Monograph:

de Lucio, J., Mínguez, R., Minondo, A. et al. (1 more author) (2014) Networks and the dynamics of firms' export portfolio: evidence for Mexico. Working Paper. SERPS (201400). Department of Economics, University of Sheffield ISSN 1749-8368

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Networks and the Dynamics of Firms' Export Portfolio: Evidence for Mexico

Juan de Lucio
Raúl Mínguez
Asier Minondo
Francisco Requena

Networks and the dynamics of firms' export portfolio: Evidence for Mexico

Juan de Lucio (High Council of Spanish Chambers of Commerce, Spain)

Raúl Mínguez (High Council of Spanish Chambers of Commerce, Spain)

Asier Minondo (Deusto Business School, Spain)

Francisco Requena (University of Sheffield, United Kingdom) *

Abstract

In this paper we use network-analysis tools to identify communities in the web of exporters' destinations. Next we use our network-based community measure as predictor of additional countries chosen by firms expanding their export destination portfolio. We defend that our network-based community measure is superior to extended gravity measures. This superiority stems from the fact that community is a revealed measure, is country-specific and can be calculated at the industry level. Using data on Mexican new exporters over the period 2003-2009, we show that the probability of choosing a new export destination multiplies almost by three if it belongs to the same community of any of the firm's previous destinations. The introduction of the network-based community variable improves the accuracy of the model up to 20% relative to a model that only includes gravity and extended gravity variables. We also show that industry-specific communities and general communities play similar roles in determining the dynamics of Mexican exporters' country portfolio.

*Corresponding author: Francisco Requena. Department of Economics, University of Sheffield, Edgar Allen House, 241 Glossop Road, Sheffield S10 2GW UK. Tel: + 44 114 222 3398, Fax: + 44 114 222 3458. Email: f.requena@sheffield.ac.uk

Introduction

Product and country export portfolio is much more concentrated in developing countries than it is in developed countries (Cadot et al, 2013). In addition, survival rather than entry into exports markets is the key to understand the growth of exports in developing countries (Besedes and Prusa, 2011). These two facts motivate this paper: there is a need to examine the dynamics of firms' export destination portfolio in developing countries. If destination path-dependence in exports exists, it is important that firms choose adequately export destinations because the chances of survival are higher and, by extension, export growth.

Until recently firm-level research in this area has mostly treated export status as a binary variable: firms are either exporting or they are not. Hence, empirical studies of entry into exporting have focused on the initial entry decision, particularly on identifying the firm-specific characteristics which set exporting firms apart from non-exporters. In this paper we focus our enquiry on a subsequent question: given that firms have the ability to export, what determine their choices about where to export?

Exporters do not add new destinations to their portfolio at random. In fact, firms have a higher probability of adding destinations that are similar to their domestic market. This similarity is governed by the so-called gravity factors, such as income, geography or culture. Firms also have a higher probability to add new destinations that are similar to their previous destinations. Previous literature contends that this similarity can be approached by the so-called extended-gravity variables that capture the geographical and cultural closeness between previous and new destinations (Morales et al., 2011).

This paper is divided into two parts. In the first part we develop a new similarity measure using the tools of complex network analysis. In particular, we identify communities of countries sharing similar characteristics within the web of export markets of Mexico. Our measure has several advantages relative to extended gravity indicators. First, it is a revealed measure, and hence it captures not only all extended-gravity proxies, but also any non-measurable or non-observable characteristic that might also affect how similar export destinations are. Second, the community measure is a country-specific measure. In the case of extended gravity variables, the similarity between two export destinations is always the same, irrespective of the location of the exporter. For example, the (extended gravity) distance between Country A and Country B is the same for exporters from Mexico and another Country (say C). However, it might be the case that Country A and Country B belong to the

same community of Mexico but not of Country C. For example, Country A and Country B might have the same preferences for Mexican products but not for products from Country C. The third advantage of our network-based measure is that we can identify industry-specific communities. It might be the case that countries belong to the same community in one industry but to a different community in another industry. For example, in the case of Mexico and regarding tequila, Country A and Country B might form a community because both countries have the same regulations on the maximum alcohol content. In this case, the tequila that has been modified to meet the requirements in Country A will also be suitable for Country B, leading these two countries to form a community. In contrast, for book sales, Country A and Country B might belong to different communities because they speak different languages. Extended-gravity measures cannot control for these differences because geographical and cultural variables do not usually vary across industries.

In the second part of the paper, we use our network-based similarity measure as determinant of the destinations-portfolio of new regular Mexican exporters over the period 2003-2009. We show that an exporter will have almost three times higher probability of choosing a new destination if it belongs to the same community of any of its previous destinations. The network variable keeps its strong predictive capacity even when we control for gravity and extended-gravity variables, and improves the accuracy of the model up to 20%. We also identify industry-specific destination-communities and general destination-communities and find that both have a similar influence on the evolution of Mexican firms' export-portfolio.

Our analysis is related to different strands of literature that analyze firms' exports dynamics. Chaney (2011) proposes a model of international network formation where firms obtain information about new potential partners from their current trading partners. The network formation game yields equilibria where firms' export destinations are path-dependent. Albornoz et al. (2012) and Nguyen (2012) develop alternative multi-market export models based on the idea that a firm's foreign demands are uncertain and correlated across markets. When faced with multiple destinations to which they can export, many firms will choose to sequentially export in order to slowly learn more about its chances for success in untested markets. Experimentation becomes an optimal strategy leading to path-dependence in firms' export destinations.

Our paper is also related to the concept of "the geographic spread of trade", a term originally proposed by Evenett and Venables (2002). They showed that geographic and linguistic proximity to an existing export-market was a consistently significant factor in

determining expansion into new markets for sector-level exports from developing countries, implying a role for learning from existing export experiences. Using firm-level export data, Morales et al. (2011) for Chile, Lawless (2011) for Ireland, and Defever et al. (2011) for China explore the role of "extended gravity" forces and show that firms tend to choose new export destinations that are similar (geographically, culturally or economically) to destinations that firms are already exporting to.

Our paper is also related with the novel literature that applies network methods to analyze international trade. Kali and Reyes (2007) map the topology of international trade and develop new measures of economic integration based on network analysis. De Benedictis and Tajoli (2011) apply network analysis to describe the evolution of international trade and to study other trade-related topics. Hidalgo et al. (2007) use the probability of exporting products in tandem to develop a measure of proximity between products, which is displayed into a map using network analysis. These authors show that the product-map determines the evolution of countries' productive specialization. Based on the product-map, Kali et al. (2013) show that countries' growth prospects are enhanced if their export basket is closer, measured using network analysis tools, to more complex products that the country does not export yet.

The remainder of the paper is organized as follows. Section 2 presents the data and applies network analysis tools to identify communities within the web of Mexican exporters' destinations. Section 3 introduces the empirical model to examine how connectedness between countries affects the choice of new destinations by Mexican new exporters that expand their export portfolio. Section 4 presents the results of the empirical analyses. Section 5 concludes.

2. Communities in the network of Mexican export markets.

In this section, we start presenting the database used to perform our empirical analysis and then explore the web of export destinations in Mexico using tools from network analysis. Next, we explain how to identify communities in a network and apply the identification algorithm of communities to the entire network of Mexican exporters. Finally, we construct industry-specific networks and implement the algorithm of identification of communities separately to each of them and show that the number of communities and its members can vary from industry to industry.

2.1 An exploration of the web of export destinations.

We use the transaction level customs data on the universe of Mexican exporters over the period 2000-2009. The database was facilitated by the World Bank's Trade and Integration Team (Cebeci et al, 2012).[1] The database provides the annual value of exports per firm, destination and Harmonized System 6-digit product code.[2] We use data over the period 2003-2009 to analyze the dynamics of exporters' destination portfolio in the next section (our dependent variable in the econometric exercise) and data over the period 2000-2002 to examine the network of export markets and construct the community measure in this section (our main explanatory variable). A detailed description of the Mexican firm-level data as well as other data sources used in the paper can be found in Appendix 1.

We begin our empirical analysis by examining the network of Mexican firms' export destinations using tools from network analysis. Figure 1 presents the network of Mexican exporters' destinations in year 2002.[3] Export destinations are nodes in the web and two nodes are connected by an edge if there is at least a firm that exports to both nodes. The size of the node is correlated with the number of firms that export to that destination and the size of the edge (weight) is correlated with the number of firms that export to both destinations linked by the edge. The network has 175 nodes and 9,576 edges. The density of the network is 0.63.[4] The most important destinations for Mexican exporters were the US (25,730 firms), Guatemala (2,534 firms), Canada (1,931 firms), Costa Rica (1,855 firms) and El Salvador (1,394 firms). The edges with the highest weights were Canada-US with 1,534 firms exporting to both destinations, Guatemala-US with 1,368 firms, Costa Rica-US with 1,084 firms, Costa Rica-Guatemala with 923 firms and Colombia-US with 913 firms. All the nodes are connected in the network; this means that there is no destination where all exporters to that destination only exported to that destination. Each node has an average degree of 55; that is, the total number of additional destinations served by firms that export to a destination is, on average, 55. As expected, the destination with the highest degree is the US: 170 edges. It is followed by Chile (164), Canada (160), Guatemala (159) and Colombia (158).

2.2 Identification of communities.

In the destination network, we want to identify communities of countries that have stronger relationships among them than with the rest of destinations in the network. These

---

[1] Data were collected by the Trade and Integration Unit of the World Bank Research Department as part of their efforts to build the Exporter Dynamics Database (http://econ.worldbank.org/exporter-dynamics-database).
[2] For example, one record of our database is the annual value of exports of "Pullovers, cardigans etc. of wool or hair, knit" (HS-6 code 6110101) by a Mexican exporting firm (identifier 15) to Italy in 2002.
[3] In the robustness section we use 2000 and 2001 data.
[4] If all nodes were connected with each other density would be 1.

tighter relationships reveal that if a firm exports to one country in the community it will also tend to export to other countries within the community. One widely used procedure to identify communities within a network is the maximization of a modularity function (Newman, 2006), which is expressed as follows:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{1}$$

where Q is the modularity index, m is the number of edges in the network, A_{ij} is the number of edges between node i and node j, k_i and k_j are the degree of nodes i and j respectively, and $\delta(c_i,c_j)$ is a delta function that takes the value of 1 if i and j belong to the same community, and zero otherwise. The term in brackets compares the number of edges between two destinations with the number of edges we would expect if edges were distributed randomly in the network, providing that the degree of each destination is not altered. Hence, the term in brackets compares the actual number of relationships with a benchmark number of relationship. If the number of edges between i and j is higher than the benchmark, these destinations will form a community. The network will be partitioned in a number of communities that maximizes the value of Q.

The procedure to determine the optimum number of communities is not trivial, as the number of possible combinations of destinations rises exponentially with the number of destinations, making the exhaustive comparison of all possibilities unfeasible. To overcome this problem, different algorithms have been proposed to maximize modularity and identify communities within a network. In this paper, we use the algorithm proposed by Blondel et al. (2008). However, as pointed out by Fortunato and Berthelemy (2007), the modularity maximization algorithm has a resolution-limit limitation, as it might aggregate small communities within a broader community. In order to resolve this limitation they suggest applying the maximization algorithm iteratively. First, the algorithm is applied on the whole network. Second, the algorithm is applied only on each community identified in the first step. The process stops once the algorithm does not find any further partition. In each step, it should be checked that the number of edges within the communities identified by the algorithm is larger than the expected number of edges. This iterative process has the advantage of identifying hierarchies of communities. At the beginning, the algorithm identifies few communities, characterized by a large number of members with only just above the average connections between them. However, with each iteration, large communities are fragmented into smaller communities characterized by stronger ties among members.

To avoid a too long iterative process and awkward relationships, we run the community detection algorithm on a sample of destinations which are served, at least, by 50 exporters in 2002. In addition to that, exporters have to export, at least, to two different destinations. After applying this filter, the number of nodes declines from 175 to 65 (representing 79% of total Mexican exports) in 2002.

Figure 2 displays the process of community-identification in Mexico. In the first iteration, the algorithm identifies two big communities. The first community is composed by countries located in the American continent, except USA and Canada; and the second community by the rest of countries. When we apply the algorithm on the community of countries located in the American continent, three final communities (shaded in yellow) emerge. The first community is formed by South American countries: Argentina, Bolivia, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay and Venezuela. The second community is formed by Central American countries and three Caribbean countries: Belize, Costa Rica, Cuba, Dominican Republic, El Salvador, Guatemala, Honduras, Nicaragua, Panama and Puerto Rico. The third community is formed by Caribbean countries: Bahamas, Barbados, Haiti, Jamaica and Trinidad and Tobago. When we apply again the modularity maximization algorithm on the rest of countries community we get a fourth final community composed by large developed countries: Canada, France, Germany, Great Britain, Italy, Spain and USA. If we further iterate the group of remaining countries we end up with six additional final communities. The fifth group is composed by small European countries: Austria, Belgium, Denmark, Finland, Netherlands, Norway, Sweden and Switzerland. The sixth community only includes three large Asian countries: India, Indonesia and Turkey. The seventh community encompasses countries located in or near the Middle-East area: Egypt, Israel, Saudi Arabia and United Arab Emirates, and two peripheral European Union countries: Portugal and Greece. The eighth community is composed by two Eastern European countries: Poland and Russia. The ninth community is formed mostly by Asian countries: China, Hong-Kong, Japan, Korea, Malaysia, Philippines, Singapore, Taiwan and Thailand; countries located in Oceania: Australia and New Zealand; and two emerging countries: Brazil and South Africa. The final community is formed by two small European Union countries: Hungary and Ireland.

We can observe that gravity variables, such as geographical location, play a role in the formation of communities. For example, most of South American countries are located in the same community and most of Central American countries are located in the same community. However, we also observe that there are some countries that do not obey the geographical

rule. For example, Brazil does not belong to the community of South American countries, but rather to a heterogeneous group that encompasses distant and emerging countries. We also observe that Canada and the US do not form a North American community, but are integrated in a broader large high-income countries' community.

These cases point out that there might be other reasons besides those captured by gravity variables that might explain similarity between countries. As the network of Mexican export-destinations is built upon the choices taken by Mexican exporters, each community identified in the network is a revealed synthetic indicator of all the variables that might determine the degree of similarity among export destinations. Hence, belonging to a community might be a superior criterion to identify similarities between countries than (extended) gravity measures.

What does a community capture besides the forces controlled by gravity variables? As mentioned before, a community captures any variable that influences the degree of similarity among countries that is difficult to observe or measure. For example, at the aggregate level, the existence of large distribution chains that happen to be present in some countries, but not in others, might explain why some exporters have a higher presence in a community of countries. It might also be the case, as suggested before, that countries share similar preferences for the exporter country products, and these similar preferences are not well captured by extended gravity measures. In addition to that, following the trade models based on firm heterogeneity (Melitz, 2003), a community can be considered as a set of countries that require a similar threshold productivity for a foreign firm to obtain profits. The threshold is high if fixed and variable costs of exporting are large or the size of the destination market is small. For example, for Mexican exporters, large European Union countries belong to a different community than small European Union countries because, due to differences in size, the former demands a lower threshold productivity to obtain profits than the latter. As variable costs of exporting are one of the determinants of the threshold productivity to obtain profits, a destination might belong to one community for a country and to a different community to another country. As long as gravity variables do not capture completely how the combination of fixed costs, variable costs and demand factors determine threshold productivities, the community variable might render a superior indicator, such that it still has explanatory power besides these variables.

To test that community captures other variables besides those controlling for gravity measures, we run a regression to determine to what extent gravitational variables explain the variation in the probability of belonging to the same community. The gravitational variables

we use to explain the similarity between two destinations are distance, sharing a border, speaking the same language, being located in the same region, belonging to the same income per capita quintile, having a common colonizer, belonging to the same regional trade agreement and the combined number of total migrants (number of citizens born in destination i that live in destination j plus the number of citizens born in destination j that live in destination i).[5] As shown in Table 1, a larger distance reduces the probability of belonging to the same community; while sharing a border, speaking the same language, being located in the same region and having a similar income per capita level increase the probability of belonging to the same community. The number of migrants born in one destination and living in other destination also raises the probability of belonging to the same community. Finally, sharing the same colonizer and belonging to the same regional trade agreement do not have a significant impact on the probability of belonging to the same community. We can observe that gravitational variables only explain 26% of the differences in the probability of belonging to the same community in Mexico. These results confirm that, besides gravity measures, community captures other variables that enhance the degree of similarity among countries.

2.3. Industry-specific communities.

As explained in the introductory section of the paper, one of the key advantages of the network-based similarity measure versus extended-gravity measures is that the former can be calculated at the industry level. In the extended gravity framework, for example, the marginal distance from an incumbent destination to a new destination is the same irrespective of the industry; however, in the network framework, the incumbent and the new destinations might belong to the same community in some industries but not in others. To capture this possibility, we classify exporters in one of seven industries: agriculture, chemicals, machinery & transport equipment, metals, non-metallic minerals, paper and textiles. For each industry, we identify the industry-specific destination-communities and the rest of industries destination-communities. The limitation of this analysis is that the number of destinations at the industry level is smaller than at the aggregate level. To keep a sufficient amount of destinations we set a less stringent criterion to determine the sample used to identify communities. In particular, we only exclude from the sample those destinations that are served only by one exporter. The sample that meets this requirement and is common for the seven industries is formed by 55 destinations. They represent 69% of Mexican exports in

---

[5] See Appendix 1 for a description of the data sources.

2002. To identify the communities we follow the same procedure as the one used to identify communities in the whole network.

Tables A2 in the Appendix display the identified communities in each of the seven industries in Mexico. We should stress that results should be taken with care as we use highly aggregated industries and few firms might be linking some marginal destinations, leading to firm-specific communities. As shown in the figures, there are differences in the number of communities among industries and countries. The highest number of communities is found in chemicals, 10, and the lowest in paper, 8. We observe that there are some broad communities present in most industries, such as a cluster of South American countries, a cluster for Central American countries and a cluster for Caribbean countries. However, the size of these clusters and its members vary from industry to industry. Moreover, in some cases, the communities of South American and Central American countries are merged. Asian countries tend to stay in the same community as other Asian countries; however, the members vary from industry to industry, and sometimes the Asian countries form different communities. This is also the case for large and high-income developed countries, and smaller and high-income European countries.

To assess the similarity among communities across industries we calculate an adjusted Rand index. This index, ranging between -1 and +1, calculates the fraction of destination pairs that belong to the same community in two different industries.[6] As shown in Table A3 of the appendix, the adjusted rand indexes lie between 0.17 and 0.41. In most cases the correlation across partitions is weak, confirming that destinations tend to belong to different communities when examining different industries. In the last column of Table A3 in the Appendix, we calculate the adjusted Rand index for the industry partitions and the rest of industries partitions for each industry. We observe a weak correlation between partitions. These results point out that the communities identified at the industry-level are different to those identified for the whole set of exporters.

3. Can communities predict the expansion path of firms' export destination portfolio?

So far we have calculated network-based communities from the entire web and industry-specific webs of Mexican firms' export destination portfolio. The rest of the paper investigates whether our network-based communities helps to predict which countries will be

---

[6] Specifically, the index calculates the fraction of correctly classified (respectively misclassified) elements to all elements. The index is adjusted to ensure that the expected value of the index for two random partitions is zero (Hubert and Arabie, 1985).

chosen by exporting firms that expand their destination portfolio. We begin by explaining the empirical model used to study the determinants of the expansion of destination portfolio by exporting firms. Next we describe sample of firms used to perform the empirical analysis.

3.1. Empirical model and econometric specification

Let $\mathbf{y_{iGt}} = (y_{i1t}, y_{i2t}, \dots, y_{iKt})$ denote the vector of a exporting firm *i's* current destination portfolio G made of K countries. We want to examine the decision about where to export when a firm expands the portfolio of destinations G. If destinations share common characteristics, having served one destination might reduce the sunk cost of entering similar destinations. Hence, previous export destinations might determine the export-path. Our interest lies in the quantification of the effect of "similarity" between countries on the probability of entering a new export destination. In particular we want to examine two types of measures: those based on gravity-type indicators and those based on our network analysis.

We derive our econometric equation from a simple model of export participation into specific foreign markets by profit-maximizing firms that produce one good in the local market and sell part of the production abroad. There are several alternative markets and firms have to decide which markets to export to. At any period, exporting firms have the choice of entering into a number of markets if it did not export to those markets in the previous period. Let π_{igt} be firm *i's* profits from exporting to market g in year t. We assume that the expected profits of exporting to country g by firm i is a linear function of factors affecting the destination choice,

$$\pi_{igt} = \mathbf{I'_{-g,t-1}}\alpha + \boldsymbol{\theta_g} + \varepsilon_{igt} \qquad (2)$$

where the vector $I_{-g,t-1}$ includes variables that measure the "mass" of information about destination g that firm i might obtain from previous exporting experience in other destinations, θ_g is a vector of destination-specific constant terms and ε_{ig} is a random term denoting the unobservable (by the researcher) unique profit advantage to the firm i from selling in country g.

An exporting firm will choose to export to a particular country if she earns the highest possible profit. Formally, the gth country is chosen by firm i as a new export destination (omitting the subscript t) if $\pi_{ig} = \max(\pi_{ik}, k = 1 \dots K)$. If the firm-specific random terms are independently distributed, each with a Type I extreme value distribution, McFadden (1974) showed that the probability of a firm i to choose a destination g is

$$P_{ig} = Pr\left(\pi_{ig} > \pi_{ik}, \ g \neq k\right) = \frac{\exp(I_{-g,t-1}'\alpha + \theta_g)}{\sum_k \exp(I_{-g,t-1}'\alpha + \theta_k)} \qquad (3)$$

where P_{ig} is the population relative frequency of exporting to destination g. The estimates are obtained by maximizing the likelihood function, $L = \prod_i \prod_g P_{ig}$. The model described above is known as conditional logit model (CLM). It is easy empirically to generalize the CLM to the case in which a firm can choose more than one destination every year.[7]

We create three sets of variables that capture the "mass" of information about destination g that firms obtained from previous exporting experience. First, we use our network-based community variable, $I_{-g,t-1}^{community}$. This variable takes the value of one if the new export market belongs to the same community of any of the firm's previous destinations and zero otherwise.[8]

As additional controls, we include gravity and extended gravity variables in order to control for "observable" similarity features between destination markets. Similarity between the domestic market (Mexico) and each new export markets is determined by standard gravity measures, such as geodesic distance between the domestic and the new export market, sharing a land border, common language, being members of the same regional trade agreement and having a large number of migrants from each country living in the other country. We also use GDP and GDP per capita as a proxy for the attractiveness of the new export market.

For extended gravity variables, we control for distance, border, language, regional trade agreement, migration, income level and geographical region. Then the variable $I_{-g,t-1}^{distance}$ characterizes the countries' geographical relationship to prior export destinations of the firm. It takes the value of one if the new export destination capital city is less than a certain number of kilometers away from the capital city of any of the previous destinations and zero otherwise. In the benchmark analysis, we use a 1,500 km. radius. We also proxy for the geographical links between countries using a common border dummy variable, $I_{-g,t-1}^{border}$, which takes the value of one if the new export destination shares a land-border with any previous destination and zero otherwise; and $I_{-g,t-1}^{region}$, which takes the value of one if the new export destination is in the same region of any of the previous destinations and zero

---

[7] Notice that CLM does not allow the inclusion of explanatory variables that are not directly related to the choices. In our case, it means that we cannot estimate a single parameter to capture the impact of firm-specific characteristics on the firm's probability of exporting to a particular destination. Another potential limitation of CLM is the risk of violation of the Independence Irrelevant Alternatives (IIA) assumption. In the sensitivity analyses section we estimate equation (3) with alternative models such as nested logit and mixed logit, that relaxes the IIA assumption and allow incorporating firm-level characteristics in the model.

[8] Notice that we constructed our "stock" community variable with 2002 data, before new exporters start expanding their "flow" of destinations since 2004 onwards.

otherwise.[9] The regional dummy variable also controls for the existence of natural trade blocs in international trade (Frankel et al., 1995). We also consider cultural closeness measures such as common language between export destinations. Specifically, the variable $I_{-g,t-1}^{language}$ takes the value of one if the new export destination speaks the same language of any of the previous destinations and zero otherwise. We also proxy for economic proximity, controlling the presence of any previous export destination located in the same income quintile of the new export market. We also consider other variables that might enhance the proximity between destinations, such as migration flows and belonging to the same regional trade agreement. Specifically, the variable $I_{-g,t-1}^{migr}$ takes the value of one if any previous destination has at least 100 immigrants and 100 emigrants in the new export destination and zero otherwise. The variable $I_{-g,t-1}^{rta}$ takes the value of one if the new export destination belongs to the same regional trade agreement of any of the previous destinations and zero otherwise.

3.2. Data

Our sample consists of all (6026) Mexican exporting firms that internationalized between 2003 and 2007 and carried on exporting until 2009. We called these firms "new regular exporters".[10] There are some interesting features that explain why we have chosen them. First, we know their entire export portfolio since we know they started exporting. Second, they account for a significant share of Mexican exports (21% of total firms, 23% of all transactions and 21% of all value of exports in 2009). Third, as Table 2 shows, they exhibit dynamics of export destination portfolio similar to the old regular exporters. Considering year-to-year changes, the largest percentage of firms (about 60 percent) does not modify the export destination portfolio. Every year some firms only enter into new destinations (15%), others opt for only exiting (9%) and, finally, others decide to enter and exit simultaneously (17%). The last column in Table 2 shows the transitions in the export destination portfolio of the old regular exporters (firms exporting before 2003) in the period 2007-2008. They are almost the same as those of the new regular exporters.

Table 3 presents a summary of our dependent variable: the number of new destinations per year served by a typical new regular exporter. The percentage of firm-year pairs that take a value of one is 56%, that is, the majority of firms in our sample that decide to expand their

---

[9] We use the seven major regions identified by the World Bank: East Asia & Polynesia, Europe & Central Asia, Latin-American & Caribbean, Middle East & North Africa, South Asia, and Sub-Saharan Africa.

[10] Strictly speaking, we know that a new regular exporter did not export in 2000, 2001 and 2002. In our set of new regular exporters, 757 started to export in 2003, 948 in 2004, 1110 in 2005, 1283 in 2006 and 1928 in 2007.

destination portfolio enter a single new destination. The number of firms that expand their destination portfolio in more than six destinations is very small (less than 3% of the firm-year).

Next we show the distribution of destinations across communities. The number of communities served by a typical new regular exporter each year is one: 75% of new regular exporters only serve one community, 14% of new regular exporters serve two communities, 6% of new regular exporters serve three communities and 5% of new regular exporters serve four or more communities. As the US is the most important destination for Mexican firms, a very large percentage of new regular exporters (84%) serve the community in which the US is integrated[11]; 22% serve the community of Central American countries, 13% serve the community of South American countries and another 13% the community of Asian countries. The community of small European countries is served by 5% of new regular exporters; the community of Middle East countries, the community formed by India, Indonesia and Turkey, and the community of Caribbean countries are only served by 2% of the new regular exporters. The rest of communities are served by very few new regular exporters.

4. Estimation results

In this section we present the results from the econometric analyses. First, we investigate whether communities identified in the whole network of Mexican exporters determine the export path of new exporters. We also analyze whether industry-specific destination-communities have a larger role in determining the export path than general destination-communities. Second, we perform a set of sensitivity analyses to test the robustness of our results.

4.1. Main results

Table 4 reports estimates of the conditional logit model allowing for simultaneous exports to multiple destinations. In specification (1) we estimate the model with community as the only independent variable. The communities used in the benchmark analysis are the final communities identified in Figure 2. The community coefficient has a large positive value and is strongly statistically significant. The transformation of the community coefficients into odds-ratios provides an easy way to interpret economically the estimates. For Mexico, the probability of choosing a new destination rises by 474% (=exp(1.557)) if it belongs to the same community of any of the firm's previous destinations.

---

[11] 79% of new regular exporters served the US market in the period 2003-2009.

Specification (2) introduces gravitational variables, such as GDP, GDP per capita, distance, common border, common language, belonging to the same regional trade agreement and migration stocks to proxy the similarity between the domestic (Mexico) and the new export market, and the attractiveness of the new export market. There is a reduction in the size of the positive community coefficient and still is strongly statistically significant. According to the new coefficients, once we control for gravity measures, the probability of choosing a new destination rises by 303% in Mexico if it belongs to the same community of any of the firm's previous destinations. Regarding gravitational variables, the larger the size of the new export market the higher the probability of choosing that market as a new export destination; and the larger the bilateral distance the lower the probability of selecting the new export destination. Speaking the same language, having a land border, belonging to the same trade agreement and a higher number of migrants raise the probability of choosing the new export destination. In contrast, the larger the new export market's income per capita the lower the probability of selecting that market.

Specification (3) also controls for extended-gravity measures and, again, the positive value of the community coefficient is further reduced: once we control for gravity and extended gravity measures, the probability of choosing a new export destination rises by 193% if it belongs to the same community of any of the firm's previous destinations.

Regarding extended gravity variables, having a previous export destination within 1,500 kilometers radius of the new destination raises the probability of selecting this destination. We also find that a new export destination has a larger probability of being chosen if there are previous destinations that speak the same language as this new destination, share a border, are located in the same income-quintile and region, and have sizable bilateral migration flows. In contrast, probability of being chosen is not affected when the new destination belongs to the same regional trade agreement of any previous destination.

To confirm that the community variable enhances the model's predictive capacity, we calculate the percentage of cases in which the observation with the highest predicted probability by the conditional logit model corresponds to the new destination selected by the exporter. We compare this percentage in a model that includes gravity and extended gravity variables and a model that also includes the community variable. We find that adding the community variable improves the accuracy of predictions by almost 20%.

Finally, specification (4) introduces destination-specific fixed effects. These fixed effects preclude the estimation of time-invariant gravity variables so, for the sake of clarity, we remove all gravity variables from the estimation. The community coefficients remain

positive and statistically significant at the 1 percent level. Once we control for gravity, extended-gravity and destination-specific effects, the probability of choosing a new export destination rises by 180%, almost a 3-fold increase, if it belongs to the same community of any of its previous destinations.

Table 5 presents the results of estimating the model with industry-specific communities and rest of industries communities. Both coefficients are positive and strongly statistically significant. As we expected, the industry-specific coefficient is larger than the rest of industries community coefficient. However, the differences are small and we cannot reject the null hypothesis that both coefficients are equal. Hence, we conclude that industry-specific communities and rest of industries communities play similar roles in determining exporters' export-path.

4.2. Sensitivity analyses

We perform a series of sensitivity analyses to assess the robustness of our results. First, we want to confirm that the communities identified by the network-analysis algorithm do not have a strong explanatory capacity by chance. To rule out this possibility we assign destinations to communities randomly. To carry out this exercise we assume that the number of communities and the number of members within each community is the same as in the benchmark estimations. We perform the exercise 50 times; each time, once the random communities are generated we run the same model as the one in Table 4 - specification (4). In all 50 estimations the community coefficient never was positive and statistically significant. These results point out that the communities generated by the network-analysis algorithm do not exert an influence on export dynamics by chance. On the contrary, it confirms that belonging to a community is a very important determinant of the evolution of the export path.

Second, we analyze whether the community coefficient is robust to the use of a larger sample to identify communities. We expand the sample including the years 2000 and 2001. To avoid marginal destinations, we exclude from the sample the destinations with less than 50 exporters during the period 2000-2002. The longer period and a less stringent threshold to admit a destination raises the number of destinations to 90 (65 previously). After applying iteratively the modularity maximization algorithm, we identify 12 communities. Compared to the sample used in the benchmark analyses, the number of communities rises in two (see Tables A4 in the Appendix). Table 6 presents the results of estimating the model with the new samples. The community coefficient remains positive and strongly statistically significant. Compared with the benchmark estimation, the value of the community coefficient increases.

According to the new estimates the probability of selecting a new export destination rises by 200% if it belongs to the same community of any of the firm's previous destinations.

Third, we analyze whether the community coefficients are robust to different extended distance variables. The variable used in the benchmark analyses is whether there is a previous export destination within a 1,500 kilometers radius of the new export destination. We re-estimate the model with a shorter distance: 500 kilometers, and a larger distance 3,000 kilometers. As shown in columns 2-3 in Table 6, the community coefficient is robust to the alternative extended-gravity distance radius.

Fourth, we use a more stringent threshold to determine whether a firm is a new exporter. Now, we define a firm as a new exporter if it does not export in 2000, 2001, 2002 and 2003. As shown in column 4 of Table 6, the community coefficient remains positive and statistically significant and is similar to those reported in the benchmark estimation.

Fifth, we convert the community and the extended gravity variables from discrete variables to count variables. For example, now, the community coefficient is the number of previous destinations that belong to the same community as the new destination. As shown in column 5 of Table 6, now the community coefficient reduces its positive value, but remains strongly statistically significant: a one unit increase in the number of destinations previous served by the exporter within the community rises the probability of choosing a new destination within the community by 128%.

Sixth, we analyze whether results for Mexico are robust to excluding all transactions with the US. As explained previously, 79% of new regular exporters in Mexico have the US in their portfolio and, hence, most of new regular exporters serve the community in which the US is integrated: 84%. To ensure that results are not driven by the large percentage of exporters having the US as destination, we remove all export transactions to the US from the new regular exporters' database. As shown in column 6 of Table 6, the community coefficient is very similar to the one reported in the benchmark analysis table so our estimations are not driven by the large percentage of new regular exporters that serve the US market.

Seventh, the results presented in the benchmark estimation use the final communities identified in Figure 2. In order to test the robustness of our results we also estimate the model with communities identified with fewer iterations of the modularity maximization algorithm. A shown in Figure 2, the maximum number of iterations needed to arrive to a final destination is five. So we can use the communities identified after four iterations, three iterations, two iterations and one iteration. As we reduce the number of iterations, the number of communities is also reduced. With a lower number of communities, clusters have a larger

number of members but we expect similarities between members to be lighter. These effects might drive the community coefficient in opposite directions. On the one hand, as there are more destinations within a community, exporters will have a larger probability of choosing a destination within a community. On the other hand, as similarities within the community are lighter, exporters will have a lower probability of choosing a destination belonging to the community.

Table 7 presents the results of the estimations for different community hierarchies. For comparison we also reproduce the results when estimating the model with the final communities (column 1). The community coefficient is always positive and statistically significant; the community coefficient only drops substantially when we only use one iteration.

Eighth, the (fixed-effects) conditional logit model imposes a strong restriction: IIA. The IIA states that the ratio of the probabilities of choosing a new export market only depends on the attributes of the two destinations, and is independent on the characteristics of other destinations. However, this restriction fails if some destinations have some (unobserved) common characteristics, making substitution among them easier. If the IIA restriction does not hold the conditional logit model leads to biased estimates. To address this limitation, we estimate alternative logit models that relax the IIA assumption: the nested-logit model and the mixed logit model. In the nested model, alternatives can be separated, at least, in two main groups. Within each group the IIA assumption holds, but across groups the IIA assumption does not need to hold. To implement the nested logit model we should determine a nesting criterion. We start assuming that firms decide, first, what major world region they want to sell to and, second, they select the destination within that region. We also used alternative nesting criteria, such as dividing destinations into close markets and distant markets, which no major changes in results. To estimate the nested logit model we assign destinations to one of the seven major regions defined by the World Bank.[12] The results after estimating the nested logit model are reported in Table 8. The community coefficient remains positive and highly statistically significant. As shown at the bottom of the table the LR test rejects the IIA assumption, validating the nesting of destinations by major regions.

A limitation of the nested logit model is that it introduces a very rigid substitution structure. For example, in our previous exercise a destination can only belong to one region. The mixed logit model overcomes this rigidity by allowing for variation in the coefficients

---

[12] See footnote 10.

across firms, unrestricted substitution patterns, and correlation in unobserved factors over time (Train, 2003). The last two columns of Table 8 present the results of estimating the mixed-logit model. One column presents the average value of each coefficient and the other column presents the average standard deviation of each coefficient. [13] The community coefficient remains positive and highly statistically significant. Looking to the standard deviation coefficient, we observe, as well, that there is a large heterogeneity in the impact of community across firms.

To sum up, the sensitivity analyses show that the positive and significant contribution of belonging to a community in determining the dynamics of firms' new export destinations is robust to the use of different samples and econometric specifications. We also show that this positive relation does not arise randomly.

5. Conclusions

How do exporters choose new export destinations? While there are many factors that are important for this decision, an empirical regularity strikes out: firms tend to choose new export markets that are similar to their prior export destinations. Network analysis, through the community-detection algorithm, provides a tool to identify destinations that share a common set of characteristics. This measure has three advantages over the gravity measures. First, it is a revealed measure and, hence, encapsulates all the observable and non-observable factors that may influence the degree of similarity among destinations. Second, it is a country-specific measure. Third, it allows the calculation of industry-specific similarity measures.

We apply this methodology to the web of Mexican exporters' destinations and find that there are ten communities in the Mexican web (of 65 countries). Next we show that belonging to the same community of a previous export destination exerts a strong influence on the dynamics of the exporter. In particular, the probability of choosing a new destination multiplies by three if it belongs to the same community of any of firm's previous destinations. We show that the strong predictive capacity of the network-based similarity measure remains once we control for gravity measures, and improves the accuracy of the model up to 20%. We also show that industry-specific community of destinations and general community of destinations exert a similar influence on the dynamics of Mexican exporters' destination-portfolio. Results are robust to different specifications and samples.

---

[13] In contrast to the conditional logit model, the nested and mixed logit models do not allow firms to choose more than one new destination per year. Hence, when estimating these latter models, we narrow the sample to exporters than only choose one new destination per year.

Appendix. Data sources.

Our firm-level data comes from transaction-level customs data on the universe of Mexican exporters over the period 2000-2009. The source for the data is detailed in the Annex of Cebeci, Fernandes, Freund and Pierola (2012) and the data was collected by the Trade and Integration Unit of the World Bank Research Department, as part of their efforts to build the Exporter Dynamics Database. In the Mexican dataset the firm identification code changes from 2007 onwards. For the year 2007 we have data with the old firm classification and with the new firm classification. Matching firm-level country and HS 6-digit specific records we can establish a correspondence between the old firm classification and the new firm classification for firms that exported in 2007. For the rest of firms that exported in 2008 and/or 2009, we cannot know whether they are new exporters or they are firms that exported in the 2000-2006 period. Since we use data on year 2002 for the network analysis and the sample of entrants that export at three consecutive years for the analysis of the dynamics of destination portfolio, this problem in the raw data does not affect our analysis. Table A1 in the Appendix display information on the number of trading firms, number of transactions and value (in millions US$) in the 2000-2009 period.

Data for the construction of the gravity and extended-gravity measures come from different sources. Data on income and population are taken from World Bank (2012). Distance, contiguity, common language, colonial relationship and same continent are obtained from Head et al. (2010). Bilateral migration stocks in 2000 are from Özden et al. (2011) and common membership in a regional trade agreement (RTA) in 2002 is obtained from de Sousa et al. (2012). The web links for databases open to the public are: World Development Indicators: http://data.worldbank.org/data-catalog/world-development-indicators; CEPII gravity dataset: http://www.cepii.fr/anglaisgraph/bdd/gravity.asp; RTA database: http://kellogg.nd.edu/faculty/fellows/bergstrand.shtml; World Bilateral Migration database: http://data.worldbank.org/data-catalog/global-bilateral-migration-database.

References

Albornoz, F., Calvo-Pardo, H., Corcos, G., and Ornelas, E. (2012). "Sequential exporting", Journal of International Economics, **88**, 1, 17-31

Besedes, T. and Prusa, Th. (2011), "The Role of Extensive and Intensive Margins and Export Growth", Journal of Development Economics, 96(2), 371-379

Blondel, V. D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E. (2008). "Fast unfolding of communities in large networks", Journal of Statistical Mechanics: Theory and Experiment, October, P10008.

Cebeci, T., Fernandes, A.M., Freund, C. and Pierola, M.D. (2012). "Exporter Dynamics Database", World Bank Policy Research Paper 6229.

Cadot, O., Iacovone, L., Pierola, M.D., Rauch, F. (2013), "Success and failure of African exporters″, Journal of Development Economics, 101, 284-296.

Chaney, T. (2011). "The network structure of international trade", NBER Working Paper 16753.

De Benedictis, L. and Tajoli, L. (2011). "The World Trade Network", The World Economy, **34**, 8, 1417-1454.

de Sousa, J., Mayer, T. and Zignano, S. (2012). "Market access in global and regional trade", Regional Science and Urban Economics, **42**, 6, 1037-1052.

Defever, F., Heid, B. and Larch, M. (2011). "Spatial exporters", CEP Discussion P. No. 1100.

Evenett, S.J. and Venables, A. J. (2002). "Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows", University of Bern Working Paper, mimeo.

Fortunato, S. and Berthélemy, M. (2007). "Resolution limit in community detection", Proceedings of the National Academy of Sciences, **104**, 1, 36-41.

Frankel, J., Stein, E. and Wei, S. (1995). "Trading Blocs and the Americas: The Natural, the Unnatural, and the Super-Natural", Journal of Development Economics, **47**, 1, 61-95.

Hidalgo, C.A., B. Klinger, A.L. Barabási and R. Hausmann, R. (2007). "The Product Space Conditions the Development of Nations", Science, **317**, 5837, 482-487.

Hubert, L. and Arabie, P. (1985). "Comparing partitions", Journal of Classification, **2**, 1, 193–218.

Kali, R. and Reyes, J. (2007). "The architecture of globalization: a network approach to international economic integration", Journal of International Business Studies, **38**, 4, 595-620.

Kali, R., Reyes, J., McGee, J. and Shirrell, S. (2013). "Growth networks", Journal of Development Economics, **101**, 216-227.

Lawless, M. (2011). "Marginal Distance: Does Export Experience Reduce Firm Trade Costs?", Central Bank of Ireland Technical Working Paper 02/RT/11.

McFadden, D. (1974). "Conditional Logit Analysis of Qualitative Choice Behavior", in P. Zarembka (eds.), Frontiers in Econometrics, New York: Academic Press, 105–142.

Morales, E., Sheu, G., and Zahler, A. (2011). "Gravity and extended gravity: Estimating a structural model of export entry", unpublished paper.

Newman, M.E. (2006). "Modularity and community structure in networks", Proceedings of the National Academy of Sciences, **103**, 23, 8577-8582

Nguyen, D. X. (2012). "Demand uncertainty: Exporting delays and exporting failures", Journal of International Economics, **86**, 2, 336-344.

Özden, C., Parsons, C.R., Schiff, M. and Walmsley, T.L. (2011). "Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000", World Bank Policy Research Working Paper 5709.

Train, K.E. (2003). Discrete Choice Methods with Simulation, Cambridge University Press, Cambridge.

Figure 1. The network of Mexican firms export destinations, 2002
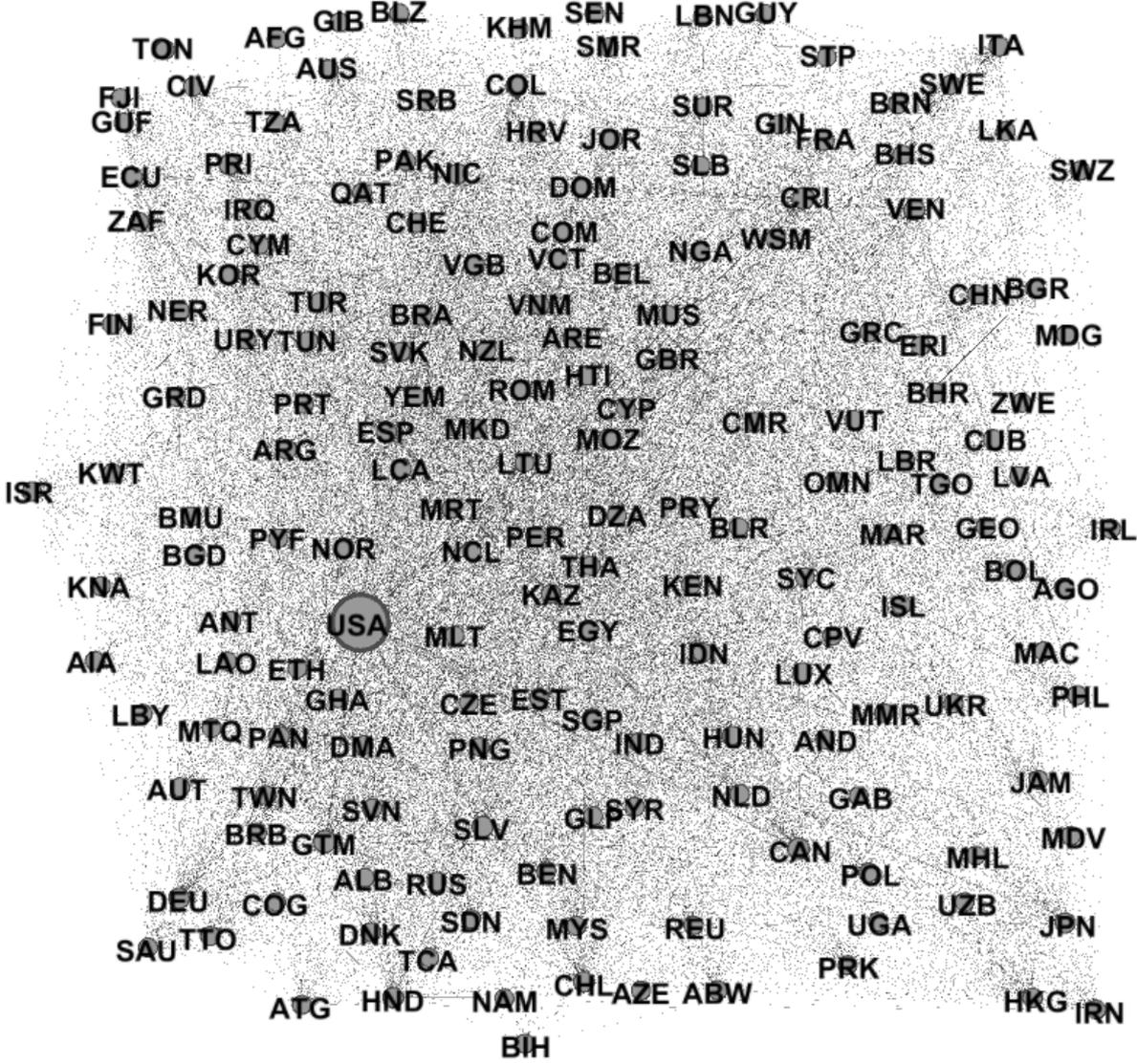
Figure 2. Community detection process in the network of Mexican exporters' destinations
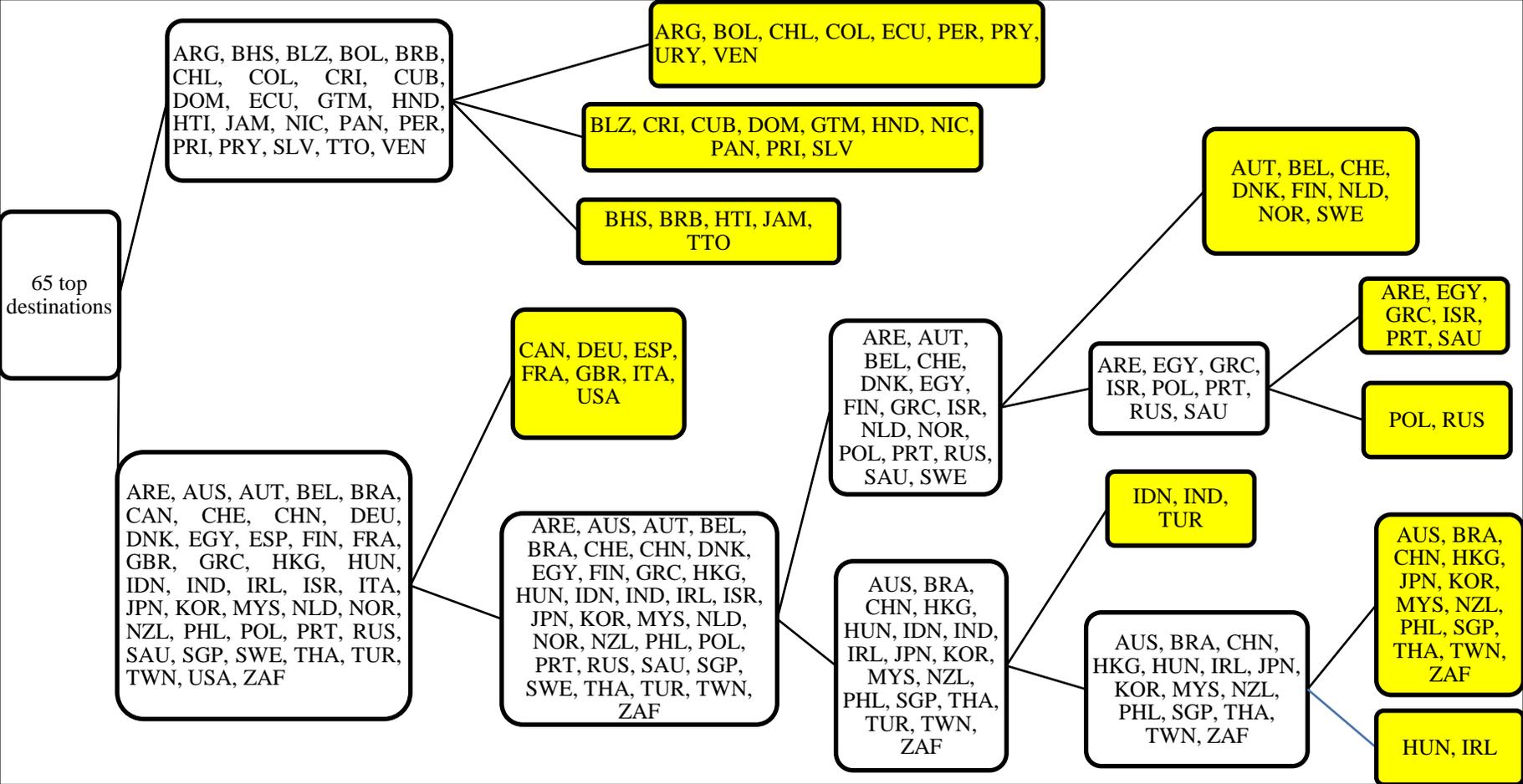
Table 1. Regression results of the probability of belonging to the same community on

gravity measures

Distance (Ln)	-0.044*** (0.009)
Border	0.133*** (0.028)
Language	0.135*** (0.013)
Region	0.177*** (0.016)
Income	0.034*** (0.009)
Common colonizer	0.034 (0.027)
Regional trade agreement	-0.020 (0.016)
Migration flows (Ln)	0.011*** (0.001)
Constant	0.359*** (0.085)
R-square	0.26
Observations	4,160

Note: Linear probabilistic model. Standard deviations in parentheses. ***, ** statistically significant at 1% and 5% respectively.

Table 2. Changes in export destination portfolio of Mexican "new regular exporters".

Type of firm	(1)	(2)	(3)	(4)	(5)	(6) Regular exporters 2000-2009
	New exporting firms since 2003 that do not stop exporting					
Years exporting	**03-04**	**04-05**	**05-06**	**06-07**	**07-08**	**07-08**
# new regular exporters	757	948	1110	1283	1928	
# firms in each period	757	1705	2815	4098	6026	5697
Changes in country portfolio (%)						
Only entries	21	17	15	14	15	16
Only exits	4	6	8	8	9	7
Simultaneous entry & exit						
same entries and exits	5	6	5	6	6	6
entries>exits	8	7	7	7	7	7
entries < exits	1	3	4	4	4	3
No change in portfolio	61	62	61	61	60	61

Source: Own elaboration using Census of Exporting Mexican Firms, 2000-2009.

Table 3. Dependent variable. Number of new export destinations per firm-year.

# entries per firm-year pair	Frequency	Cum. Freq.
1	4104	56.63
2	1524	21.03
3	701	9.67
4	363	5.01
5	212	2.93
6	129	1.78
7	54	0.75
8	49	0.68
9	41	0.57
10	38	0.52
11 or more	32	0.44
Total	7247	100.00

Table 4. Main results. Conditional logit estimations. Baseline results. (Dependent variable: Choice of new export destination)

Specification	(1)	(2)	(3)	(4)
Community	1.557*** (0.025)	1.108*** (0.027)	0.662*** (0.013)	0.587*** (0.031)
GDP		0.487*** (0.009)	0.530*** (0.009)	
GDP pc		-0.187*** (0.012)	-0.212*** (0.012)	
Distance		-0.995*** (0.026)	-0.827*** (0.026)	
Border		0.265*** (0.040)	0.573*** (0.042)	
Language		0.444*** (0.031)	0.484*** (0.032)	
RTA		0.107*** (0.024)	0.154*** (0.024)	
Migrants		0.034*** (0.004)	0.017*** (0.005)	
I_distance 1500			0.069** (0.024)	0.334*** (0.030)
I_border			0.426*** (0.030)	0.436*** (0.030)
I_language			0.338*** (0.026)	0.331*** (0.037)
I_RTA			-0.039 (0.031)	0.216*** (0.034)
I_income			0.130*** (0.031)	0.174*** (0.030)
I_migration			0.810*** (0.058)	0.503*** (0.059)
I_region			0.425*** (0.037)	0.523*** (0.039)
Country dummies	No	No	No	Yes
Observations	446,161	446,161	446,161	446,161
Nº of firms	3,339	3,339	3,339	3,339
Nº of countries	65	65	65	65
R2	0.07	0.16	0.18	0.21

Note: Clustered standard errors at the firm-level in parentheses. ***, ** significant at 1 percent and 5 percent respectively.

Table 5. Estimations with sector-specific communities

Dependent variable: Choice of new export destination.

	Mexico
I _Community sector	0.447*** (0.030)
I _Community no-sector	0.370 *** (0.036)
I _distance 1500	0.313*** (0.030)
I _border	0.379*** (0.031)
I _language	0.363*** (0.038)
I _RTA	0.208*** (0.035)
I _income	0.139*** (0.031)
I _migration	0.465*** (0.062)
I _region	0.583*** (0.040)
Observations	407,742
Nº of firms	3,440
Nº of countries	55
R2	0.21

Note: All estimations include destination-specific fixed effects. Clustered standard errors at the firm-level in parentheses. *** significant at 1 percent.

Table 6. Sensitivity Analyses I (Dependent variable: Choice of new export destination)

Specification	(1) Larger sample	(2) N_distance 500	(3) N_distance 3000	(4) New exporters	(5) Count	(6) No US
Community	0.693***	0.587***	0.595***	0.560***	0.250***	0.572***
	(0.032)	(0.031)	(0.031)	(0.034)	(0.012)	(0.031)
I_distance 1500	0.304***			0.325***	0.070***	0.299***
	(0.030)			(0.033)	(0.020)	(0.030)
D_distance 500		0.266***				
		(0.046)				
D_distance 3000			0.090***			
			(0.034)			
D_border	0.400***	0.347***	0.434***	0.390***	0.294***	0.444***
	(0.030)	(0.033)	(0.030)	(0.034)	(0.022)	(0.031)
D_language	0.037***	0.336***	0.340***	0.379***	0.036***	0.397***
	(0.009)	(0.037)	(0.037)	(0.041)	(0.009)	(0.040)
D_RTA	0.189***	0.233***	0.234***	0.250***	0.027***	0.248***
	(0.034)	(0.034)	(0.034)	(0.038)	(0.005)	(0.038)
D_income	0.150***	0.155***	0.173***	0.175***	0.066***	0.180***
	(0.030)	(0.030)	(0.030)	(0.034)	(0.010)	(0.032)
D_migration	0.536***	0.470***	0.490***	0.483***	0.007	0.477***
	(0.059)	(0.058)	(0.058)	(0.064)	(0.009)	(0.061)
D_region	0.581***	0.619***	0.601***	0.541***	0.071***	0.517***
	(0.039)	(0.038)	(0.039)	(0.044)	(0.011)	(0.042)
Observations	468,883	446,161	446,161	357,715	446,161	333,126
Nº of firms	3,328	3,320	3,320	2,826	3,320	2,750
Nº of countries	90	65	65	65	65	64
R2	0.23	0.20	0.20	0.21	0.19	0.18

Note: All regressions include destination-specific fixed-effects. Clustered standard errors at the firm-level in parentheses. ***, ** significant at 1 percent and 5 percent respectively.

Table 7. Sensitivity Analyses II: Estimations for different community hierarchies (Dependent variable: Choice of new export destination).

Community type	Final	4 iterations	3 iterations	2 iterations	1 iteration
Community	0.587***	0.592***	0.612***	0.625***	0.473***
	(0.031)	(0.030)	(0.031)	(0.030)	(0.044)
D_distance 1500	0.334***	0.332***	0.335***	0.335***	0.366***
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
D_border	0.436***	0.435***	0.436***	0.434***	0.493***
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
D_language	0.331***	0.332***	0.339***	0.332***	0.437***
	(0.037)	(0.037)	(0.037)	(0.037)	(0.036)
D_RTA	0.216***	0.215***	0.204***	0.199***	0.320***
	(0.034)	(0.034)	(0.034)	(0.034)	(0.033)
D_income	0.174***	0.171***	0.166***	0.157***	0.294***
	(0.030)	(0.030)	(0.031)	(0.031)	(0.030)
D_migration	0.503***	0.498***	0.477***	0.473***	0.395***
	(0.059)	(0.058)	(0.059)	(0.058)	(0.058)
D_region	0.523***	0.525***	0.524***	0.556***	0.608***
	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
Observations	446,161	446,161	446,161	446,161	446,161
Nº of firms	3,339	3,339	3,339	3,339	3,339
Nº of countries	65	65	65	65	65
R2	0.21	0.21	0.21	0.21	0.20

Note: All estimations include destination-specific fixed effects. Clustered standard errors at the firm-level in parentheses. ***, **, * significant at 1 percent and 5 percent respectively.

Number of observations=446161. Number of firms: 3339. Number of countries: 65.

Table 8. Sensitivity Analyses III: Nested-Logit and Mixed -Logit model estimations (Dependent variable: Choice of new export destination).

Specification	Nested-Logit	Mixed-Logit Mean	Mixed-logit Standard deviation
Community	0.760*** (0.061)	0.613*** (0.054)	0.959*** (0.113)
D_distance 1500	0.561*** (0.086)	0.193** (0.083)	0.544*** (0.197)
D_border	0.932*** (0.093)	0.493*** (0.067)	0.599*** (0.164)
D_language	0.587*** (0.073)	0.340*** (0.048)	0.368* (0.191)
D_RTA	0.226*** (0.068)	-0.094* (0.048)	0.094 (0.136)
D_income	0.262*** (0.063)	0.210*** (0.056)	0.708*** (0.133)
D_migration	0.801*** (0.131)	1.051*** (0.181)	0.742** (0.303)
D_region		0.392*** (0.066)	0.561*** (0.153)
LR-Test for IIA (phi value)	0.0000		

Note: The mixed-logit estimation also includes gravity-type controls. Standard deviation in parentheses. ***, **, * significant at 1 percent, 5 percent and 10 percent respectively. Number of observations=250157. Number of firms: 2625. Number of countries: 65.

Table A1. Mexican exporters database (2000-2009)

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Number of exporting firms	35.509	34.318	31.592	30.420	30.441	30.984	30.171	30.283	29.796	28.690
Number of transactions	183.586	181.343	171.029	161.899	167.905	173.620	174.308	187.267	190.584	183.036
Value of exports (current US$ million)	165.974	158.539	160.669	164.941	187.736	213.902	249.510	270.776	290.160	228.728
Regular exporters (N=5697) (% firms)	16%	17%	18%	19%	19%	18%	19%	19%	19%	20%
% total transactions	35%	38%	40%	40%	43%	46%	46%	46%	46%	46%
% value of total exports	60%	64%	65%	66%	66%	66%	67%	67%	68%	69%
Regular new exporters Number				757	948	1.110	1.283	1.928		6.026
(% in total firms)				2%	3%	4%	4%	6%		21%
Transactions				3.207	3.907	5.081	5.145	7.177		41.571
(% in total transactions)				2%	2%	3%	3%	4%		23%
Value of exports				30.890	29.351	28.498	33.411	62.596		47.805
(% in total exports)				19%	16%	13%	13%	23%		21%
Memorandum										
Merchandise exports (Source: WDI)	166.367	158.547	160.682	165.396	187.980	214.207	249.961	271.821	291.265	229.712

Note: The term "firm" refers to any individual operator that makes a transaction in a year. The dataset contains all the transactions with a value above 1,500US$. The area in dark grey refers to the number of firms, number of transactions and value of exports in 2009 of the new regular exporters, that is, firms that started exporting after 2003 and did not stop exporting until 2009 (our Mexican firms sample).

Table A2. Communities by industries. Mexico 2002

Agriculture	Chemicals	Machinery & Transport equipment	Metals	Non-metallic minerals	Paper	Textiles
ARG	BOL	CHL	BOL	ARG	CHL	CHL
BOL	CHL	COL	CHL	BRA	COL	COL
BRA	COL	CRI	COL	CHL	CRI	CRI
CHL	CUB	CUB	ECU	COL	CUB	CUB
COL	ECU	DOM	PER	ECU	DOM	DOM
DOM	PAN	ECU	VEN	ESP	ECU	ECU
ECU	PER	GTM	BHS	PER	GTM	GTM
PER	VEN	HND	BLZ	VEN	HND	HND
PRI	BHS	NIC	BRB	BLZ	JAM	NIC
VEN	BRB	PAN	HTI	GRC	NIC	PAN
BHS	JAM	PER	JAM	HTI	PAN	PER
BLZ	TTO	SLV	TTO	JAM	SLV	PRI
BRB	CRI	USA	CRI	TTO	VEN	SLV
CUB	DOM	VEN	CUB	BHS	ARG	VEN
HTI	GTM	BHS	DOM	CHE	BRA	BZL
JAM	HND	BLZ	GTM	CRI	CAN	BOL
TTO	NIC	BOL	HND	CUB	HKG	BRB
URY	SLV	BRB	NIC	DOM	IDN	HTI
CRI	ARE	HTI	PAN	GTM	KOR	JAM
GTM	CHN	JAM	PRI	HND	MYS	TTO
HND	HKG	PRI	SLV	NIC	PER	CHN
NIC	IDN	TTO	HKG	PAN	PHL	HKG
PAN	KOR	ARG	HUN	PRI	URY	HUN
SLV	THA	IDN	MYS	SLV	AUS	THA
CHN	TWN	MYS	NLD	USA	AUT	TWN
KOR	AUS	NZL	SGP	ARE	BEL	BRA
PHL	GRC	PHL	THA	IDN	BHS	CAN
THA	MYS	SGP	TWN	KOR	BRB	CHE
HKG	NLZ	THA	AUS	THA	CAN	DEU
IDN	PHL	TWN	AUT	TWN	DEU	IDN
NZL	SGP	URY	BEL	CHN	ESP	KOR
SGP	DEU	ZAF	BRA	HKG	FIN	PHL
TWN	ESP	AUS	CHN	JPN	FRA	SWE
AUS	FRA	BEL	GBR	PHL	GBR	USA
BEL	GBR	BRA	JPN	ZAF	GRC	ARE
CAN	ITA	CAN	KOR	CAN	HUN	BEL
CHE	JPN	CHN	ZAF	FRA	ITA	BHS
DEU	BLZ	DEU	CAN	GBR	JPN	ESP
ESP	CAN	ESP	CHE	ITA	NLD	FRA
GRA	HTI	FRA	DEU	SAU	NZL	GRC
GBR	PRI	GBR	ESP	SWE	PRI	ITA
ITA	USA	HKG	FRA	FIN	PRT	NLD
JPN	FIN	ITA	ITA	MYS	SAU	PRT
NLD	HUN	JPN	USA	NLZ	SWE	AUS
SWE	SAU	KOR	ARE	NLD	TWN	FIN
USA	ARG	NLD	SAU	BOL	USA	GBR
ARE	BRA	ARE	ARG	HUN	ARE	JPN
GRC	PRT	FIN	IDN	SGP	BLZ	MYS
MYS	URY	GRC	PHL	URY	BOL	SGP
PRT	ZAF	SAU	URY	AUS	SGP	ZAF
SAU	AUT	AUT	FIN	AUT	THA	ARG
AUT	BEL	CHE	GRC	BEL	TTO	AUT
FIN	CHE	HUN	NZL	BRB	CHE	NZL
HUN	NLD	PRT	PRT	DEU	HTI	SAU
ZAF	SWE	SWE	SWE	NLD	ZAF	URY
				PRT		

Table A3. Adjusted Rand indexes.

	Agriculture	Chemicals	Machinery & Transport equipment	Metals	Non-metallic minerals	Paper	Textiles	Industry partitions vs. rest of industries partitions
Agriculture	**1.00**							0.36
Chemicals	0.31	**1.00**						0.14
Machinery&Transport	0.34	0.32	**1.00**					0.38
Metals	0.37	0.31	0.38	**1.00**				0.25
Non-metallic minerals	0.26	0.24	0.24	0.32	**1.00**			0.31
Paper	0.23	0.16	0.28	0.18	0.17	**1.00**		0.31
Textiles	0.26	0.25	0.41	0.35	0.23	0.25	**1.00**	0.22

Table A4. Community detection process in the network of Mexican exporters' destinations using the 2000-2002 sample

ARG, BOL, CHL, COL, ECU, PER, PRY, URY, VEN
BLZ, CRI, CUB, DOM, GTM, HND, NIC, PAN, PRI, SLV
DMA, HTI, JAM, SUR, TTO
ABW, ANT, ATG, BHS, BMU, BRB, CYM, LCA, VCT, GUY
BEL, BRA, CAN, CHE, DEU, ESP, FRA, GBR, ITA, JPN, NLD, USA
ARE, EGY, JOR, KWT, LBN, SAU, SYR
DZA, LKA, MAR, PAK, TUR
DNK, FIN, NOR, SWE
AUT, CZE, HUN, IRL
NGA, VGB
CHN, HKG, IDN, IND, KOR, MYS, PHL, PRK, SGP, THA, TWN, VNM
AUS, NZL, ZAF