of social connections between the home and the destination country on migration decisions of the home country residents, so as to infer the effects of initial variation in the number of North American contacts of the Southern and Northern Italians on subsequent migration flows.

Modelling job information sharing between people with a network has several advantages. First, networks are acknowledged to be a natural representation of the interpersonal information exchange and in particular, "one of the most extensively documented roles for social networks in economics is that of contacts in labor markets" (Jackson (2006), p.5). Secondly, the explicit design of the network of social relations allows for studying

---

[3]For example, Gilani et al. (1981) found in a study on Pakistan that about three-fourths of the migrants had obtained their Middle East jobs through friends and relatives. Massey and Espinosa (1997) in a survey of Mexican immigrants to the United States, showed that almost forty percent of migrants received jobs through informal networks of friends. Similar results are found for immigrants from Bangladesh and India to the Middle East (over 63 percent according to Mahmood (1991) and Nair (1991)), from Pakistan and Cyprus to Britain (Joly (1987) and Josephides and Rex (1987)), from Portugal to France (Hily and Poinard (1987)), and from Turkey to Germany (Wilpert (1988)).

the *mechanism* through which social interaction between agents in the labor market is transformed into migration decisions, and the impact which even a small change in the pattern of this interaction may have on resulting migration flow. These considerations, critical for understanding the determinants of migration decisions, have been largely absent from the existing migration analysis.

The migration decisions of the home country residents are modelled as an outcome of a one-shot simultaneous move game. The home country residents (players) maximize their lifetime expected income (payoff) and in the beginning of the first time period chose one of two strategies: either migrate to the destination country or stay at home. To study the equilibrium of the game, I first show that for any configuration of the social network, there exists a unique stationary distribution of agents' employment states in every country. Moreover, the stationary probability of employment of an agent is increasing in the number of his/her social connections in the country. Therefore, at least in case of a *complete* network, where all agents are connected with each other, the long-run probability of employment in the home country is *decreasing* with migration, whereas the probability of employment in the destination country is *increasing* with migration. As a result, when agents assign substantial weight to their long-run expected earnings, their migration decisions are *strategic complements*. The theory of the games in strategic complements then suggests that there exists a unique *maximal Nash equilibrium* in pure strategies.[4] In this paper, the result on strategic complementarity of players' decisions and the existence of a maximal Nash equilibrium is proved analytically for the case of a complete network structure. Then the numerical work supplements the analysis by addressing other network topologies.

The main finding of the paper is that the marginal change in the number of social connections between the home and the destination country/region may cause large variation in agents' migration decisions. So, the substantially higher migration to North America from the south of Italy than from the north of Italy could indeed have resulted from minor superiority in the initial number of North American connections of Italians from the south. Specifically, for the case when the network of social relations is complete and agents primarily care about their long-run expected income, I find that there exists a threshold value for the number of social links to the destination country (North America), such that no one migrates in the maximal equilibrium from a region or a social group in the home country where the actual number of these links is below the threshold (northern Italy), but everyone migrates from a region or a social group where the number of these links is above the threshold (southern Italy). This conclusion is

---

[4]By definition in Calvó-Armengol and Jackson (2004), a maximal equilibrium is such that the set of players staying at home is maximal meaning that the set of players staying at home in any other equilibrium of the game is a subset of those staying in at the maximal equilibrium.

also supported by the results of numerical simulations for the case of other network structures: the higher the number of agents' contacts in the destination country, the higher the equilibrium migration flow, and even a marginal increase in the number of these contacts may lead to a substantial increase in migration.

The remainder of the paper is organized as follows. The model describing the economy, evolution of employment, and migration decision game is presented in section 3.2. The solution of the game and its dependence on the structure of links in the network are demonstrated in section 3.3. In this section I first describe the results of analytical work for the case when the network is complete and agents assign a high weight to their long-run expected earnings. Then I present the results of numerical simulations for arbitrary network topologies. Finally, the concluding remarks are presented in section 3.4.

3.2 The Model

3.2.1 Economy

There are two countries/regions: the home country/region (H) and the destination country/region (D). The economy in (H) and (D) is infinite horizon in discrete time, $t \in \{1, 2, \dots\}$.

In the beginning of period 1, some residents of (H) migrate to (D). An agent in (H) migrates to (D) as soon as his expected life-time income in (D), net of the sunk moving cost c, is higher than his expected life-time income in (H). Having decided to migrate, the agent leaves (H) *immediately* and becomes *initially unemployed* resident of (D). The migration decision is made just once and no reentry is allowed for agents choosing to migrate.[5] Two ideas are implicit in these conditions. First, the majority of *costs of staying* in the labor market of a particular country (education, acquiring labor market specific skills and opportunity) are usually borne at an early stage of an agent's career and are sunk, so that there is little incentive to change the migration decision. Secondly, *costs of moving* are also sunk and usually high, which precludes from return migration. Furthermore, all *native* residents of (D) stay in their country permanently; they do not make any migration decisions.[6]

Let H be the set of all initial residents in (H), D be the set of all initial residents in (D) and $\mathcal{M}$ be the set of migrants. Let $a_H = |H|$, $a_D = |D|$, and $m = |M|$. So,

---

[5]Similar assumptions are made for the drop-out decision game in Calvó-Armengol and Jackson (2004).
[6]The precise description of the migration process is presented in section 3.2.5.

after migration the total number of residents in (H) is $a_H - m$ and the total number of residents in (D) is $a_D + m$.

For convenience of exposition, I assume that residents of both countries are assigned a number so that $H = \{1, \ldots, a_H\}$ and $D = \{a_H + 1, \ldots, a_H + a_D\}$. Besides, to address residents of (H) and (D) *after* migration I define two bijections

$$\alpha : H \setminus \mathcal{M} \to \{1, .., a_H - m\},$$
$$\gamma : D \cup \mathcal{M} \to \{1, .., a_D + m\},$$

such that α associates a certain number $\alpha(i) \in \{1, .., a_H - m\}$ with each agent i who is a *permanent* resident of (H) and γ associates a certain number $\gamma(j) \in \{1, .., a_D + m\}$ with each agent j who is either a migrant from (H) or an initial resident of (D).

Agents in (H) and (D) are connected by the network of social relations, G. Network G is undirected and for any $i \neq j$, $G_{ij} = 1$ indicates that agents i and j know each other, and $G_{ij} = 0$ indicates that they do not know each other. Also, by convention, $G_{ii} = 0$ for any agent i. I assume that network G is given exogenously and its structure does not change over time.

All residents of (H) and (D) are labor market participants. Jobs within each country are identical with the wage rate w_H per period in country (H) and the wage rate w_D per period in country (D). Both wages are exogenous, constant over time and $w_D > w_H$. Unemployed agents in each country earn nothing and receive no unemployment benefits.

For any period $t \geq 1$, the end-of-period t employment state of the labor market in (H) and (D) is characterized by vectors h_t and d_t, respectively. If agent i from (H) (or (D)) is employed *at the end of period t*, $h_t(\alpha(i)) = 1$ $(d_t(\gamma(i)) = 1)$, otherwise $h_t(\alpha(i)) = 0$ $(d_t(\gamma(i)) = 0)$. Thus, at any $t \geq 1$ h_t is a 0-1 vector of length $a_H - m$, $h_t \in \{0, 1\}^{a_H - m}$, and d_t is a 0-1 vector of length $a_D + m$, $d_t \in \{0, 1\}^{a_D + m}$.

Furthermore, let vectors h_0, d_0 characterize the *initial* employment state of the labor market in (H) and (D) or equivalently, the initial distribution of employment among agents in the two countries. Precisely, let h_0 be a vector of starting employment statuses in (H) of agents from the set H of *initial* (H) residents and d_0 be a vector of starting employment statuses in (D) of agents from the set $D \cup H$ of initial (D) *and* (H) residents. For the purposes of further analysis, the initial employment statuses in (D) of agents from H are set to unemployment, the status which these agents would have in the beginning of period 1 *if* they migrated to (D). Just like vectors h_t, d_t for $t \geq 1$, vectors h_0, d_0 are composed of 0's and 1's, where 1 indicates initial employment and 0 indicates initial unemployment; $h_0 \in \{0, 1\}^{a_H}$ and $d_0 \in \{0, 1\}^{a_H + a_D}$.

In the following, I denote by E_H all initially employed and by U_H a initially unemployed in (H) agents from H. Similarly, E_D denotes initially employed and U_D – initially unemployed in (D) agents from D. Formally,

$$E_H = \{i \in H \text{ s.t. } h_0(i) = 1\}, \quad U_H = \{i \in H \text{ s.t. } h_0(i) = 0\},$$
$$E_D = \{i \in D \text{ s.t. } d_0(i) = 1\}, \quad U_D = \{i \in D \text{ s.t. } d_0(i) = 0\}.$$

Clearly, $E_H \cup U_H = H$ and $E_D \cup U_D = D$. Let $e_H = |E_H|$, $e_D = |E_D|$, $u_H = |U_H|$, and $u_D = |U_D|$.

Furthermore, let E_H^i and U_H^i denote the sets of direct initially employed and unemployed in (H) contacts of agent i from H. Likewise, E_D^i and U_D^i denote the sets of direct initially employed and unemployed in (D) contacts of agent i from D.

$$E_H^i = \{j \in H \text{ s.t. } G_{ij} = 1, h_0(j) = 1\}, \quad U_H^i = \{j \in H \text{ s.t. } G_{ij} = 1, h_0(j) = 0\},$$
$$E_D^i = \{j \in D \text{ s.t. } G_{ij} = 1, d_0(j) = 1\}, \quad U_D^i = \{j \in D \text{ s.t. } G_{ij} = 1, d_0(j) = 0\}.$$

Let $e_H^i = |E_H^i|$, $e_D^i = |E_D^i|$, $u_H^i = |U_H^i|$, and $u_D^i = |U_D^i|$.

Below I describe the transmission of job information and dynamics of employment in the network. They are modelled as in Calvó-Armengol and Jackson (2004).

3.2.2 Timing of events

Each period $t \geq 1$ starts with some agents being employed and others unemployed as described by the employment states h_{t-1}, d_{t-1} of the previous period. If $t = 1$, then in the beginning of the period agents in (H) make migration decisions and those who decide to migrate actually leave (H) and become permanent (D) residents. After that the timing of events in period 1 and in any period $t > 1$ is the following. First, residents of both, (H) and (D) obtain information about new job openings *in their own country*. Any agent in (H) and (D) hears of a new job opening *directly* with a probability $p_a \in (0, 1)$. This job arrival process is independent across agents. If the agent is unemployed, he accepts the job offer. If the agent is already employed, he passes the offer along to one of the direct unemployed contacts *in the same country*. This is where the network structure of relationships between people becomes important: it determines the job information exchange process and ultimately affects long-run employment prospects of any agent.

Finally, the last event that happens in a period is that some employed agents lose their jobs. This happens randomly with a breakup probability $p_b \in (0, 1)$ which is also independent across all agents in (H) and (D).

3.2.3   Job information exchange

I impose four important assumptions on the job information exchange between agents.

First, the job information is transmitted in *at most one stage*[7]: if the job offer is neither taken by the agent who hears of the job directly nor by the direct contacts of this agent, the job is wasted rather than being passed on to some other needy party. Secondly, if several job offers arrive to one unemployed agent, he randomly selects only one of them and all other job offers remain unfilled.[8] Thirdly, an employed agent passes the job information to one of his direct unemployed contacts with *uniform randomization*.[9] Finally, *after migration* the information may only flow between agents *within the same country*.

The last assumption means that after migration taking place in the beginning of the first period, links connecting an agent from (H) with agents in (D) are only "active" if the agent from (H) has migrated to (D). At the same time, links connecting a migrant with his friends in (H) "stop functioning". This assumption can be justified as follows. Since migration decisions are irreversible, sending job information to agents in the other country after migration has taken place is essentially vain and job offers are doomed to be wasted. Indeed, agents in (D), either initial (D) residents or newcomers, cannot take on a job in (H) as they stay in (D) forever. Similarly, agents in (H) who decided to stay cannot take on a job in (D). Therefore, in this model, only job information exchanged between residents of the same country can be used to effectively result in employment.

Notice that such specification of the model suggests that although migration does not change the structure of the whole network G, it actually alters the structure of the labor market in (H) and in (D). In particular, when someone migrates, she/he is de facto removed from the set of agents in (H) and cannot share job information with agents who stayed in (H). This aspect of the model is of essential difference from the specification of the drop-out decision game in Calvó-Armengol and Jackson (2004). In their model, the network structure of the labor market is not altered when an agent drops out; instead the dropout's employment status is set to be zero forever but her/his social contacts continue to pass on information about jobs.

Lastly, this difference in information transmission via links connecting residents of the same country and links connecting residents of different countries plays a major role in

---

[7] The terminology of Boorman (1975).

[8] The first two assumptions rule out the study of chain effects in the transmission of job information. However, the one-stage information transmission is strongly supported by the empirical study in Granovetter (1974). The finding of the study is that information chains of length two or more account for only 15.6 percent of the sample.

[9] This is the so-called *equal neighbors treatment* assumption in Calvó-Armengol and Jackson (2004).

separating labor markets of (H) and (D). In particular, it predetermines irrelevance of the employment status and network structure in one market to employment conditions in the other market.

3.2.4 Transition between employment states

The evolution of employment of any agent i over time is predetermined by the sequence of stochastic events in each time period. For example, the transition from unemployment at the end of period $t - 1$ to employment at the end of period t is possible if agent i either receives a job offer directly, at rate p_a, or hears about a job offer from one of his direct employed contacts. The chance to hear about a job offer from an employed friend is increasing in probability p_a that this friend will receive an extra job offer but is decreasing in the number of other unemployed contacts of this friend. Moreover, having received a job, the agent can only keep it till the end of a period with probability $1 - p_b$. Somewhat simpler, the transition from employment at the end of period $t - 1$ to unemployment at the end of period t is fully determined by the breakup rate, p_b, and does not depend on the pattern of social connections of an agent. The other two transitions, when the employment status of an agent does not change over the period, are defined by the corresponding complementary events. Formally, for any agent $i \in H \setminus \mathcal{M}$ and for any $t \geq 2$, the four probabilities of transition between employment states are given by:

$$P(h_t(\alpha(i)) = 1 | h_{t-1}(\alpha(i)) = 0) =$$

$$= \begin{cases} (1 - p_b) \left[1 - (1 - p_a) \displaystyle\prod_{j \in E^i_{H\setminus\mathcal{M},t-1}} \left(1 - \frac{p_a}{|U^j_{H\setminus\mathcal{M},t-1}|} \right) \right] & \text{if } E^i_{H\setminus\mathcal{M},t-1} \neq \emptyset \\ (1 - p_b) p_a & \text{if } E^i_{H\setminus\mathcal{M},t-1} = \emptyset \end{cases} =$$

$$= (1 - p_b) \left[1 - (1 - p_a) \left(\prod_{j \in E^i_{H\setminus\mathcal{M},t-1}} \left(1 - \frac{p_a}{|U^j_{H\setminus\mathcal{M},t-1}|} \right) \right)^{\mathbf{1}_{E^i_{H\setminus\mathcal{M},t-1}}} \right] \tag{3.1}$$

$$P(h_t(\alpha(i)) = 0 | h_{t-1}(\alpha(i)) = 0) = 1 - P(h_t(\alpha(i)) = 1 | h_{t-1}(\alpha(i)) = 0) \tag{3.2}$$

$$P(h_t(\alpha(i)) = 0 | h_{t-1}(\alpha(i)) = 1) = p_b \tag{3.3}$$

$$P(h_t(\alpha(i)) = 1 | h_{t-1}(\alpha(i)) = 1) = 1 - P(h_t(\alpha(i)) = 0 | h_{t-1}(\alpha(i)) = 1) = 1 - p_b \tag{3.4}$$

where $E^i_{H\setminus\mathcal{M},t-1}$ is the set of direct contacts of agent i in $H \setminus \mathcal{M}$ who are employed in (H) at the end of period $t - 1$, $U^j_{H\setminus\mathcal{M},t-1}$ is the set of direct contacts of agent j ($j \in E^i_{H\setminus\mathcal{M},t-1}$) in $H \setminus \mathcal{M}$ who are unemployed in (H) at the end of period $t - 1$, and $\mathbf{1}_{E^i_{H\setminus\mathcal{M},t-1}} = 1$ if $E^i_{H\setminus\mathcal{M},t-1} \neq \emptyset$ but $\mathbf{1}_{E^i_{H\setminus\mathcal{M},t-1}} = 0$ otherwise.

Similarly, for any agent $i \in D \cup \mathcal{M}$ and any $t \geq 2$, the probabilities of transition between employment states are defined by four equations:

$$P(d_t(\gamma(i)) = 1 | d_{t-1}(\gamma(i)) = 0) =$$

$$= \begin{cases} (1 - p_b) \left[1 - (1 - p_a) \prod\limits_{j \in E^i_{D \cup \mathcal{M}, t-1}} \left(1 - \frac{p_a}{|U^j_{D \cup \mathcal{M}, t-1}|} \right) \right] & \text{if } E^i_{D \cup \mathcal{M}, t-1} \neq \emptyset \\ (1 - p_b) p_a & \text{if } E^i_{D \cup \mathcal{M}, t-1} = \emptyset \end{cases} =$$

$$= (1 - p_b) \left[1 - (1 - p_a) \left(\prod\limits_{j \in E^i_{D \cup \mathcal{M}, t-1}} \left(1 - \frac{p_a}{|U^j_{D \cup \mathcal{M}, t-1}|} \right) \right)^{\mathbf{1}_{E^i_{D \cup \mathcal{M}, t-1}}} \right] \quad (3.5)$$

$$P(d_t(\gamma(i)) = 0 | d_{t-1}(\gamma(i)) = 0) = 1 - P(d_t(\gamma(i)) = 1 | d_{t-1}(\gamma(i)) = 0) \quad (3.6)$$

$$P(d_t(\gamma(i)) = 0 | d_{t-1}(\gamma(i)) = 1) = p_b \quad (3.7)$$

$$P(d_t(\gamma(i)) = 1 | d_{t-1}(\gamma(i)) = 1) = 1 - P(d_t(\gamma(i)) = 0 | d_{t-1}(\gamma(i)) = 1) = 1 - p_b \quad (3.8)$$

where $E^i_{D \cup \mathcal{M}, t-1}$ is the set of direct contacts of agent i in $D \cup \mathcal{M}$ who are employed in (D) at the end of period $t - 1$, $U^j_{D \cup \mathcal{M}, t-1}$ is the set of direct contacts of agent j ($j \in E^i_{D \cup \mathcal{M}, t-1}$) in $D \cup \mathcal{M}$ who are unemployed in (D) at the end of period $t - 1$, and $\mathbf{1}_{E^i_{D \cup \mathcal{M}, t-1}} = 1$ if $E^i_{D \cup \mathcal{M}, t-1} \neq \emptyset$ but $\mathbf{1}_{E^i_{D \cup \mathcal{M}, t-1}} = 0$ otherwise.

Notice that essentially the same equations determine the probabilities of transition from the *initial* employment status to the employment status at the end of period 1 in both countries. The equations should only be corrected for the fact that the initial employment status of any agent $i \in H \setminus \mathcal{M}$ is given by $h_0(i)$ and the initial employment status of $i \in D \cup \mathcal{M}$ is given by $d_0(i)$.

Remark *Time t employment status of any agent i in a given country is fully determined by the job arrival and breakup rates, p_a and p_b, by the network structure and the time $t - 1$ employment status of agents in a two-links-away neighborhood of agent i in that country. The probability of being employed at period t for agent i who is unemployed at the end of period $t - 1$ is increasing in the number of direct employed contacts and decreasing in the number of two-links-away unemployed contacts who share at least one of their direct employed contacts with i.*

The Remark suggests that in a *one-period-ahead* perspective, direct employed contacts of unemployed agent i improve his prospects for hearing about a job offer while unemployed two-links-away contacts are agent i's "competitors" for job information and therefore, decrease his chances of employment. More distant indirect relationships do not have

an impact on one-period-ahead employment prospects of i. However, in a longer time frame, the larger network and status of other agents affect employment status of agent i through the effect they have on the employment status of i's connections.

Given transition probabilities between employment states of any agent in (H) and in (D), transition probability between two employment states of the whole labor market in each country is simply a product of the transition probabilities between the corresponding employment states of agents. So, for any two employment states h, h' in (H) and d, d' in (D) and for any $t \geq 2$,

$$P(h_t = h'|h_{t-1} = h) = \prod_{i \in H \setminus \mathcal{M}} P(h_t(\alpha(i)) = h'(\alpha(i))|h_{t-1}(\alpha(i)) = h(\alpha(i))) \quad (3.9)$$

$$P(d_t = d'|d_{t-1} = d) = \prod_{i \in D \cup \mathcal{M}} P(d_t(\gamma(i)) = d'(\gamma(i))|d_{t-1}(\gamma(i)) = d(\gamma(i))) \quad (3.10)$$

The Remark, (3.9) and (3.10) imply that employment state h_t of the labor market in (H) and employment state d_t of the labor market in (D) follow two separate finite state *Markov processes*. I denote by $M(G[H \setminus \mathcal{M}], p_a, p_b)$ the Markov process for h_t and by $M(G[D \cup \mathcal{M}], p_a, p_b)$ the Markov process for d_t.[10]

Markov processes $M(G[H \setminus \mathcal{M}], p_a, p_b)$ and $M(G[D \cup \mathcal{M}], p_a, p_b)$ have several important characteristics. First, they are both *homogenous*. This follows from the fact that job arrival and breakup probabilities, p_a and p_b, are constant and the structure of the network in (H) and (D) does not change after migration taking place in the very beginning of period 1. Secondly, Markov chains for h_t and d_t are *irreducible* and *aperiodic*, since all transition probabilities are strictly positive.

These properties of Markov chains $M(G[H \setminus \mathcal{M}], p_a, p_b)$ and $M(G[D \cup \mathcal{M}], p_a, p_b)$ lead to the following statement:

Proposition 3.1. *There exists a unique stationary distribution of employment states in (H) and a unique stationary distribution of employment states in (D).*

The stationary distribution of employment in (H) and (D) defines a *unique* steady-state probability of employment of any agent in every country:

$$p_{ss}^i = \sum_{h \text{ s.t. } h(\alpha(i))=1} \mu(h) \quad \forall \quad i \in H \setminus \mathcal{M} \quad (3.11)$$

$$q_{ss}^i = \sum_{d \text{ s.t. } d(\gamma(i))=1} \nu(d) \quad \forall \quad i \in D \cup \mathcal{M} \quad (3.12)$$

---

[10]$G[A]$ denotes the network induced by graph G on the set of agents A.

where μ is the stationary distribution of employment states in (H) and ν is the stationary distribution of employment states in (D).

3.2.5 Migration decision game

Description of the game   Migration decision making of residents in (H) is modelled as a *one-shot simultaneous move game*, Γ, where the structure of the game is common knowledge. The game takes place in the beginning of period 1 when no information about new job openings has yet arrived in either of the two countries. The players are all residents of (H). They simultaneously choose one of two actions: staying in (H), denoted by s, or migrating to (D), denoted by m, so as to maximize their expected life-time income. In the strategic form, $\Gamma = [H, \Sigma, \{\pi^i\}_{i \in H}]$, where H is a set of players, $\Sigma = \{s, m\}$ is each player's set of pure strategies, and π^i is a payoff function of player i, the expected income of i in (H) or (D).

Strategies and payoff function   Let $\sigma = (\sigma(1), \ldots, \sigma(a_H))$ be a profile of *pure strategies* of all players in H and $\sigma(-i)$ be a profile of pure strategies of player i's opponents. Let $v_H^i(\sigma(-i))$ and $v_D^i(\sigma(-i))$ represent the expected life-time income of player i in (H) and (D), provided that i's opponents play $\sigma(-i)$. Then for each player $i \in H$ and any strategy profile σ, the payoff function π^i is defined by:

$$\pi^i(s, \sigma(-i)) = v_H^i(\sigma(-i)),$$
$$\pi^i(m, \sigma(-i)) = v_D^i(\sigma(-i)).$$

To write functions v_H^i and v_D^i explicitly, I impose an important simplifying assumption. For the rest of the paper I assume that agents in both countries start in period 1, and then jump to the steady state in the next "period".[11] This gives a rough representation of the life-time optimization problem of players, but enough to see the effects of their interaction in the short and in the long run.

Under this assumption, the expected life-time income of player i in either country is the sum of the expected income in period 1 and the present discounted value of the future expected income starting from period 2 onwards. The future expected income is represented by the infinite flow of the *identical* one-period expected earnings determined by the product of the steady-state probability of employment and constant wage rate.

---

[11] The same simplification is made in Calvó-Armengol and Jackson (2004) for their study of the drop-out decision game.

So, each player i who stays in (H) receives a payoff

$$v_H^i(\sigma(-i)) = \begin{cases} w_H + \beta w_H p_{ss}^i(\sigma(-i)) + \beta^2 w_H p_{ss}^i(\sigma(-i)) + \cdots \\ w_H p^i(\sigma(-i)) + \beta w_H p_{ss}^i(\sigma(-i)) + \beta^2 w_H p_{ss}^i(\sigma(-i)) + \cdots \end{cases} =$$

$$= \begin{cases} w_H + \frac{\beta}{1-\beta} w_H p_{ss}^i(\sigma(-i)) & \text{if } h_0(i) = 1 \\ w_H p^i(\sigma(-i)) + \frac{\beta}{1-\beta} w_H p_{ss}^i(\sigma(-i)) & \text{if } h_0(i) = 0 \end{cases} \quad (3.13)$$

where p^i and p_{ss}^i are the probabilities of player i to be employed in (H) at the beginning of period 1 (if i is initially unemployed) and at the steady-state,[12] and $0 < \beta < 1$ is a constant discount factor.

More compactly (3.13) can be written as:

$$v_H^i(\sigma(-i)) = w_H \left[(1 - h_0(i)) p^i(\sigma(-i)) + h_0(i) \right] + \frac{\beta}{1-\beta} w_H p_{ss}^i(\sigma(-i)). \quad (3.14)$$

Likewise, if player i moves to (D), he receives a payoff

$$\begin{aligned} v_D^i(\sigma(-i)) &= w_D q^i(\sigma(-i)) - c + \beta w_D q_{ss}^i(\sigma(-i)) + \beta^2 w_D q_{ss}^i(\sigma(-i)) + \cdots = \\ &= w_D q^i(\sigma(-i)) - c + \frac{\beta}{1-\beta} w_D q_{ss}^i(\sigma(-i)) \end{aligned} \quad (3.15)$$

where q^i and q_{ss}^i are the probabilities of player i to be employed in (D) at the beginning of period 1 and at the steady-state.[13]

Notice that probabilities p^i and q^i, in the above equations, differ from the probabilities of transition between initial unemployment and end-of-period-1 employment as defined in section 3.2.4. In fact, p^i and q^i are higher than the corresponding transition probabilities, since they describe the chance of getting employed in the beginning of a period *irrespective* of the fact that the job can be lost in the end of the period:

$$\begin{aligned} p^i &= \frac{P(h_1(\alpha(i)) = 1 | h_0(i) = 0)}{1 - p_b}, \quad &(3.16) \\ q^i &= \frac{P(d_1(\gamma(i)) = 1 | d_0(i) = 0)}{1 - p_b} \quad &(3.17) \end{aligned}$$

Equilibrium The equilibrium of game Γ is a standard Nash equilibrium in pure strategies. Below I study the properties of this equilibrium in order to infer how migration decisions of individuals are affected by the structure of their social relations and in particular, by the number of their contacts in the destination country.

---

[12]Recall that the steady-state probabilities of employment of player i in (H) and in (D), p_{ss}^i and q_{ss}^i, were defined in (3.11) and (3.12).

[13]Due to the assumption that any migrant is initially unemployed in (D), v_D^i represents the expected life-time income of initially *unemployed* agent in (D).

3.3 Results

First, I consider the specific case when network G of social connections is *complete* (or represents a union of complete components), so that all individuals in the network are linked with each other. Furthermore, I assume that discount factor β is sufficiently close to 1, that is, agents mainly care about their expected earnings in the long run. In fact, when β is sufficiently "large", the weight $\frac{\beta}{1-\beta}$ assigned by agents to their expected earnings from period 2 onwards is substantially larger than the weight assigned to their expected income in period 1. As a result, migration decisions of home country residents are essentially driven by their long-run income considerations.

For this admittedly restrictive setting I solve the model analytically. I prove the existence of equilibrium and study the change in the equilibrium migration choices as the number of links to the destination country changes just marginally. After that I relax the assumptions on the network structure and the discount parameter and examine the outcomes of the migration decision game in some simple numerical examples.

3.3.1 Steady-state probability of employment and maximal Nash equilibrium in the complete network

The long-run expected income in (H) and in (D), which determines migration choices of players when β is large enough, hinges on the steady-state probability of employment. Below I consider the dependence of the steady-state probability of employment in the complete network on the total number of agents, or alternatively, on the number of direct social contacts of each agent.[14]

The steady-state probability of employment can be derived using the *subdivision of periods* procedure proposed in Calvó-Armengol and Jackson (2004). The idea of this procedure is that as we divide job arrival rate, p_a, and job breakup rate, p_b, by some larger and larger factor, we are essentially looking at arbitrarily short time periods. Then one can begin to differentiate between the short- and long-run effects.[15,16]

---

[14]Clearly, since the network is complete, the steady-state probability of employment is fully defined by the total number of agents in the network.

[15]In the limit, the resulting process for employment approaches a continuous time (Poisson) process, which is a natural situation where the short-run effects of agents' interaction are inconsequential and employment prospects of any individual are determined by the long-run effects.

[16]Formally, by dividing p_a and p_b by some common factor T, we obtain an associated Markov process $M(g, \frac{p_a}{T}, \frac{p_b}{T})$ for the employment state on a complete network g. In terminology of Calvó-Armengol and Jackson (2004), this Markov process is called the *T-period subdivision* of Markov process $M(g, p_a, p_b)$. Lemma 3.2 addresses the situation when T is sufficiently large, so that the true steady-state probability of employment under $M(g, p_a, p_b)$ and $M(g, \frac{p_a}{T}, \frac{p_b}{T})$ can be calculated using the approximate Markov process. Further details are provided in the Appendix.

Lemma 3.2. *Under fine enough subdivision of periods, the steady-state probability of employment in a complete network of size n (formed by n agents) is equal to*

$$\theta(n) = \frac{1 + \sum_{l=1}^{n-1} \prod_{k=1}^{l} \frac{p_b}{p_a}\left(1 - \frac{k}{n}\right)}{1 + \frac{p_b}{p_a} + \frac{p_b}{p_a}\sum_{l=1}^{n-1}\prod_{k=1}^{l}\frac{p_b}{p_a}\left(1 - \frac{k}{n}\right)}. \tag{3.18}$$

It is strictly increasing in n.

The statement of Lemma 3.2 is illustrated with Figure 3.1. The strictly monotonic pattern of the steady-state probability is preserved for all combinations of parameters p_a, p_b; higher p_a and lower p_b result in better employment prospects in the long run.
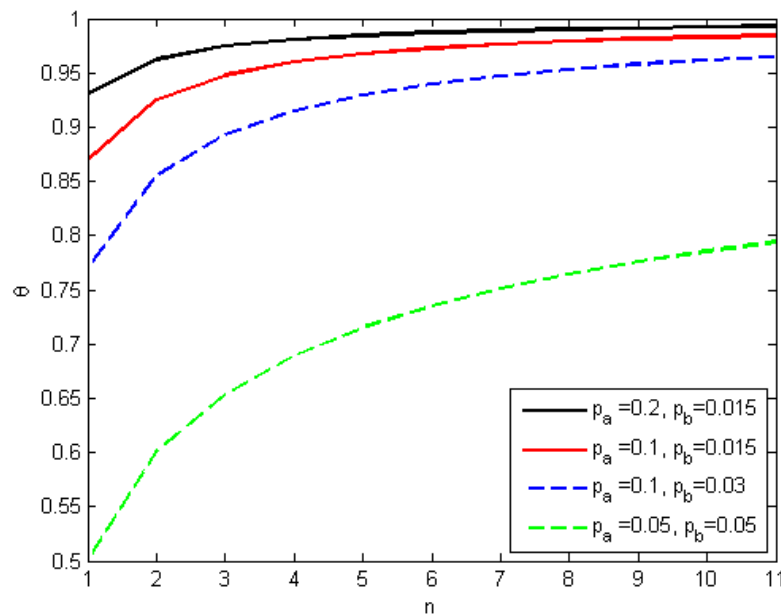
FIGURE 3.1: Steady-state probability of employment in the complete network of size
n.

Thus, despite the short-run competition for jobs between unemployed agents,[17] in the long run, having more agents in the network is good news for everyone as it effectively improves future employment prospects of any agent. In particular, this implies that in the complete network, migration has a long-lasting positive effect on employment probability in (D) but is also detrimental for the long-run employment probability in (H).[18] As a consequence, if the weight assigned to the long-run expected income is high enough, i.e. β is sufficiently close to 1, the incentive of any agent to migrate is increasing as more of the other players migrate. Likewise, the greater the number of players who stay, the stronger the incentive for others to stay. This inference gives rise to Proposition 3.3.

---

[17]See the Remark in section 3.2.4.

[18]This result supports the idea that migrants can promote the welfare of the destination country.

Proposition 3.3. *Given complete network G and fine enough subdivision of periods, there exists $\bar{\bar{\beta}}$ such that for all $\beta \geq \bar{\bar{\beta}}$:*

1. *Decisions of players in H are strategic complements, that is, game Γ is supermodular*

2. *There exists a unique maximal Nash equilibrium in pure strategies*

The existence of a maximal Nash equilibrium follows from the theory of supermodular games. The maximal equilibrium is defined as in Calvó-Armengol and Jackson (2004). It is an equilibrium where the set of players staying in (H) is maximal, that is, the set of players staying in (H) at any other equilibrium of the game is a subset of those staying in (H) at the maximal equilibrium.

3.3.2 The effects of links between countries on equilibrium migration flow

Strategic complementarity of players' decisions and interchangeability of players' positions in the complete network imply that only four outcomes are possible at equilibrium: initially unemployed players either all migrate or all stay and initially employed players either all migrate or all stay. Moreover, since irrespectively of players' initial employment status, expected earnings in period 1 are negligible compared to those from period 2 onwards (β is close to 1), not only the strategies of players with identical initial status are the same but the strategies of *all* players are the same. This leads to Proposition 3.4.

Proposition 3.4. *Given complete network G and fine enough subdivision of periods, either no one or everyone migrates at the maximal Nash equilibrium. The situation when no one migrates is the maximal Nash equilibrium if and only if*

$$w_H \theta(a_H) \geq w_D \theta(a_D + 1). \tag{3.19}$$

Condition (3.19) states that for any agent, the steady-state expected income in (H) is at least as high as in (D), provided that all other players stay. Thus, the situation when no one migrates is sustainable at the Nash equilibrium.[19] Furthermore, the situation when everyone migrates is *always* a Nash equilibrium (although not necessarily maximal), since for any $w_D > w_H$,

$$w_H \theta(a_H) < w_D \theta(a_D + a_H)$$

---

[19]Recall that when β is high enough, it is the long-run expected income in (H) and in (D) that determines equilibrium decisions of players.

It is easy to see that the larger the number a_D of players' contacts in D, the more likely it is that in the maximal equilibrium, all players migrate. Formally, given a_H, w_H and w_D, there exists a threshold value $\bar{a}_D(a_H, w_H, w_D)$ of a_D such that as soon as a_D exceeds this threshold, the inequality $w_H \theta(a_H) \geq w_D \theta(a_D + 1)$ does not hold and the outcome with all players staying in (H) is *not* an equilibrium. Therefore, as a corollary of Proposition 3.4, we obtain:

Proposition 3.5. *The number of migrants in the maximal equilibrium of Γ is not decreasing in the number of players' contacts in D. Moreover, for any a_H, w_H and w_D, there exists a threshold value $\bar{a}_D(a_H, w_H, w_D)$ such that no one migrates in the maximal Nash equilibrium when the number of players' contacts in D is equal to $\bar{a}_D(a_H, w_H, w_D)$, but everyone migrates if this number of contacts exceeds $\bar{a}_D(a_H, w_H, w_D)$ at least by 1. The threshold $\bar{a}_D(a_H, w_H, w_D)$ is defined by the equation:*

$$\frac{w_H}{w_D}\theta(a_H) = \theta(a_D + 1) \tag{3.20}$$

Thus, under certain conditions on the network structure and on the ratio of wages in (H) and in (D), just a minor increase in the number of social connections with (D) leads to a sharp increase in migration at the maximal Nash equilibrium.

The result of Proposition 3.5 is certainly very stylized. However, it still demonstrates the desired effects. In particular, for the case of 19th century Italian emigration, it suggests that the striking difference in migration flows to North America from the south and from the north of Italy could have resulted from a small initial difference in the number of social links to North America from each of the two regions.[20]

More formally, suppose that H_N and H_S are two social groups in the south and in the north of Italy respectively and that D_N and D_S are two groups in North America, such that D_N has social connections with H_N and D_S has social connections with H_S. Suppose that communities in the south and in the north of Italy are "separated", so that they do not have an impact on the long-run expected earnings of each other. For example, let $G[H_N]$, $G[H_S]$ form separate components of the network in Italy and $G[D_N]$, $G[D_S]$ form separate components of the network in North America. Also, let both $G[H_N \cup D_N]$ and $G[H_S \cup D_S]$ be complete networks and $|H_N| = |H_S| = a_H$. Furthermore, assume that $|D_S| > \bar{a}_D(a_H, w_H, w_D) \geq |D_N|$, that is, the number of North American contacts of the Southern Italians is larger than the threshold $\bar{a}_D(a_H, w_H, w_D)$ defined above and the number of North American contacts of the Northern Italians is lower or equal to this threshold. Then according to Proposition 3.5, in the maximal

---

[20]Moreover, the preceding analysis suggests that if Italian potential migrants were mainly concerned about their earnings in the long run, the gap in *initial* employment conditions favoring North America over Italy did not play a major role in determining the migration decisions.

Nash equilibrium of the migration decision game, everyone migrates from the group H_S of Southern Italians but no one migrates from the group H_N of the Northern Italians, provided that both, Northern and Southern Italians mainly care about their expected income in the long run.

For example, for the specific parameter values, $p_a = p_b = 0.05$, $a_H = 10$ and $\frac{w_H}{w_D} \in [0.6366, 0.7639)$, condition (3.19) suggests that in the maximal Nash equilibrium, no one migrates from H_N, if agents in H_N do not have any social connections in North America ($|D_N| = 0$), but everyone migrates from H_S, if agents in H_S have just one acquaintance in North America ($|D_S| = 1$).[21]

3.3.3 The effects of links between countries on migration in non-complete networks: Numerical exercise

In this section I relax the assumptions that network G is complete and discount factor β is close to 1. I show that the basic insights from the previous analysis about the impact of social links between agents on the steady-state probability of employment and on the size of migration flow continue to hold.

First, just as in case of a complete network structure, I find that the steady-state probability of employment of any agent in a non-complete network is increasing in the number of the agent's direct contacts. This result is demonstrated by examples in Table 3.1. Table 3.1 displays the values of the steady-state probability of employment of agent 1. As the number of connections between agent 1 and other agents increases, the steady-state probability of employment grows.[22] Moreover, everything else being equal, higher job information arrival rate, p_a, and lower job breakup rate, p_b, result in higher probability of employment in the long run.

As before, the finding of a positive impact of direct social contacts on each agent's steady-state probability of employment suggests that as soon as discount factor β is high enough, directly linked players tend to choose identical strategies. However, strategic complementarity of *all* players' decisions and the existence of a (maximal) Nash equilibrium in the game are contingent on the particularity of the network structure and chosen parameter values.

---

[21]This result is implied by inequality
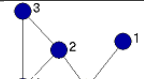
$$w_D\theta(1) \leq w_H\theta(a_H) < w_D\theta(2)$$

where $\theta(1) \approx 0.5$, $\theta(2) \approx 0.6$, and $\theta(a_H) = \theta(10) \approx 0.7854$ according to Figure 3.1 (see the plot corresponding to $p_a = p_b = 0.05$).

[22]The upper part of Table 3.1 displays the change in the steady-state probability of employment of agent 1 when the number of his links with other agents increases but the size of the network remains fixed; the lower part of the table continues the exercise for the case when an increase in the number of links of agent 1 comes along with an increase in the size of the network.

TABLE 3.1: Steady-state probability of employment of agent 1 for different values of p_a and p_b

network G (size of G is fixed)	$p_a = 0.2$ $p_b = 0.015$	$p_a = 0.1$ $p_b = 0.015$	$p_a = 0.1$ $p_b = 0.03$	$p_a = 0.05$ $p_b = 0.05$
	0.9292	0.8678	0.7638	0.4872
	0.9583	0.9219	0.8463	0.5748
	0.9686	0.9426	0.8818	0.6269
	0.9770	0.9608	0.9163	0.6971

network G (size of G changes)	$p_a = 0.2$ $p_b = 0.015$	$p_a = 0.1$ $p_b = 0.015$	$p_a = 0.1$ $p_b = 0.03$	$p_a = 0.05$ $p_b = 0.05$
	0.9584	0.9221	0.8469	0.5760
	0.9686	0.9425	0.8818	0.6294
	0.9740	0.9544	0.9052	0.6800
	0.9772	0.9616	0.9196	0.7154

Remark: The probabilities are expressed in percentage points.

If Nash equilibrium exists, it can be defined by simple conditions. To state these conditions formally for a *generic* network structure and *any* $0 < \beta < 1$, I introduce new notation. Let C_H^i, C_D^i be the sets of those players in H who are competitors or potential competitors of player i for the first-period job offers in (H) and in (D), respectively:

$$\forall i \in U_H \quad C_H^i = \bigcup_{j \in E_H^i} C_{H,\text{from } j}^i, \text{ where } C_{H,\text{from } j}^i = \left\{ l \in U_H \text{ s.t. } l \neq i, j \in E_H^l \right\},$$

$$\forall i \in H \quad C_D^i = \bigcup_{j \in E_D^i} C_{D,\text{from } j}^i, \text{ where } C_{D,\text{from } j}^i = \left\{ l \in H \text{ s.t. } l \neq i, j \in E_D^l \right\}$$

If player i is initially employed in (H), I assume that the set of his competitors in (H) is empty:

$$\forall i \in E_H \quad C_H^i = \emptyset$$

Using this notation, an equilibrium of game Γ can be defined as a strategy profile such that for any player i who migrates, $i \in \mathcal{M}^*$, the following inequality holds:

$$w_H \left[(1 - h_0(i)) p^{i*} + h_0(i) \right] + \frac{\beta}{1-\beta} w_H p_{ss}^{i*} \le w_D q^{i*} - c + \frac{\beta}{1-\beta} w_D q_{ss}^{i*}$$

where

$$p^{i*} = 1 - (1 - p_a) \left[\prod_{j \in E_H^i \backslash \mathcal{M}^*} \left(1 - \frac{p_a}{u_H^j - c_{H,\text{from } j}^i} \right) \right]^{\mathbf{1}_{E_H^i \backslash \mathcal{M}^*}},$$

$$q^{i*} = 1 - (1 - p_a) \left[\prod_{j \in E_D^i} \left(1 - \frac{p_a}{u_D^j + 1 + c_{D,\text{from } j}^i} \right) \right]^{\mathbf{1}_{E_D^i}},$$

$c_{H,\text{from } j}^i = |C_{H,\text{from } j}^i \cap \mathcal{M}^*|$, $c_{D,\text{from } j}^i = |C_{D,\text{from } j}^i \cap \mathcal{M}^*|$, and p_{ss}^{i*}, q_{ss}^{i*} are the steady-state probabilities of employment of player i on $G[\{H \backslash \mathcal{M}^*\} \cup \{i\}]$ and $G[D \cup \mathcal{M}^*]$, respectively. At the same time, for any player k who does not migrate, $k \in H \backslash \mathcal{M}^*$, the opposite inequality or equality holds.
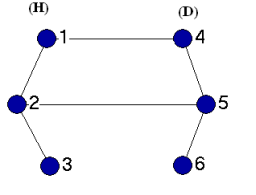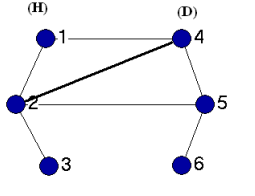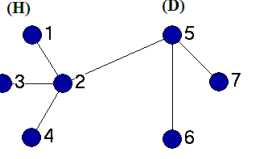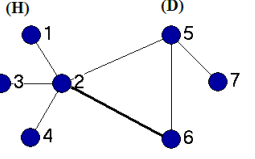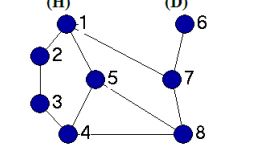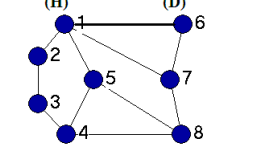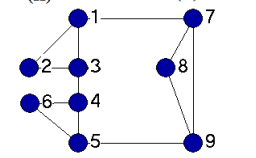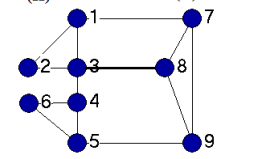
I use this definition of equilibrium in order to find equilibrium of the game numerically for a range of network structures and particular parameter values. I focus on a change in equilibrium strategies of players caused by a single link increase in the number of connections between (H) and (D).

Table 3.2 presents the results for four specific examples. The values of parameters used to calculate these examples are $w_H = 5$, $w_D = 7$, $c = 1$, $p_a = 0.1$, $p_b = 0.015$, and $\beta = 0.7$.[23] For these parametric assumptions and for a large set of combinations of initial employment states in (H) and (D), I find that there exists a *unique maximal Nash equilibrium* of the migration decision game in each of the four examples. Moreover, for a subset of initial employment states in two countries, a minimal increase in the number of social links between (H) and (D) causes a substantial increase in the amount of agents who choose to migrate at the maximal equilibrium. For example, for some employment conditions, not only the agent whose own number of links increases changes his strategy in favor of migrating but also some other agents in (H) do so.[24]

---

[23]The same values of p_a and p_b are used in the numerical examples of Calvó-Armengol and Jackson (2004). The authors argue that "if we think about these numbers from the perspective of a time period being a week, then an agent looses a job roughly on average once in 67 weeks, and hears (directly) about a job on average once in every 10 weeks" (p. 430).

[24]See examples 1, 3, and 4 of Table 3.2.

TABLE 3.2: Equilibrium strategy profiles of players in H. Changes caused by adding a link

---

Example 1: $H = \{1, 2, 3\}$, $D = \{4, 5, 6\}$, link $2 - 4$ is added

Initial employment states in H and D		
$h_0 = (0,1,1), d_{0|D} = (1,1,0)$	(m,s,s)	(m,m,s)
$h_0 = (0,1,1), d_{0|D} = (1,1,1)$	(m,s,s)	(m,m,s)
$h_0 = (1,1,0), d_{0|D} = (1,1,1)$	(s,s,m)	(m,m,m)
$h_0 = (1,1,1), d_{0|D} = (1,1,0)$	(s,s,s)	(m,m,s)
$h_0 = (1,1,1), d_{0|D} = (1,1,1)$	(s,s,s)	(m,m,s)

Example 2: $H = \{1,2,3,4\}$, $D = \{5,6,7\}$, link $2 - 6$ is added

Initial employment states in H and D		
$h_0 = (0,1,0,1), d_{0|D} = (1,1,0)$	(m,s,m,s)	(m,m,m,s)
$h_0 = (0,1,1,0), d_{0|D} = (1,1,0)$	(m,s,s,m)	(m,m,s,m)
$h_0 = (0,1,1,1), d_{0|D} = (1,1,0)$	(m,s,s,s)	(m,m,s,s)
$h_0 = (1,1,0,0), d_{0|D} = (1,1,0)$	(s,s,m,m)	(s,m,m,m)
$h_0 = (1,1,0,1), d_{0|D} = (1,1,0)$	(s,s,m,s)	(s,m,m,s)
$h_0 = (1,1,0,1), d_{0|D} = (1,1,1)$	(s,s,m,s)	(s,m,m,s)
$h_0 = (1,1,1,0), d_{0|D} = (1,1,0)$	(s,s,s,m)	(s,m,s,m)

Example 3: $H = \{1,2,3,4,5\}$, $D = \{6,7,8\}$, link $1 - 6$ is added

Initial employment states in H and D		
$h_0 = (1,0,0,0,1), d_{0|D} = (1,1,0)$	(s,m,m,m,s)	(m,m,m,m,m)
$h_0 = (1,0,0,0,1), d_{0|D} = (1,1,1)$	(s,m,m,m,s)	(m,m,m,m,m)
$h_0 = (1,0,0,1,1), d_{0|D} = (1,1,0)$	(s,m,m,s,s)	(m,m,m,s,s)
$h_0 = (1,0,0,1,1), d_{0|D} = (1,1,1)$	(s,m,m,s,s)	(m,m,m,s,s)
$h_0 = (1,0,1,0,1), d_{0|D} = (1,1,0)$	(s,m,s,m,s)	(m,m,s,m,m)
$h_0 = (1,0,1,0,1), d_{0|D} = (1,1,1)$	(s,m,s,m,s)	(m,m,s,m,m)
$h_0 = (1,0,1,1,1), d_{0|D} = (1,1,0)$	(s,m,s,s,s)	(m,m,s,s,s)
$h_0 = (1,1,0,0,1), d_{0|D} = (1,1,0)$	(s,s,m,m,s)	(m,s,m,m,m)
$h_0 = (1,1,0,0,1), d_{0|D} = (1,1,1)$	(s,s,m,m,s)	(m,s,m,m,m)
$h_0 = (1,1,1,0,0), d_{0|D} = (1,1,1)$	(s,s,s,m,m)	(m,s,s,m,m)
$h_0 = (1,1,1,0,1), d_{0|D} = (1,1,0)$	(s,s,s,m,s)	(m,s,s,m,m)
$h_0 = (1,1,1,0,1), d_{0|D} = (1,1,1)$	(s,s,s,m,s)	(m,s,s,m,m)

Example 4: $H = \{1,2,3,4,5,6\}$, $D = \{7,8,9\}$, link $3 - 8$ is added

Initial employment states in H and D		
$h_0 = (0,1,1,0,0,1), d_{0|D} = (0,1,1)$	(m,s,s,m,m,s)	(m,s,m,m,m,s)
$h_0 = (0,1,1,0,0,1), d_{0|D} = (1,1,1)$	(m,s,s,m,m,s)	(m,s,m,m,m,s)
$h_0 = (0,1,1,0,1,0), d_{0|D} = (1,1,1)$	(s,s,s,m,m,m)	(m,m,m,m,m,m)

Examples of Table 3.2 can be viewed as an illustration of Italian emigration case. Indeed, let H represents a group of people in the *north* of Italy and D represents a group in North America. Suppose that a group of people in the *south* of Italy has the same structure of social relations with each other and the same initial employment state as a group of people in the north of Italy, H. Then the examples of Table 3.2 suggests that for a certain combination of initial employment states of Italians and North Americans, a substantially higher proportion of Southern Italians migrates to North America as soon as the number of social connections between the Southern Italians and North Americans is just marginally higher than the number of connections between the Northern Italians and North Americans.

3.4 Conclusion

The aim of this paper is to examine whether the puzzling lack of migration from the north of Italy to North America, despite the evident economic benefits of that move, and simultaneous large migration from the south of Italy could have resulted from the *initial small difference* in the number of social contacts in North America of the Northern and Southern Italians. By studying the *origins* of variation in migration, I intend to contribute to the research on chain migration which concentrates on explaining the *continuation* of migration, provided that substantial amount of people have migrated previously.

To address the case of Italian emigration, I consider an explicit network model with two countries, a home and a destination country, and study the effects of a small variation in the number of social links between these countries on the size of the migration flow. Agents in the network are labor market participants. They use their links with direct contacts to exchange information about job opportunities in the local labor market. Employment of every agent in each time period is subject to the idiosyncratic risk. If an agent is unemployed, then having links to employed agents improves his prospects for hearing about a job offer. In contrast, having two-links-away unemployed competitors for job information worsens these prospects. In the beginning of period 1, residents of the home country compare their employment prospects at home and in the destination country and take once-and-for-all migration decisions. These migration decisions are modelled as a pure-strategy Nash equilibrium of a one-shot simultaneous move game. I study the properties of this equilibrium.

The analytical work in the paper is restricted to the case when the network of social connections is complete and discount factor β is sufficiently close to 1. I find that in

such a setting, the steady-state probability of employment of any agent is increasing in the size of the network. This implies that migration improves the long-run employment prospects in the destination country but worsens those in the home country. As a result, migration decisions of the home country residents turn out to be strategic complements which guarantees the existence of a unique maximal Nash equilibrium in the game.

The main finding of the paper is that substantially higher migration to North America from the south than from the north of Italy could indeed be an outcome of originally minor superiority in the number of social links with North America of the Southern Italians over those of the Northern Italians. In particular, in the stylized setting of the model, when the network of social connections is complete and β is close to 1, there exists a threshold value for a number of links to North America, such that as soon as the actual number of links to North America of a group of the Northern Italians is lower than this threshold and the actual number of such links of a group of the Southern Italians is higher than this threshold, no one migrates from the group of the Northern Italians but everyone migrates from the group of the Southern Italians at the maximal Nash equilibrium.

More generally, numerical simulations for various network structures show that higher number of social connections between countries leads to higher migration flow. Moreover, even a small increase in the number of these connections may produce a large increase in migration, at least as soon as certain conditions on the network structure and/or initial employment states in the two countries are fulfilled.

The results of the paper demonstrate the importance of the structure of social relations in determining migration decisions and emphasize the role of initial connections between North Americans and Italians in explaining the subsequent variation in emigration from the north and from the south of Italy.

3.5 Appendix

To calculate the steady-state probability of employment, I analyze the *subdivided dynamics* of employment introduced in Calvó-Armengol and Jackson (2004). Consider *T-period subdivision* $M^T(g, \frac{p_a}{T}, \frac{p_b}{T})$ of the Markov process $M(g, p_a, p_b)$ for the arbitrary network structure g. Let P^T denote the matrix of transitions between different employment states under M^T. That is, $P^T_{ss'}$ is the probability that $s_t = s'$ conditional on $s_{t-1} = s$. It is easy to see that for large T,

$$
P^T_{ss'} = \begin{cases}
\frac{p_b}{T} + o(\frac{1}{T^2}) & \text{if } s \text{ and } s' \text{ are adjacent and } s(i) > s'(i) \text{ for some } i \\
\frac{p_i(s)}{T} + o(\frac{1}{T^2}) & \text{if } s \text{ and } s' \text{ are adjacent and } s(i) < s'(i) \text{ for some } i \\
o(\frac{1}{T^2}) & \text{if } s \text{ and } s' \text{ are non-adjacent and } s \neq s' \\
1 - \#s\frac{p_b}{T} - \sum\limits_{i \text{ s.t. } s(i)=0} \frac{p_i(s)}{T} + o(\frac{1}{T^2}) & \text{if } s = s'
\end{cases}
$$

where $\#s$ denotes the number of agents who are employed in state s, $\#s = \sum\limits_k s(k)$, and $p_i(s)$ is the probability that agent i who is *unemployed* in state s hears about a new job offer and at most ones:

$$
p_i(s) = p_a + p_a \cdot \sum\limits_{j \text{ s.t. } s(j)=1,\, g_{ij}=1} \frac{1}{\sum\limits_{k \text{ s.t. } s(k)=0} g_{jk}},
$$

The definition of $P^T_{ss'}$ suggests that Markov process $M^T(g, \frac{p_a}{T}, \frac{p_b}{T})$ is a Poisson process. In particular, when T is high enough, the probability of even a single shock to employment state, s, at every subperiod is very low (of order $\frac{1}{T}$). The probability of two or more shocks is even lower (of order $\frac{1}{T^2}$ or lower). Therefore, instead of $M^T(g, \frac{p_a}{T}, \frac{p_b}{T})$, I consider an *approximate* Markov process $\widehat{M}^T(g, \frac{p_a}{T}, \frac{p_b}{T})$ where only one-shock transitions are retained while the transitions involving two or more shocks are disregarded. Matrix $\widehat{P}^T$ of transitions under $\widehat{M}^T$ is defined as

$$
\widehat{P}^T_{ss'} = \begin{cases}
\frac{p_b}{T} & \text{if } s \text{ and } s' \text{ are adjacent and } s(i) > s'(i) \text{ for some } i \\
\frac{p_i(s)}{T} & \text{if } s \text{ and } s' \text{ are adjacent and } s(i) < s'(i) \text{ for some } i \\
0 & \text{if } s \text{ and } s' \text{ are non-adjacent and } s \neq s' \\
1 - \#s\frac{p_b}{T} - \sum\limits_{i \text{ s.t. } s(i)=0} \frac{p_i(s)}{T} & \text{if } s = s'
\end{cases}
$$

$$(3.21)$$

In the following, I calculate the stationary probability distribution of employment using the approximate Markov process, $\widehat{M}^T$, instead of Markov process M^T. This substitution is justified since for large enough values of T, transition probabilities of the approximate Markov process, $\widehat{M}^T$, are close to those of the true Markov process, M^T. As a result,

the stationary probability distributions under M^T and under $\widehat{M}^T$ are also close, which is proved formally in Calvó-Armengol and Jackson (2004):

$$\lim_{T \to \infty} \mu^T = \lim_{T \to \infty} \hat{\mu}^T$$

where μ^T is the steady-state distribution of employment under M^T and $\hat{\mu}^T$ is the steady-state distribution of employment under $\widehat{M}^T$.

Proof of Lemma 3.2

To simplify the notation, below I use a for $\frac{p_a}{T}$ and b for $\frac{p_b}{T}$.

The steady-state probability of employment of any agent i in the complete network of size n can be defined as:

$$\theta(n) = \sum_{s \in \{0,1\}^n \text{ s.t. } s(i)=1} \hat{\mu}^T(s) = \sum_{k=1}^{n} \binom{n-1}{k-1} \hat{\mu}^T(s : \#s = k), \tag{3.22}$$

where $\hat{\mu}^T(s : \#s = k)$ is the steady-state probability of employment state s, such that exactly k agents are employed. To derive $\hat{\mu}^T(s : \#s = k)$ for any $k \in 0 : n$ I use the definition of transition probabilities, $\widehat{P}^T_{ss'}$, in (3.21) applied to the case of a complete network:[25]

$$\widehat{P}^T_{ss'} = \begin{cases} b & \text{if } s \text{ and } s' \text{ are adjacent and } s(i) > s'(i) \text{ for some } i \\ a(1 + \frac{\#s}{n-\#s}) & \text{if } s \text{ and } s' \text{ are adjacent and } s(i) < s'(i) \text{ for some } i \\ 0 & \text{if } s \text{ and } s' \text{ are non-adjacent and } s \neq s' \\ 1 - \#sb - an\mathbf{1}_{\{\#s \leq n-1\}} & \text{if } s = s' \end{cases} \tag{3.23}$$

(3.23) enables representation of $\hat{\mu}^T$ in the following form:

$$\begin{aligned}
\hat{\mu}^T(s' : \#s' = k) &= ak\left(1 + \frac{k-1}{n-k+1}\right) \hat{\mu}^T(s : s(-i) = s'(-i) \text{ and } s(i) < s'(i)\mathbf{1}_{\{k \geq 1\}}) + \\
&+ b(n-k)\hat{\mu}^T(s : s(-i) = s'(-i) \text{ and } s(i) > s'(i)\mathbf{1}_{\{k \leq n-1\}}) + \\
&+ (1 - bk - an\mathbf{1}_{\{k \leq n-1\}})\hat{\mu}^T(s' : \#s' = k), \quad k \in 0 : n
\end{aligned} \tag{3.24}$$

The first term in the sum on the right-hand side of (3.24) corresponds to a change in the employment status of agent i from unemployment in s, $s(i) = 0$, to employment in s', $s'(i) = 1$. There are k such agents and each of them may find a job either directly, with probability a, or using the information about a job offer from one of his $k - 1$ employed

---

[25]The derivation of the expression for $\hat{\mu}^T$ in this proof is based on the unpublished manuscript by A. Calvó-Armengol (2007).

friends, with probability $a\frac{k-1}{n-k+1}$, since every friend chooses i as a job offer recipient randomly among $n-k+1$ unemployed candidates.

The second term on the right-hand side of (3.24) corresponds to a change in the employment status of agent i from employment in s to unemployment in s'. There are $n-k$ such agents and each of them looses a job at exogenous rate b.

Finally, the third term in the sum corresponds to the case when none of the agents changes his employment status.

Cancelling the term $\hat{\mu}^T(s' : \#s = k)$ on the left- and on the right-hand side of (3.24) results in the system of equations (3.25):

$$
\begin{aligned}
(a\mathbf{1}_{\{k \leq n-1\}}n \;+\; bk)\hat{\mu}^T(s' : \#s' = k) =& \\
=&\; ak\left(1 + \frac{k-1}{n-k+1}\right)\hat{\mu}^T(s : s(-i) = s'(-i) \text{ and } s(i) < s'(i)) \\
+&\; b(n-k)\hat{\mu}^T(s : s(-i) = s'(-i) \text{ and } s(i) > s'(i)), \quad k \in 0 : n \quad (3.25)
\end{aligned}
$$

To solve this system, I first rewrite it in terms of probabilities μ_k, $k \in 0 : n$, that k out of n agents are employed at the steady state. Due to interchangeability of nodes in the complete network,

$$
\mu_k = \binom{n}{k}\hat{\mu}^T(s : \#s = k), \quad k \in 0 : n \tag{3.26}
$$

Then (3.25) reduces to a system of $n+1$ equations with $n+1$ unknowns, $\mu_0, \ldots, \mu_n$:

$$
(an\mathbf{1}_{\{k \leq n-1\}} + bk)\mu_k = \mathbf{1}_{\{k \geq 1\}}an\mu_{k-1} + \mathbf{1}_{\{k \leq n-1\}}b(k+1)\mu_{k+1}, \quad k \in 0 : n
$$

The solution can be found recursively:

$$
\begin{aligned}
\mu_{n-1} &= \frac{b}{a}\mu_n \\
\mu_{n-2} &= \frac{b}{a}\frac{n-1}{n}\mu_{n-1} \\
\mu_{n-3} &= \frac{b}{a}\frac{n-2}{n}\mu_{n-2} \\
&\;\;\vdots \\
\mu_1 &= \frac{b}{a}\frac{2}{n}\mu_2 \\
\mu_0 &= \frac{b}{a}\frac{1}{n}\mu_1
\end{aligned}
$$

Hence,

$$\mu_{n-l} = \left[\frac{b}{a}\right]^l \frac{(l-1)!\binom{n-1}{l-1}}{n^{l-1}}\mu_n = \left[\frac{b}{a}\right]^l \frac{(n-1)!}{n^{l-1}(n-l)!}\mu_n, \qquad 1 \le l \le n \qquad (3.27)$$

Using (3.27), probability μ_n can be found from the normalization condition, $\sum_{k=0}^n \mu_k = 1$. We have:

$$
\begin{aligned}
1 = \sum_{k=0}^n \mu_k &= \sum_{l=0}^n \mu_{n-l} = \sum_{l=0}^n \left[\frac{b}{a}\right]^l \frac{(n-1)!}{n^{l-1}(n-l)!}\mu_n = \mu_n\left[1 + \sum_{l=1}^n \left[\frac{b}{a}\right]^l \frac{(n-1)!}{n^{l-1}(n-l)!}\right] = \\
&= \mu_n\left[1 + \frac{b}{a} + \sum_{l=2}^n \left[\frac{b}{a}\right]^l \frac{(n-1)\cdots(n-(l-1))}{n^{l-1}}\right] = \\
&= \mu_n\left[1 + \frac{b}{a} + \sum_{l=2}^n \left[\frac{b}{a}\right]^l \left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{p-1}{n}\right)\right] = \\
&= \mu_n\left[1 + \frac{b}{a} + \frac{b}{a}\sum_{l=2}^n \prod_{k=1}^{l-1} \frac{b}{a}\left(1 - \frac{k}{n}\right)\right] = \\
&= \mu_n\left[1 + \frac{b}{a} + \frac{b}{a}\sum_{l=1}^{n-1} \prod_{k=1}^{l} \frac{b}{a}\left(1 - \frac{k}{n}\right)\right]
\end{aligned}
$$

So,

$$\mu_n = \frac{1}{1 + \frac{b}{a} + \frac{b}{a}\sum_{l=1}^{n-1}\prod_{k=1}^{l}\frac{b}{a}\left(1 - \frac{k}{n}\right)}, \qquad n \ge 1$$

Now, the closed-form expressions for the remaining n probabilities, $\mu_0, \ldots, \mu_{n-1}$ follow from (3.27). At last, plugging μ_k, $k \in 0 : n$, into (3.26) and using the obtained sequence of $\hat{\mu}^T(s : \#s = k)$, $k \in 0 : n$, in (3.22), leads to the following expression for the steady-state probability of employment:

$$
\begin{aligned}
\theta(n) &= \sum_{k=1}^n \frac{k}{n}\mu_k = \sum_{l=0}^{n-1} \frac{n-l}{n}\mu_{n-l} = \sum_{l=0}^{n-1} \left[\frac{b}{a}\right]^l \frac{(n-1)!}{n^l(n-l-1)!}\mu_n = \\
&= \mu_n\left[1 + \sum_{l=1}^{n-1}\left[\frac{b}{a}\right]^l \frac{(n-1)\cdots(n-l)}{n^l}\right] = \mu_n\left[1 + \sum_{l=1}^{n-1}\left[\frac{b}{a}\right]^l\left(1 - \frac{1}{n}\right)\cdots(1 - \frac{l}{n})\right] = \\
&= \mu_n\left[1 + \sum_{l=1}^{n-1}\prod_{k=1}^{l}\frac{b}{a}\left(1 - \frac{k}{n}\right)\right] = \frac{1 + \sum_{l=1}^{n-1}\prod_{k=1}^{l}\frac{b}{a}\left(1 - \frac{k}{n}\right)}{1 + \frac{b}{a} + \frac{b}{a}\sum_{l=1}^{n-1}\prod_{k=1}^{l}\frac{b}{a}\left(1 - \frac{k}{n}\right)} = \\
&= \frac{1 + \sum_{l=1}^{n-1}\prod_{k=1}^{l}\frac{p_b}{p_a}\left(1 - \frac{k}{n}\right)}{1 + \frac{p_b}{p_a} + \frac{p_b}{p_a}\sum_{l=1}^{n-1}\prod_{k=1}^{l}\frac{p_b}{p_a}\left(1 - \frac{k}{n}\right)}, \qquad n \ge 1
\end{aligned}
$$

Given $\theta(n)$, it is now straightforward to show that $\theta(n)$ is strictly increasing in n for any $n \ge 1$. Reducing $\theta(n)$ and $\theta(n+1)$ to a common denominator and subtracting $\theta(n)$ from $\theta(n+1)$, we obtain a ratio where denominator is unambiguously positive and the

nominator is equal to

$$\left(1 + \sum_{l=1}^{n} \prod_{k=1}^{l} \frac{p_b}{p_a} \left(1 - \frac{k}{n+1}\right)\right) \left(1 + \frac{p_b}{p_a} + \frac{p_b}{p_a} \sum_{l=1}^{n-1} \prod_{k=1}^{l} \frac{p_b}{p_a} \left(1 - \frac{k}{n}\right)\right) -$$

$$- \left(1 + \sum_{l=1}^{n-1} \prod_{k=1}^{l} \frac{p_b}{p_a} \left(1 - \frac{k}{n}\right)\right) \left(1 + \frac{p_b}{p_a} + \frac{p_b}{p_a} \sum_{l=1}^{n} \prod_{k=1}^{l} \frac{p_b}{p_a} \left(1 - \frac{k}{n+1}\right)\right)$$

Simple algebra transforms this expression to

$$\prod_{k=1}^{n} \frac{p_b}{p_a} \left(1 - \frac{k}{n+1}\right) + \sum_{l=1}^{n-1} \left[\frac{p_b}{p_a}\right]^{l} \left[\prod_{k=1}^{l} \left(1 - \frac{k}{n+1}\right) - \prod_{k=1}^{l} \left(1 - \frac{k}{n}\right)\right] > 0$$

Hence, $\theta(n+1) - \theta(n) > 0$ for any $n \geq 1$. In particular, as the size of the network, n, becomes arbitrarily large, the steady-state probability of employment approaches 1. ∎

3.6 Bibliography

Baines, D. (1994) *"European Emigration, 1815-1930: Looking at the Emigration Decision Again,"* Economic History Review, 47, pp. 525-544.

Baily, S. L. (1983), "Italian Immigrants in Buenos Aires and New York", *American Historical Review*, 88, pp. 281-305.

Böcker, A. (1994) *"Chain Migration over Legally Closed Borders: Settled Immigrants as Bridgeheads and Gatekeepers,"* The Netherlands Journal of Social Sciences, 30(2), pp. 87-106.

Boyd, M. (1989), *"Family and Personal Networks in International Migration: Recent Developments and New Agendas,"* International Migration Review, 23(3), pp. 638-670.

Boorman, S.A. (1975) *"A Combinatorial Optimization Model for Transmission of Job Information through Contact Networks,"* Bell Journal of Economics, 6(1), pp. 216-249.

Calvó-Armengol, A. (2007), unpublished manuscript.

Calvó-Armengol, A. and M.O.Jackson (2004) *"The Effects of Social Networks on Employment and Inequality,"* American Economic Review, 94(3), pp. 426-454.

Dahya, B. (1973) *"Pakistanis in Britain: Transients or Settlers,"* Race, 14, pp. 241-247.

Faist, T. (2000) *"The Volume and Dynamics of International Migration and Transnational Social Spaces,"* Oxford: Oxford University Press.

Gilani, I., Khan, F. M. and M. Iqbal (1981) *"Labor Migration from Pakistan to the Middle East and Its Impact on the Domestic Economy,"* Research Report Series, nos. 126 and 127, Islamabad: Pakistan Institute of Development Economics.

Granovetter, M. (1973) *"The Strength of Weak Ties,"* American Journal of Sociology, 78(6), pp. 1360-80.

Granovetter, M. (1995) *"Getting a job: A study of Contacts and Careers,"* 2nd Ed. Chicago: University of Chicago Press.

Gurak, D.T. and F. Caces (1992) *"Migration Networks and the Shaping of Migration Systems,"* in International Migration Systems: A Global Approach, eds. Kritz, M.M., Lim, L.L. and Zlotnik, H., Oxford: Clarendon Press, pp. 150-176.

Harris, J. R. and M. P. Todaro (1970) *"Migration, Unemployment, and Development: A Two-Sector Analysis,"* American Economic Review, 60, pp. 126-142.

Hatton, T.J. and J. G. Williamson (1998) *"The Age of Mass Migration: Causes and Economic Impact,"* New York: Oxford University Press.

Hily, M.A. and M. Poinard (1987) *"Portuguese Associations in France,"* in Immigrant Associations in Europe, eds. Rex, J., Joly, D. and C. Wilpert, Aldershot, England: Gower, pp. 126-165.

Jackson, M.O. (2006) *"The Economics of Social Networks,"* Chapter 1 in Volume I of: Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society, eds. Blundell R., Newey W., and T. Persson, Cambridge University Press.

Joly, D. (1987) *"Associations Amongst the Pakistani Population in Britain,"* in Immigrant Associations in Europe, eds. Rex, J., Joly, D. and C. Wilpert, Aldershot, England: Gower, pp.62-85.

Josephides, S. and J. Rex (1987) *"Asian and Greek Cypriot Associations and Identity,"* in Immigrant Associations in Europe, eds. Rex, J., Joly, D. and C. Wilpert, Aldershot, England: Gower, pp.11-41.

Klein, H. S. (1983) *"The Integration of Italian Immigrants into the United States and Argentina: A Comparative Analysis,"* American Historical Review, 88, pp. 306-29.

Lewis, W. A. (1954) *"Economic Development with Unlimited Supplies of Labor,"* The Manchester School of Economic and Social Studies, 22, pp. 139-191.

Lowell, B. L. (1987) *"Scandinavian Exodus: Demography and Social Development of 19th Century Rural Communities,"* Boulder, Col: Westview Press.

Mahmood R. A. (1991) *"Bangladesh Return Migrants from the Middle East: Process, Achievement, and Adjustment,"* in Migration to the Arab World, Experience of Returning Migrants, ed. Gunatilleke, G., The United Nations University, Tokyo, Japan.

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A. and J. E. Taylor (1993) *"Theories of International Migration: A Review and Appraisal,"* Population and Development Review, 19(3), pp. 431-466.

Massey, D.S. and K. Espinosa (1997) *"Whats Driving Mexico-U.S. Migration? A Theoretical, Empirical and Policy Analysis,"* American Journal of Sociology, 102(4), pp. 939-999.

Nair, E. R.G. (1991) *"Asian Migration to the Arab World: Kerala (India),"* in Migration to the Arab World, Experience of Returning Migrants, ed. Gunatilleke, G., The United Nations University, Tokyo, Japan.

Ranis, G. and J. C. H. Fei (1961) *"A Theory of Economic Development,"* American Economic Review, 51, pp. 533-565.

Sjaastad, L.A. (1962) *"The Costs and Returns of Human Migration,"* Journal of Political Economy, 70S, pp. 80-93.

Todaro, M. P. (1969) *"A Model of Labor Migration and Urban Unemployment in Less Developed Countries,"* American Economic Review, 59, pp. 138-148.

Todaro, M. P. (1976) *"Internal Migration in Developing Countries,"* Geneva: International Labor Office.

Todaro, M. P. (1989) *"Economic Development in the Third World,"* New York: Longman.

Todaro, M. P. and L. Maruszko (1987) *"Illegal Migration and US Immigration Reform: A Conceptual Framework,"* Population and Development Review, 13, pp. 101-114.

Walker, M. (1964) *"Germany and the Emigration, 1816-1885,"* Cambridge, Mass: Harvard University Press.

Wilpert, E.C. (1988) *"Work and the Second Generation: the Descendants of Migrant Workers in the Federal Republic of Germany,"* in Entering the Working World, ed. Wilpert, E.C., Aldershot: Gower, pp. 111-149.