Preliminary and Incomplete

GRAVITY AND MATCHING*

Johannes Moenius **
University of Redlands

James E. Rauch †
University of California, San Diego & NBER

Vitor Trindade ‡
University of Missouri, Columbia

November 14, 2007

We develop a gravity model of international trade in which border effects, impacts of migrants, and effects of past trading relationships are all based in networks of entrepreneurs. In our model workers leave their former employers to become entrepreneurs, and found new firms by partnering with former colleagues or with workers who left a different employer. In the absence of migration, the first type of partnership creates trade only within a country and the second type creates trade both within and across countries, so that total trade displays country border effects. Migrants, however, can match with former colleagues across country boundaries. Similarly, past trading relationships facilitate search for partners across country boundaries. It is shown that this model generates a decomposition of bilateral trade into number of partnerships or matches and value per match. Standard gravity model variables are shown to affect number of matches and value per match differently; distance, for example, is predicted to decrease number of matches but leave value per match unaffected. Following Besedes and Prusa (2006), who count "relationships" between the United States and its trading partners by the number of product varieties for which positive trade is observed, we use these "links" and "value per link" as our empirical proxies for number of matches and value per match. Preliminary estimates using OECD data on trade and migration, both within the OECD and between the OECD and non-OECD countries, and U.S. Department of Commerce trade data, support the sharpest predictions of the theory for the impacts on number of links and value per link of distance, migrants, colonial ties, and the interactions between them.

---

* Our thanks to seminar participants at University of British Columbia, University of Toronto, University of Redlands, University of California, San Diego, and University of Michigan, and to Jennifer Poole for excellent research assistance. We also thank Jonathan Eaton, Samuel Kortum, and Francis Kramarz for kinding supplying their firm-level trade data for France. We are responsible for any errors.
** School of Business, University of Redlands, Redlands, CA 92373-0999.
Email: johannes_moenius@redlands.edu. Tel: (909) 748-8779.
† Department of Economics, University of California, San Diego, La Jolla, CA 92093-0508.
Email: jrauch@weber.ucsd.edu. Tel: (858) 534-2405.
‡ Department of Economics, University of Missouri, Columbia, MO 65211.
Email: trindadev@missouri.edu. Tel: (573) 882-9925.

1. Introduction

A common feature of many varieties of social networks is what we call a "cluster and bridge" structure, in which groups of densely tied agents (clusters) are connected by sparse ties (bridges), as opposed to either completely isolated groups or a uniform density of ties among all agents.[1] Narrowing our focus to social networks in which the agents are engaged in economic activity, we can identify a second common feature: higher rewards to agents whose ties span clusters than to agents whose ties are confined within a given cluster. Evidence has accumulated in diverse settings that agents with bridge ties perform better than agents with cluster ties (see Burt 2000 for a survey): firms that bridge clusters in interfirm networks show higher profits, managers that bridge clusters in intrafirm networks receive higher pay and more rapid promotions. This may be due to opportunities for arbitrage ("brokerage") across clusters of differences in information or resources, or could reflect selection of the most able agents into bridging positions. In our model below a combination of gains from trade and selection will be at work.

Let us think of the links in an economic network displaying these two common features as representing flows of goods, and arbitrarily divide up the network by drawing political boundaries. If the boundaries tend not to intersect clusters, flows of goods will be more dense within boundaries than across them. In other words, the flows of goods in a cluster and bridge network will exhibit "border effects" of the type first identified by McCallum (1995). Note that the cost of crossing these boundaries is not relevant to the argument, and indeed border effects are found even for political boundaries that impose no apparent costs on trade, such as those

[1] Jackson and Rogers (2005) list the relevant references and also develop an abstract model of network formation that displays this feature. Their "islands" play the same role as our "parent firms" below.

between U.S. states (Wolf 2000). Moreover, the agents involved in facilitating the flows of goods along bridge links will tend to be more successful than the agents who facilitate the flows of goods along cluster links. Since bridge links are much more likely than cluster links to cross international borders, this common feature of economic networks fits well with robust findings that exporting firms are larger and have higher productivity than non-exporting firms.[2] The stylized facts of international trade thus match the stylized facts of socioeconomic networks.

In this paper we develop a model of network formation that delivers these stylized facts and also generates new insights regarding the gravity model of international trade.[3] These insights will concern the roles of distance, colonial ties/past trade, and migrants, the latter two of which have been treated in an ad hoc manner in the literature to date. Our new predictions receive support from data on internal and external OECD trade and US trade.

2. Model intuition

Workers turned entrepreneurs will be the agents of network formation in our model. Clusters will form among entrepreneurs who spin off from a common "parent firm." Having already worked together, such entrepreneurs know each other's capabilities and needs and are thus at least weakly tied at "birth."[4] It is thus relatively easy for them to form partnerships with each other or do business with each other as independent firms.

It is widely recognized that spinoffs or "entrepreneurial spawning" are a major source of entrepreneurship. Bhide (2000, p. 94) reports that 71 percent of the firms in the Inc 500 (a list of

---

[2] Fafchamps et al. (forthcoming) list the relevant references, and also report that Moroccan exporting firms were typically exporters at start-up or very soon thereafter. This is consistent with our model below, in which firms are born through the formation of either bridge ties or cluster ties.

[3] Some of the welfare and policy implications of the model are worked out in Rauch and Watson (2007).

[4] For a definition of "weak ties" along these lines see Rauch (2001, p. 1179).

young, fast-growing firms) were founded by entrepreneurs who "replicated or modified an idea encountered through previous employment." This process has been especially well studied in the high-tech, venture capital context, where the classic example is the spinoffs from Fairchild Semiconductor in Silicon Valley (Braun and Macdonald 1982). Gompers et al. (2003, p. 3) explain the fertility of this process as follows: "Working in such firms exposes would-be entrepreneurs to a network of suppliers of labor, goods, and capital, as well as a network of customers. Because starting a new venture requires suppliers and customers to make relationship-specific investments before it is guaranteed that the venture will get off the ground, networks can be particularly useful in alleviating this chicken-and-egg problem." They report that the share of U.S. venture-capital backed entrepreneurs in the period 1986-99 who previously worked for publicly traded firms is around 45 percent.

There is no need to appeal to a high-tech, venture-capital backed environment to explain entrepreneurial spawning, however. It is also generated by a more mundane process of "client-based entrepreneurship" (Rauch and Watson 2003), in which employees try to wrest the value of client relationships from their employers by setting up their own firms and taking their clients with them. This can occur in any industry in which client relationships are important, including manufacturing, business services, and personal services. According to the 1992 Economic Census of the United States (U. S. Department of Commerce 1997, p. 86), 45.1 percent of nonminority male business owners "previously worked for a business whose goods/services were similar to those provided by the [current] business."

In our model, workers leaving their firms to become entrepreneurs may take the relatively easy avenue of forming partnerships with their former colleagues, or might at greater cost seek partnerships with unknown workers leaving other firms to become entrepreneurs. Those that c lsucceed in the latter endeavor form bridges, whereas those that do not form clusters. Specifically, we assume that a cluster partnership serves as the fallback option when deciding whether to accept a potential bridge partnership. Selection then ensures that accepted bridge partnerships will be of higher quality and thus perform better on average than cluster partnerships, and the extent to which this is true will increase with the average of the quality of potential bridge partnerships relative to cluster partnerships, representing the potential for gains from trade.

Entrepreneurs tend to form their firms in the communities in which they live.[5] A cluster consisting of entrepreneurs who spun off from a common parent firm will therefore tend to be geographically localized. This tendency connects our model of entrepreneurial network formation to border effects in interregional and international trade. Our model will display country border effects because cluster partnerships are formed only within countries whereas bridge partnerships are formed both within and across countries.

We will impose the assumption of random matching on the search technology employed by entrepreneurs seeking bridge partnerships. Output generated by bridge partnerships between any two countries will then be proportional to the product of their economic masses, the "gravity" relationship that is another robust feature of international trade patterns. In this respect, random matching plays the role in our model that CES preferences play in standard trade models.

---

[5] Indeed, we conjecture that a worker who leaves his employer to become an entrepreneur is more likely to stay in the same community than a worker who leaves his employer for another employer. This conjecture receives some support from Michelacci and Silva (2007), who find that in both the Italy and United States the fraction of entrepreneurs who start up their business in the region where they were born is significantly higher than the fraction of workers employed in the region where they were born.

When international bridge partners repatriate their profits they generate bilateral trade. This bilateral trade is proportional to the product of the number of matches between the two countries times the average value per match. To see the new empirical predictions that this decomposition yields, we need to formalize the model we have sketched so far.

3. Model specifics

We study a multi-country overlapping generations model. There are I countries in the world. In every country i a new generation of risk-neutral agents with mass N_i is born in each period and lives until the end of the next period. Each young agent supplies one unit of labor. Old agents either supply one unit of labor or manage firms. Specifically, a firm will be owned by two old agents who have formed a partnership. Each partnership will be distinguished by a match quality, denoted by $z \in [z_0,\infty)$. A firm needs to hire labor in order to actively engage in production, with output being a function of the quality of the partners' match and the labor employed, as in Rauch and Trindade (2003). The effectiveness of the match in contributing to production is assumed to be diminishing in the distance d_{rs} between partners r and s with a constant elasticity β. This reflects difficulties with communication and transportation. We assume the production function F:

$$y_{rs} = F\left(x, zd_{rs}^{-\beta}\right), \qquad \textbf{(1)}$$

where x is the mass of young workers hired and y_{rs} is the output of the firm formed by agents r and s. The function F exhibits constant returns to scale.

Each firm maximizes profits under perfect competition, taking both the price of output and the wage rate w as given. Output is our numeraire. With constant returns to scale, it is easy to show that total profits of firm rs with match quality z can be written as

$$\Pi(d_{rs}, w, z) = z d_{rs}^{-\beta} \pi(w), \qquad \textbf{(2)}$$

where the function $\pi(w)$ is decreasing and convex in w. If the partners reside in different countries, they employ labor in the country where it is cheaper. Without loss of generality we will rank the countries in the order of their equilibrium wage rates from lowest to highest, so that $i < j$ implies $w_i < w_j$. It follows that a partnership formed between old agents from countries i and j, $i < j$, will employ labor in country i at wage w_i. We will denote the distance between these partners by d_{ij}, since in the empirical work below we will not be able to observe where an agent is located within a country. By the same token, the distance between any two partners within country i is denoted by d_{ii}.

At the end of a given period, when young agents employed as workers in existing firms are about to become old, each draws an opportunity to spin off and form his own firm with probability λ_i. We provisionally assume that this opportunity is always accepted; conditions that assure this outcome will be specified shortly. These nascent entrepreneurs engage in a matching process culminating in the formation of the firms that they will manage in the next period. An agent can match with someone in the same original firm as himself (thus forming by definition a cluster match) or with someone who is currently working in a different firm (forming a bridge match).

The matching process runs as follows. All spin-off entrepreneurs meet at a giant convention and match randomly once. By the law of large numbers, the number of matches (and therefore of potential partnerships) between agents in countries i and j is $\lambda_i N_i \lambda_j N_j / N$, where $N = \sum_i \lambda_i N_i$ is the world mass of entrepreneurs. When an agent finds a potential partner from another firm, their match quality is drawn according to the "bridge" cumulative distribution function η. Knowing this quality level, the agent and his potential partner then decide whether to form a firm.

If an agent fails to form a partnership with someone from another firm, then this agent freely obtains a cluster match with someone in his current firm. The quality of the cluster match is the constant z^C, and the distance between the two former colleagues is normalized to unity. After all matches are consummated, firms hire labor and engage in production.

We assume that firm partners divide their profits according to the Nash bargaining solution with equal bargaining weights. When they form a bridge partnership their outside options are to find partners within their own firms, and when they form cluster partnerships their outside options are zero.[6] In a cluster partnership the two partners are in symmetrical positions, implying that each receives half of the surplus created by their match. Thus, the value of a cluster match to an agent from country i is $z^C \pi(w_i)/2$. This constitutes the threat point for the agent when he obtains a match for a bridge partnership. Suppose that this agent is matched with an agent from country j, $i < j$. Given that the potential partnership yields the value $z d_{ij}^{-\beta} \pi(w_i)$, the two agents will form an actual firm if and only if $z d_{ij}^{-\beta} \pi(w_i) \geq \left(z^C/2\right)\left(\pi(w_i) + \pi(w_j)\right)$. Note that this implies a cut-off match quality $\underline{z}_{ij}$ such that only matches with $z \geq \underline{z}_{ij}$ between agents from countries and i and j result in a firm. The cut-off match qualities can be written as:

$$\underline{z}_{ij} = \left(z^C/2\right) d_{ij}^{\beta} \left(1 + \pi(W_{ij})/\pi(w_{ij})\right),$$

using the notation $w_{ij} \equiv \min(w_i, w_j)$ and $W_{ij} \equiv \max(w_i, w_j)$ (with $w_{ii} = W_{ii} \equiv w_i$).

[6] In the next paragraph we see that nothing would change if we allowed the outside option to be employment at the equilibrium wage.

We see that within any country cluster partnerships are the least profitable firms. To ensure that all spin-off opportunities are accepted, it is therefore sufficient that $z^C \pi(w_i)/2 > w_i$ for all i. This equilibrium condition is assured by choosing λ_i sufficiently small for all i, since this decreases demand for labor in every country, shrinking w_i and raising $\pi(w_i)$.

In any bridge partnership the profits of the partner from the high wage country will be repatriated, yielding the value of trade generated by any pair of internationally matched agents with match quality z, which we denote by $V_{ij}^B(z)$:

$$V_{ij}^B(z) = zd_{ij}^{-\beta}\pi(w_{ij})/2 - \left(z^C/4\right)\left(\pi(w_{ij}) - \pi(W_{ij})\right). \qquad \textbf{(3)}$$

This is simply equal to the value of his threat point plus half of the surplus created by the bridge match. The average value of trade generated by a successful bridge match is therefore:

$$V_{ij}^B = \frac{\int_{\underline{z}_{ij}}^{\infty} V_{ij}^B(z)\mathrm{d}\eta(z)}{\int_{\underline{z}_{ij}}^{\infty} \mathrm{d}\eta(z)}. \qquad \textbf{(4)}$$

In order to compute total bilateral trade, we then need to multiply equation (4) by the number of accepted bridge matches between countries i and j, which is given by

$$n_{ij} = (\lambda_i N_i \lambda_j N_j / N)\int_{\underline{z}_{ij}}^{\infty} \mathrm{d}\eta(z). \qquad \textbf{(5)}$$

Using equations (3), (4) and (5), total bilateral trade then equals:

$$V_{ij} = (\lambda_i N_i \lambda_j N_j / N)\left(1/2\right)\left\{\int_{\underline{z}_{ij}}^{\infty} z\, d_{ij}^{-\beta}\pi(w_{ij})d\eta(z) - \left(z^C/2\right)\left(\pi(w_{ij}) - \pi(W_{ij})\right)\left(1-\eta(\underline{z}_{ij})\right)\right\} \qquad \textbf{(6)}$$

The volume of domestic trade generated by bridge partnerships is given by equation (6) evaluated at $i = j$, but this does not equal the total volume of domestic trade because it does not include cluster partnerships. The probability that an agent from country i falls back on a cluster match is given by $\sum_j \lambda_j N_j \eta(\underline{z}_{ij}) \Big/ N$. The number of cluster partnerships in country i is then

$$\left(\lambda_i N_i/2\right)\sum_j \lambda_j N_j \eta(\underline{z}_{ij}) \Big/ N \,, \qquad \textbf{(7)}$$

where we have to divide by two because both agents in the partnership are from country *i*. Finally, using the convention that half the profits in any intranational partnership are "repatriated" to one of the partners, we have the volume of internal trade created by cluster partnerships in country *i*:

$$V_{ii}^{C} = \left(\lambda_i N_i/2\right)\left(z^{C}\pi(w_i)/2\right)\sum_j \lambda_j N_j \eta(\underline{z}_{ij}) \Big/ N \,.$$

Total domestic trade is then given by:

$$V_{ii} = \left(\lambda_i N_i/2\right)\left(z^{C}\pi(w_i)/2\right)\sum_j \lambda_i N_j \eta(\underline{z}_{ij}) \Big/ N + \left(\left(\lambda_i N_i\right)^2/N\right)\left(1/2\right)\int_{\underline{z}_{ii}}^{\infty} z\, d_{ii}^{-\beta}\pi(w_i)\, d\eta(z) \,. \qquad \textbf{(8)}$$

Figure 1 illustrates the network of trading relationships established by the workers-turned-entrepreneurs who spin off from one parent firm. Two-thirds of the agents have formed partnerships with their former colleagues and one-third have formed bridge ties. Since bridge ties are the result of random matching, they are formed in proportion to the sizes of the countries: country A (in which this parent firm is located) contains half of all entrepreneurs and thus receives half of all bridge ties, country B contains one-third of all entrepreneurs and receives one-third of all bridge ties, and country C receives the remaining one-sixth of all bridge ties.

To close our model we need to determine the wage rates that clear each country's labor market. We begin by computing labor demand generated by cluster partnerships in each country. Using equation (2), we derive the labor demand by a cluster firm in country *i* to be $-z^{C}\pi'(w_i)$. Multiplying this by the total number of cluster partnerships (equation 7), we obtain total labor demand by cluster firms in country *i*:

$$L_i^{C} = -z^{C}\pi'(w_i)\left(\lambda_i N_i/2\right)\sum_j \lambda_j N_j \eta(\underline{z}_{ij}) \Big/ N \,.$$

We must add to this the demand for country i's labor from bridge partnerships. Since a bridge partnership between country i and country j with $j \geq i$ creates demand $-zd_{ij}^{-\beta}\pi'(w_i)$ for country i labor, and a bridge partnership between country i and country j with $j < i$ creates no demand for country i labor, we can write total labor demand from bridge partnerships as:

$$L_i^B = -\pi'(w_i)\sum_{j\geq i}\left(\lambda_i N_i \lambda_j N_j / N\right)\int_{\underline{z}_{ij}}^{\infty} z\, d_{ij}^{-\beta}\, d\eta(z).$$

Adding labor demand from cluster partnerships to labor demand from bridge partnerships and equating to labor supply, we obtain I equations for labor market equilibrium (one per country):

$$-\pi'(w_i)\left\{z^C\left(\lambda_i N_i/2\right)\sum_j \lambda_j N_j \eta(\underline{z}_{ij}) + \sum_{j\geq i}\lambda_i N_i \lambda_j N_j \int_{\underline{z}_{ij}}^{\infty} z\, d_{ij}^{-\beta}\, d\eta(z)\right\} \Big/ N = N_i\left(2-\lambda_i\right).$$

These equations form a system of I equations in the I unknown wages. Their solution is not trivial since the cut-offs $\underline{z}_{ij}$ depend on the wages. However, since in this paper we are not interested in the impact of trade on wages, we can avail ourselves of a useful simplification. Note that every country's labor demand contains a term from its own bridge partnerships, $\left(-\pi'(w_i)\left(\lambda_i N_i\right)^2/N\right)\int_{\underline{z}_{ii}}^{\infty} z\, d_{ii}^{-\beta}\, d\eta(z)$, which depends only on w_i. We can ensure that this term dominates the determination of wages by assuming that $d_{ij} >> d_{ii}$ for all i and any $j \neq i$, with β sufficiently large and z^C sufficiently small. Country i's wage ranking then increases with λ_i and with N_i and decreases with d_{ii}, and this ranking is unaffected by international trade.

4. Extending the model to colonial ties and international migrants

Head, Mayer, and Ries (2007) use bilateral trade data from 1948 to 2003 to examine the effect of independence on post-colonial trade. They find an immediate reduction of trade between the colony and former colonial power that they attribute to elimination of preferential access, followed by a steady decline that they attribute to the decay of trading networks. Our results in Rauch and Trindade (2002) also support a network interpretation of the impact of past colonial ties on international trade. One natural way to incorporate this into our model is to allow the current generation of entrepreneurs to inherit some knowledge from the experience of the previous generation that could be useful in the matching process. In particular, colonial ties could be modeled as a stochastic dominant shift in η, the distribution of bridge match quality, for the relevant country pair. This inherited knowledge could be allowed to decay to zero over the length of a generation, so that each generation benefits only from the experience of the generation immediately preceding it. The stochastic dominant shift in η would of course increase both the number of matches and the average match value for the relevant country pair.

We also extend the model to the presence of migrants. We define a migrant as a worker who moves from one parent firm to another before becoming an entrepreneur who spins off from the second parent firm. A migrant is both disadvantaged and advantaged in the matching process. Having spent less time working at his (second) parent firm, we suppose he has less to offer potential partners. Specifically, we assume that the bridge match quality for a potential partnership involving a migrant is drawn from a distribution η^m that is first-order stochastically dominated by η, and that the quality of a cluster match involving a migrant and a former colleague from his second parent firm is $z^{Cm} < z^C$. At the same time, a migrant is also weakly tied to his former colleagues from his original parent firm, so we assume they provide an alternative fallback option for him. This fallback option consists of his ability to use his familiarity with his former colleagues to pick the maximum among them from his bridge distribution, which we will denote by z^m. As with a cluster match, the fallback option for each party is zero, hence each earns:

$$\Pi_{ij}^m / 2 = z^m d_{ij}^{-\beta} \pi(w_{ij}) / 2, \tag{9}$$

where *i* and *j* are the origin and host countries, respectively. For simplicity we assume $z^m d_{ij}^{-\beta} > z^{Cm}$ for all country pairs, so that a migrant always prefers to find a partner from his country of origin if his bridge match fails.

Let us denote the measure of old agents who migrated from country *i* to country *j* when young by M_{ij}. We assume that their migration was exogenous.[7] We could then compute the number of migrants whose random bridge matches succeed. That will only happen to those migrants who draw bridge matches with high quality relative to z^m, access to cheap labor, and short distances relative to the distance to the migrants' communities of origin. In order to save space we shall assume that these events are sufficiently rare that we can ignore them, and simply treat migrants as though all of their potential bridge matches fail in favor of matches with former colleagues from their original parent firms.[8]

With this simplification in place, the probability that a non-migrant from country *j* will draw a bridge match quality from the distribution η with a potential partner from country *i* is given by $\lambda_i N_i^B / N$, where:

$$N_i^B = N_i - M_{ii} - \sum_{j \neq i} M_{ji} . \tag{10}$$

[7] Interestingly, empirical studies that have instrumented for migration have found *larger* effects of migration on trade (Konecny 2007) and FDI (Javorcik et al. 2006).
[8] The main practical consequence of this simplification is that it eliminates any dependence on distance to their origin countries of the impact of migrants on the number of matches between their host and origin countries. We will need to be aware of this in the empirical work below.

In this expression N_i is the measure of old agents resident in country *i*, M_{ii} is the measure of old agents in country *i* who migrated internally, and $\sum_{j \neq i} M_{ji}$ is the measure of old foreign-born in country *i*. The total number of partnerships between country *i* and country *j* is then given by:

$$n_{ij} = (\lambda_i N_i^B \lambda_j N_j^B / N) \int_{\underline{z}_{ij}}^{\infty} \mathrm{d}\eta(z) + \left(\lambda_j M_{ij} + \lambda_i M_{ji}\right), \tag{11}$$

If countries *i* and *j* share a colonial tie, we replace η in this expression with η', where η' first-order stochastically dominates η.

Using equation (11) in combination with equations (3), (4), and (9), we can derive the equivalent of equation (6), the total value of bilateral trade, for our expanded model:

$$\begin{aligned} V_{ij} = (\lambda_i N_i^B \lambda_j N_j^B / N)(1/2)\Big\{\int_{\underline{z}_{ij}}^{\infty} z\, d_{ij}^{-\beta} \pi(w_{ij}) d\eta(z) - \left(z^C/2\right)\left(\pi(w_{ij}) - \pi(W_{ij})\right)\left(1 - \eta(\underline{z}_{ij})\right)\Big\} \\ + z^m d_{ij}^{-\beta} \pi(w_{ij}) \left(\lambda_j M_{ij} + \lambda_i M_{ji}\right)/2, \quad i \neq j. \end{aligned} \tag{12}$$

Again, if countries *i* and *j* share a colonial tie, we replace η in this equation with η'. Finally, the average value per match (partnership) between countries *i* and *j* is computed by simply dividing equation (12) by equation (11).

To complete the analysis, we must derive the equivalent of equation (8), the total value of intranational trade, for our expanded model:

$$\begin{aligned} V_{ii} = \left(\lambda_i N_i^B/2\right)\left(z^C \pi(w_i)/2\right)\left(\sum_j \lambda_j \left(N_j^B \eta\left(\underline{z}_{ij}\right) + \sum_k M_{kj}\right)\right) \Big/ N \\ + \left(\left(\lambda_i N_i^B\right)^2 / N\right)(1/2) \int_{\underline{z}_{ii}}^{\infty} z\, d_{ii}^{-\beta} \pi(w_i) d\eta(z) + z^m d_{ii}^{-\beta} \pi(w_i) \lambda_i M_{ii}. \end{aligned} \tag{13}$$

We shall make use of equations (11), (12) and (13) in the empirical analysis, to which we turn next.

5. Empirical strategy

In order to implement equations (11) - (13) empirically we need to specify a distribution of match quality η. We shall assume a Pareto distribution, $\eta(z) = 1 - (z/z_0)^{-1/\theta}$, where $\theta < 1$ is a "shape" parameter and we recall that z_0 is the lower bound of the distribution of match qualities. Colonial ties will be modeled as an increase in θ for the relevant country pair, generating a first-order stochastic dominant shift in the distribution of match qualities.

Using equation (11), we can now write the number of matches as:

$$n_{ij} = (\lambda_i N_i^B \lambda_j N_j^B / N)(\underline{z}_{ij}/z_0)^{-1/\theta} + (\lambda_j M_{ij} + \lambda_i M_{ji}).$$

The impact of an increase in the mass of migrants M_{ij} on n_{ij} is

$\partial n_{ij}/\partial M_{ij} = \lambda_j - \lambda_i \lambda_j N_i^B \left[(z^C/2) d_{ij}^{\beta}(1 + \pi(W_{ij})/\pi(w_{ij}))/z_0\right]^{-1/\theta}/N$, where $\underline{z}_{ij}$ has been replaced by its definition. This expression brings out two subtle points to look for in the empirical estimation. First, since $\underline{z}_{ij}/z_0 > 1$, colonial ties (an increase in θ) not only increase the number of matches due to non-migrants but also reduce the impact of migrants on the number of matches. Second, distance increases the impact of migrants on the number of matches.

To facilitate comparison with the standard gravity model we will log-linearize the number of matches above. We can write

$n_{ij} = (\lambda_i N_i^B \lambda_j N_j^B / N)(\underline{z}_{ij}/z_0)^{-1/\theta}\left[1 + (\lambda_j M_{ij} + \lambda_i M_{ji})(\underline{z}_{ij}/z_0)^{1/\theta} / (\lambda_i N_i^B \lambda_j N_j^B / N)\right]$. The term in brackets can be replaced using the approximation $1 + x \approx \exp(x)$ for small *x*. Substituting in the expression for $\underline{z}_{ij}$, we then take logarithms of both sides to obtain:

$$\log n_{ij} \approx (1/\theta)\log(2z_0/z^C) - \log N - (1/\theta)[\beta \log d_{ij} + \log(1+\pi(W_{ij})/\pi(w_{ij}))]$$
$$+ (\lambda_j M_{ij} + \lambda_i M_{ji}) N \left[z^C d_{ij}^{\beta}(1+\pi(W_{ij})/\pi(w_{ij}))/2z_0\right]^{1/\theta} / (\lambda_i N_i^B \lambda_j N_j^B) \tag{14}$$
$$+ \log(\lambda_i N_i^B) + \log(\lambda_j N_j^B).$$

It is now instructive, instead of using equation (12), to derive the volume of trade as the number of non-migrant matches times the average non-migrant trade per match plus the number of migrant matches times the average migrant trade per match. Using equation (4), the average value of trade per non-migrant match can be shown to equal

$$(z^C/4)(\theta\pi(w_{ij}) + (2-\theta)\pi(W_{ij}))/(1-\theta).$$

Note that this expression does not depend on distance. The intuition is that distance has two offsetting effects on average value per match: a negative effect for given match quality and a positive selection effect on match quality. [9] For the Pareto distribution these two effects exactly cancel. [10] For the volume of trade for $i \neq j$ we can now calculate:

$$V_{ij} = (\lambda_i N_i^B \lambda_j N_j^B / N)(\underline{z}_{ij}/z_0)^{-1/\theta}(z^C/4)(\theta\pi(w_{ij}) + (2-\theta)\pi(W_{ij}))/(1-\theta) + (\lambda_j M_{ij} + \lambda_i M_{ji}) z^m d_{ij}^{-\beta}\pi(w_{ij})/2.$$

We can then log-linearize this expression, employing the same manipulations that we used for equation (14), to obtain

$$\log V_{ij} \approx \log z_0/\theta - (1/\theta - 1)\log(z^C/2) - \log(2N) - (1/\theta)[\beta\log d_{ij} + \log(1+\pi(W_{ij})/\pi(w_{ij}))]$$
$$+ \log[(\theta\pi(w_{ij}) + (2-\theta)\pi(W_{ij}))/(1-\theta)] + \log(\lambda_i N_i^B) + \log(\lambda_j N_j^B) \tag{15}$$
$$+ \frac{(\lambda_j M_{ij} + \lambda_i M_{ji}) N z^m z_0^{-1/\theta}\pi(w_{ij})(z^C/2)^{1/\theta - 1} d_{ij}^{\beta(1/\theta - 1)}(1+\pi(W_{ij})/\pi(w_{ij}))^{1/\theta}}{(\lambda_i N_i^B)(\lambda_j N_j^B)(\theta\pi(w_{ij}) + (2-\theta)\pi(W_{ij}))/(1-\theta)}, \quad i \neq j.$$

---

[9] That is, firms that had to overcome greater distances will on average have greater "raw" productivity.

[10] However, average value per *attempted* bridge match does decline with distance because the proportion of accepted matches that are cluster matches rises, but we cannot observe this in the data.

We can see from equation (15) that the impact of migrants is again reduced by colonial ties (higher θ) and increased by distance.

Finally, we find the log of average value per match by subtracting equation (14) from equation (15):

$$\log V_{ij} - \log n_{ij} \approx \log\left(z^C/2\right) - \log 2 + \log\left[\left(\theta\pi(w_{ij}) + (2-\theta)\pi(W_{ij})\right)/\left(1-\theta\right)\right]$$
$$+\frac{(\lambda_j M_{ij} + \lambda_i M_{ji})\left(1+\pi(W_{ij})/\pi(w_{ij})\right)^{1/\theta}}{\left(\lambda_i N_i^B \lambda_j N_j^B / N\right) z_0^{1/\theta}} \left\{ \frac{z^m \pi(w_{ij})\left(z^C d_{ij}^{\beta}/2\right)^{1/\theta - 1}(1-\theta)}{\theta\pi(w_{ij}) + (2-\theta)\pi(W_{ij})} - \left(z^C d_{ij}^{\beta}/2\right)^{1/\theta} \right\}, \; i \neq j. \quad \textbf{(16)}$$

Since we cannot sign the term in braces, we cannot sign the effect of migrants on average value per match. Intuitively, since migrants lose productivity but gain information in the bridge matching process relative to non-migrants, we should not expect them to have a clear impact on average value per match. Note that there is no impact of distance on value per match, except possibly through the migrant effect.

We can write the volume of trade as in the next equation, which is valid both for $i \neq j$ and $i = j$ (see equation 13 and the expression before equation 15):

$$V_{ij} = m_{ij}\gamma_i^{D_{ij}} z^m d_{ij}^{-\beta}\pi(w_{ij})/2 + \alpha D_{ij}\lambda_i N_i^B \pi(w_{ii}) +$$
$$\left(\lambda_i N_i^B \lambda_j N_j^B / N\right)\left(\underline{z}_{ij}/z_0\right)^{-1/\theta}\left(z^C/4\right)\left(\theta\pi(w_{ij}) + (2-\theta)\pi(W_{ij})\right)/(1-\theta),$$

where D_{ij} is an indicator variable such that $D_{ij} = 0$ if $i \neq j$, and $D_{ii} = 1$; the variable m_{ij} is defined as $m_{ij} \equiv (\lambda_j M_{ij} + \lambda_i M_{ji})$ for $i \neq j$, $m_{ii} \equiv \lambda_i N_i^B$ for $i = j$; γ_i is defined as $\gamma_i \equiv 2M_{ii}/N_i^B$; and we consider the following parameter to be approximately a constant:

$$\alpha \equiv (1/2)\left(z^C/2\right)\left(\sum_j \lambda_j \left(N_j^B \eta\left(\underline{z}_{ij}\right) + \sum_k M_{kj} \right)\right) \Big/ N \quad .$$

We again log-linearize, employing the same manipulations that we used for equations (14) and (15), to obtain:

$$
\begin{aligned}
\log V_{ij} \approx\; & \log z_0/\theta - (1/\theta - 1)\log\left(z^C/2\right) - \log(2N) - (1/\theta)\left[\beta \log d_{ij} + \log\left(1 + \pi(W_{ij})/\pi(w_{ij})\right)\right] \\
& + \log\left[\left(\theta\pi(w_{ij}) + (2-\theta)\pi(W_{ij})\right)/(1-\theta)\right] + \log\left(\lambda_i N_i^B\right) + \log\left(\lambda_j N_j^B\right) \\
& + \frac{\left(m_{ij}\gamma^{D_{ij}} z^m d_{ij}^{-\beta}\pi(w_{ij})/2 + \alpha D_{ij}\lambda_i N_i^B \pi(w_{ii})\right) N z_0^{-1/\theta}\left(z^C/2\right)^{1/\theta-1} d_{ij}^{\beta/\theta}\left(1+\pi(W_{ij})/\pi(w_{ij})\right)^{1/\theta}}{\lambda_i N_i^B \lambda_j N_j^B \left(\theta\pi(w_{ij}) + (2-\theta)\pi(W_{ij})\right)/(1-\theta)}.
\end{aligned}
\qquad \textbf{(17)}
$$

6. Data and preliminary results

For the purposes of this draft we employed data sets we already used in other work, plus the OECD migration data set (see below). This leads to two glaring deficiencies, to be remedied in future work. First, we have no wage data, so all equations are estimated as though wages were equal across all countries, causing all the wage terms in equations (14) – (17) to disappear or become constants. For now, we can only hope that this is not causing omitted variables bias of any importance. Second, we have no instruments for migration, so we simply report OLS results.

We have shown that bilateral trade can be decomposed into number of matches and average value of trade per match. Successful matches in our model imply formation of trading firms. Data on bilateral numbers of trading firms and value of trade per firm are currently available only for France in 1986 and 1992, for manufacturing exporters. There are two problems with using these data. First, we have no data on migrants resident in France for those years. Second, without a matrix of bilateral observations we cannot use country fixed effects in estimation. Nevertheless, in the next draft of this paper we will use these data in estimations that will supplement the regressions reported below. In this draft, we will only use these data to check on the quality of the proxies for number of matches and value of trade per match that we now describe.

We follow Besedes and Prusa (2006), who count "relationships" between the United States and its trading partners by the number of product varieties for which positive trade is observed. By employing the maximum level of disaggregation available (10-digit Harmonized System), they hope to approximate firm-level relationships. We recognize that there will typically be more than one partnership per variety for most pairs of countries, and there may also often be more than one variety per partnership. Varieties are simply the best proxy we have for trade-generating partnerships that is available across a large number of pairs of countries.

For our dependent variables we use the NBER-UN Trade Database (Feenstra et al 2005), which contains bilateral trade flows at the 4-digit SITC level, revision 2. Of course these data are insufficiently disaggregated, and we will use HS6 data in future work. The data cover the years 1962-2000 and more than 200 countries and territories. However, in order to match these data to our migration data, we could only use the year 2000 and needed to combine some small countries, reducing the number of countries to 181. We have generated two additional variables from these trade value data to serve as admittedly imperfect empirical correlates for matches and average value per match, respectively. We call the first variable "links." To construct it, we defined the number of unidirectional links from an exporting country to an importing country as the number of 4-digit SITC industries whose exports from the former to the latter are non-zero. For the bidirectional number of links we added the links in the reverse direction, but subtracted all those that were doubly counted. For example, if the United States exports goods in 635 industries to Canada, and Canada exports goods in 600 industries to the United States, 500 of which show exports in both directions, then the number of US to Canada links is 635; the number of Canada to US links is 600; and the number of bilateral links is 735. We also constructed a second variable called "value per link," which is simply calculated as total trade value for each country pair divided by the number of links.

How good are links and value per link as proxies for number of matches and value per match, as measured by the French export data? The correlation coefficients between the logarithm of number of export links and the logarithm of the number of French exporting firms are 0.94 in 1986 (for 113 French trading partners) and 0.95 in 1992 (for 111 French trading partners). The correlation coefficients between the logarithm of value per link and the logarithm of value of exports per French exporting firm are 0.80 in 1986 and 0.76 in 1992. Our results for number of links are therefore a more reliable test of our theory than our results for value per link.

Except for migration, all of our right-hand side variables come from the CEPII data (Mayer and Zignago 2006). "Distance" between countries measures the shortest arc distance between the most populated cities of each country. Internal distances are calculated as $d_{ii} = 0.67\sqrt{area_i/\pi}$, where $area_i$ is the surface area of country *i*. The dummy variable "Colonial tie" indicates a colonial relationship that existed after 1945, where one country governed the other. For example, both India and Kenya have a colonial tie with Great Britain, but not with each other. The variable "Contiguity" assumes a value of one if two countries share a common land border. The variable "Common language" takes the value one if two countries share a primary or official language.

Turning to migration, the OECD has compiled data on the number of foreign born by source country in each of its member countries (see Dumont and Lemaître 2005 for documentation). Unfortunately, the numbers of migrants hosted by non-OECD trading partners are unavailable. We decided to make a virtue of necessity and limit the data we used to estimate equations (14) , (16), and (17) to OECD country exports and migrants hosted by the exporting countries. Our idea is that our entrepreneurial networks explanation for the impact of migrants from trading partner countries on OECD exports is more compelling than the same explanation for the impact of these migrants on imports to the OECD, which may also be driven by the migrants' tastes for products from their countries of origin. Another advantage of working with OECD data is that an OECD country's exports to itself are readily estimated using the STAN database. As Wei (1996) points out, a country's internal trade can be calculated as the difference between its production and exports of goods to the rest of the world. The STAN database provides production and export data for all OECD countries except Turkey. In order to match the data to the NBER-UN trade data, the values for Belgium and Luxembourg have to be combined, leaving us with 28 observations for intranational trade.

Our final data set used for estimation covers 28 exporting OECD countries and 154 importing countries (including the OECD countries), each of which receives exports from at least one OECD country that hosts migrants from that importing country. This leads to a total of 28 x 154 = 4312 possible exporting relationships, out of which 3125 are captured by our data. 1187 trading relationships are either zero or have missing data. These are dropped in the estimates reported in Table 1. The zeros will be used in future work, as discussed in the next section.

Equations (14) for number of links, (16) for value per link, and (17) for volume of trade contain two variables for which data are unlikely to become available for a wide cross-section of countries any time soon: λ_i, the fraction of old workers who become entrepreneurs (in the traded goods sector), and M_{ii}, the number of old workers who migrate internally. The terms $\log\left(\lambda_i N_i^B\right)$ are absorbed into country fixed effects. However, λ_i and M_{ii} are also embedded in the coefficients on M_{ji}, the migrants hosted by OECD countries, and D_{ij}, the intranational dummy variable. In this draft we simply ignore this econometric problem and estimate fixed coefficients on D_{ij}, M_{ji} , and the interactions of M_{ji} with colonial ties and distance, deflating all these variables by the product of country populations as suggested (loosely) by equations (14), (16), and (17).

Table 1 shows the results of our estimates for the volume of trade, number of links, and value per link, respectively. We will discuss the results on intranational trade, distance, migrants, and colonial ties in light of the predictions made by our theory. We begin with the first three columns of the table.

The significance of our intranational trade variable indicates the presence of "excess" trade within a country's borders, which we attribute to cluster matching. We can see from equation (17) that this effect should increase with internal distance and general "remoteness" (if we allow α to vary), but with only 28 observations we did not explore interaction terms.[11]

The results for the distance coefficients are slightly disappointing. The impact of distance on the number of links is clearly smaller than its impact on the value of trade, with the difference plainly explained by an impact on value per link that is not predicted by our theory. However, it could be that the impact of distance on value per link simply indicates that 4-digit SITC is too highly aggregated, so that the negative effect of distance really reflects its impact on the number of matches rather than on average value per match.[12]

With this possibility in mind, we turned to U.S. Department of Commerce trade data disaggregated to the HS10 level, the same level of disaggregation used by Besedes and Prusa (2006). For consistency with Table 1, we use data for the year 2000. Because these data are for U.S. imports and exports only, we cannot include importer and exporter dummies in our

[11] As discussed in Rauch and Watson (2007), the real empirical interest in modeling cluster matching is predictions regarding the impact of different legal treatment of spinoff entrepreneurship across U.S. states on spatial patterns of entrepreneurship and, consequently, trade. The U.S. Census data needed to test these predictions are expected to become available in about one year.

[12] Qualitatively unchanged results are obtained for the distance coefficients if we drop migration and intranational trade and use the full matrix of bilateral exports (or total bilateral trade) for the 181 countries available in the NBER-UN Trade Database.

regressions, and instead simply include all the variables included in standard gravity equations (the results on distance are robust to various lists of these). The distance coefficients, estimated separately for U.S. total trade, exports, and imports, are reported in Table 2. Unlike in Table 1, these results confirm our theoretical predictions. This is remarkable, given that a standard CES derivation of the gravity equation yields the opposite predictions: all of the impact of distance on the value of trade should come from its impact on value per link and none through the number of links.

Based on equations (17) and (14), we expect positive coefficients on our (appropriately specified) migrant variable in the regressions for volume of trade and number of links, and these expectations are met in the first and second columns of Table 1.[13] Our model makes no prediction regarding the impact of migrants on value per link, and the third column of Table 1 shows that the coefficient on the migrant variable is statistically insignificant. As we mentioned in the previous section, we can also expect the coefficient on our migrant variable in the regressions for volume of trade and numbers of links to increase with distance (not the logarithm of distance). Including an interaction between the migrant variable and distance in these regressions leads to positive but insignificant coefficients on both the migrant variable and its interaction with distance, indicating a problem with multicollinearity, but F-tests show that the two variables are jointly significant in both equations.

[13] The numerically huge coefficients simply compensate for the deflation of the number of migrants by the product of populations in equations (16) and (14). However, we do not back out the quantitative estimate of the impact of migrants on exports for comparison to the previous literature, preferring to wait until we have instrumental variables estimates.

Concluding with colonial ties, we see that they have positive impacts on the volume of trade and number of links, as predicted by our theory, but not on value per link. [14] We then add an interaction term between migrants and colonial ties in the fourth through sixth columns of Table 1. Based on equations (17) and (14), we expect this interaction to be negative in the regressions for volume of trade and number of links. Intuitively, migrants and colonial ties both provide information about bridge match partners for the country-pairs to which they apply, so they substitute for each other. We see that the interaction is indeed negative and significant. [15]

7. Work in progress

Many improvements and extensions of our work to date are possible. There are two that are currently in progress.

Estimating the effect of past trade on current trade. All pairs of countries build up networks through past trade, not just country pairs with colonial ties. However, we cannot simply include past trade as an explanatory variable for current trade since it inevitably captures a host of country-pair specific omitted variables. It is necessary to include country-pair fixed effects, and therefore move to panel estimation. Our theoretical model suggests a generational time frame and therefore long differences. The NBER-UN data go back to 1962, allowing us to compute two differences of 15+ years or three differences of 10+ years. We can also use the

[14] We find a positive and significant effect of colonial ties on value per link if we drop migration and intranational trade and use the full matrix of bilateral exports (or total bilateral trade) for the 181 countries available in the NBER-UN Trade Database.

[15] Equations (17) and (14) indicate that there should be a positive interaction between colonial ties and the logarithm of distance in the volume of trade and number of links regressions. If we substitute this interaction for the standard colonial ties specification reported in Table 1, the coefficient on the interaction is strongly positive and significant and the R-squared is essentially unchanged. Simply adding the interaction to the specification in Table 1, however, created hopeless multicollinearity problems. Our colonial ties dummy takes on the value one for only 74 observations in Table 1. As described in the next section, we hope that in future work we will get better results for value per link and for interaction terms by using past trade directly instead of colonial ties.

much more disaggregated US data, which were useless for estimating the impact of colonial ties. These switch from HS10 to TS7 (at least for imports) when we go back before 1989, but this is not a problem since we can use the long difference from the latter data to explain the long difference for the former data.

Since past trade, unlike colonial ties, is available for all pairs of countries, we should obtain much better estimates of impact on value per link and also of interactions with distance than those reported in this draft. Focusing on number of past links for concreteness, we can specify the exponent of the Pareto distribution $-1/\theta$ as $\log(n_{ijt-1}/n_0)/\theta_0$, where $n_0 > n_{ijt-1}$ and θ_0 sufficiently small ensure $0 < \theta < 1$. This functional form makes θ increasing in the number of past links at a decreasing rate, and also yields a linear dependence of $\Delta \log n_{ijt}$ on $\Delta \log n_{ijt-1}$ and a concave dependence of $\Delta \log(V_{ijt}/n_{ijt})$ on $\Delta \log n_{ijt-1}$.

Including the zeros. In our empirical work so far we have dropped all country pairs for which observed trade is zero, thereby losing valuable information. Dealing with zeros is a potential strength of our model, but to realize that potential we must move away from the assumption of a continuum of agents, which yields a deterministic value for the number of matches. Working with the simplest version of our model, with a finite number of agents we can view the number of matches n_{ij} given by equation (5) as an expected value. It is easily shown that n_{ij} follows a binomial distribution with number of trials $\lambda_i N_i \lambda_j N_j / N$ and probability of success $1 - \eta(\underline{z}_{ij})$. Using the method of maximum likelihood, we can then estimate the same parameters that we estimated previously, this time including the observations for which the number of links is zero.

With this approach we can compute the probability that all matches between countries *i* and *j* fail to be $\left(\eta(\underline{z}_{ij})\right)^{\lambda_i N_i \lambda_j N_j / N}$. This yields the sensible result that the probability of zero matches (links) between two countries decreases with each of their sizes and increases with the distance between them (through $\underline{z}_{ij}$). In actual estimation we can replace $\lambda_i N_i$ with a country fixed effect, as we did in the previous section.

References

Besedes, Tibor and Prusa, Thomas. 2006. "Product Differentiation and Duration of U.S. Import Trade." *Journal of International Economics* 70(2): 339-358.

Bhide, Amar. 2000. *The Origin and Evolution of New Businesses.* New York: Oxford University Press.

Braun, Ernest and Macdonald, Stuart. 1982. *Revolution in Miniature: The History and Impact of Semiconductor Electronics Re-Explored in an Updated and Revised Second Edition* (New York: Cambridge University Press).

Burt, Ronald S. 2000. "The Network Structure of Social Capital." In Robert I. Sutton and Barry M. Staw, eds., *Research in Organizational Behavior* 22 (Greenwich, CT: JAI Press).

Dumont, Jean-Christophe and Georges Lemaître. 2005. "Counting Immigrants and Expatriates in OECD Countries: a New Perspective." OECD Social Employment and Migration Papers, n. 25, OECD Publishing.

Fafchamps, Marcel; El Hamine, Said; and Zeufack, Albert. Forthcoming. "Learning to Export: Evidence from Moroccan Manufacturing." *Journal of African Economies*.

Feenstra, Robert C., Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, and Hengyong Mo. 2005. "World Trade Flows: 1962-2000." National Bureau of Economic Research Working Paper no. 11040 (January).

Gompers, Paul; Lerner, Josh; and Scharfstein, David. 2003. "Entrepreneurial Spawning: Public Corporations and the Genesis of New Ventures, 1986-1999." National Bureau of Economic Research Working Paper No. 9816 (July).

Head, Keith; Mayer, Thierry; and Ries, John. 2007. "The Erosion of Colonial Trade Linkages after Independence." University of British Columbia Working Paper.

Jackson, Matthew O. and Rogers, Brian W. 2005. "The Economics of Small Worlds." *Journal of the European Economic Association* 3 (April-May): 617-627.

Javorcik, Beata et al. 2006. "Migrant Networks and Foreign Direct Investment." World Bank Policy Research Working Paper 4046 (November).

Konecny, Tomas. 2007. "Can Immigrants Hurt Trade?" CERGE-EI Working Paper, May.

Mayer, Thierry and Soledad Zignago. 2006. "Notes on CEPII's Distances Measures." CEPII Working Paper, available at www.cepii.fr/distance/noticedist_en.pdf (May).

McCallum, John. 1995. "National Borders Matter: Canada-U.S. Regional Trade Patterns." *American Economic Review* 85 (June), pp. 615-623.

Michelacci, Claudio and Sivla, Olmo. "Why So Many Local Entrepreneurs?" *Review of Economics and Statistics* 89(4): 615-633.

Rauch, James E. 2001. "Business and Social Networks in International Trade." *Journal of Economic Literature* 39 (December): 1177-1203.

Rauch, James E. and Alessandra Casella. 2003. "Overcoming Informational Barriers to International Resource Allocation: Prices and Ties." *Economic Journal*, 113 (January): 21-42.

Rauch, James E. and Trindade, Vitor. 2002. "Ethnic Chinese Networks in International Trade." *Review of Economics and Statistics* 84 (February): 116-130.

Rauch, James E. and Trindade, Vitor. 2003. "Information, International Substitutability, and Globalization." *American Economic Review* 93 (June): 775-791.

Rauch, James E. and Watson, Joel. 2007. "Clusters and Bridges in Networks of Entrepreneurs." In James E. Rauch, ed., *The Missing Links: Formation and Decay of Economic Networks* (New York: Russell Sage Foundation).

U. S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census. 1997. 1992 Economic Census. Characteristics of Business Owners. Washington, D. C.: Government Printing Office.

Wei, Shang-Jin. 1996. "Intra-National versus Inter-National Trade: How Stubborn Are Nations in Global Integration?" NBER Working Paper 5531 (April).

Wolf, Holger C. 2000. "Intranational Home Bias in Trade." *Review of Economics and Statistics* 82(4), pp. 555-563.

Table 1: Regressions for Trade Volume, Links, and Value per Link

Dependent variable / Independent variables	Trade Volume	Links	Value per Link	Trade Volume	Links	Value per Link
Log of distance	-1.32 (-17.53)	-0.85 (-13.91)	-0.42 (-16.56)	-1.30 (-16.56)	-0.84 (-12.82)	-0.42 (-16.65)
Contiguity	0.19 (0.94)	-0.26 (-1.40)	0.49 (6.34)	0.16 (0.81)	-0.31 (-1.74)	0.49 (6.19)
Common language	0.54 (3.41)	0.45 (3.45)	0.02 (0.34)	0.51 (3.26)	0.42 (3.30)	0.02 (0.30)
Colonial tie	1.34 (3.71)	1.11 (4.92)	0.16 (1.28)	1.39 (3.79)	1.16 (5.02)	0.16 (1.31)
Migrants	24.66 E+06 (2.28)	21.46 E+06 (2.87)	-1.87 E+06 (-0.64)	174.3 E+06 (5.49)	149.0 E+06 (5.62)	8.30 E+06 (0.64)
Migrants x Colonial tie				-157.0 E+06 (-5.04)	-133.7 E+06 (-5.08)	-10.7 E+06 (-0.88)
Intranational	6.25 E+10 (4.13)			6.84 E+10 (4.19)		
Importer and exporter dummies	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.89	0.89	0.76	0.89	0.89	0.76
Observations	3125	3097	3097	3125	3097	3097

Estimates for 2000, sample restricted to 28 OECD exporters (Belgium and Luxembourg combined, no data on Iceland for immigrants; no data for Turkey on internal trade), all countries in the Feenstra NBER-UN database that could be matched to our gravity data (154 countries) remain as importers. Robust t-statistics in parentheses, clustered around each importing country.

Table 2: Distance Coefficients for US Trade at the 10-digit HS Level

	Imports + Exports	Exports	Imports
Value	-0.84 (-3.94)	-1.27 (-6.31)	-0.61 (-2.14)
Links	-0.98 (-8.08)	-1.12 (-8.36)	-0.68 (-4.70)
Value per Link	0.14 (1.04)	-0.15 (-1.35)	0.07 (0.33)
Controls used in all regressions	Log population Log per capita GDP Dummy for preferential trading agreement Dummy for any U.S. sanctions Dummy for use of the dollar as currency Dummy for English language Dummy for shared land border		
Observations	176	176	175

The table reports the coefficients on distance for nine different regressions, where the dependent variables are either imports and exports, only imports, or only exports, measured either in volume, number of links, or total value per link. Data are for the year 2000. Robust t-statistics in parentheses.

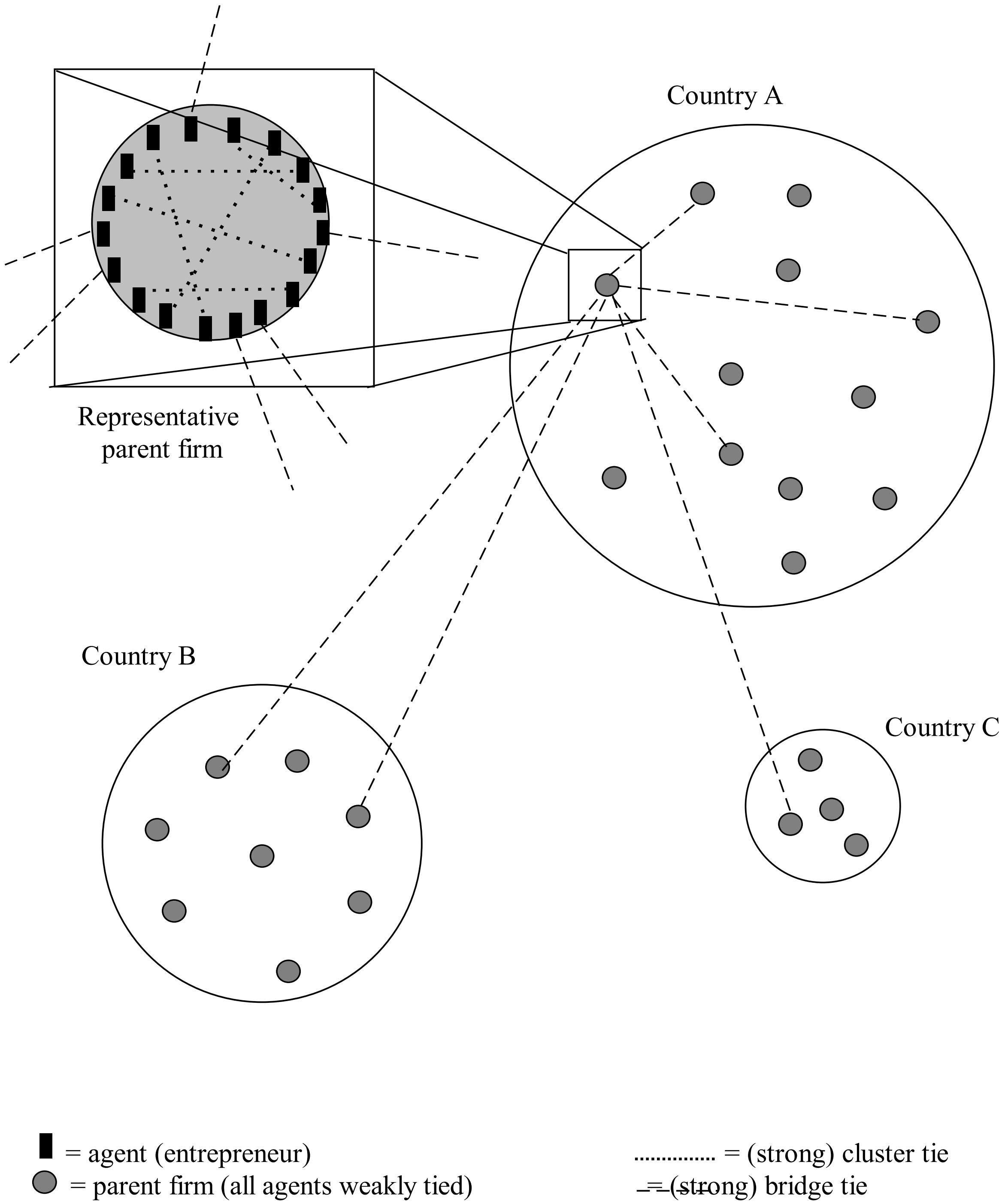

Figure 1: Underlying Network Structure