



ARTICLE

Received 20 May 2015 | Accepted 18 Apr 2016 | Published 24 May 2016

DOI: 10.1057/palcomms.2016.21

OPEN

Assessing the role of migration as trade-facilitator using the statistical mechanics of cooperative systems

Adriano Barra¹, Andrea Galluzzi², Daniele Tantari³, Elena Agliari² and Francisco Requena-Silvente⁴

ABSTRACT Interactions between natives and foreign-born individuals may help to stimulate the development and the diversification of bilateral trade relationships. In fact, migrants act as trade facilitators reducing transaction costs in export activities and, consequently, more local firms are able to establish new trade relationships abroad. The pro-trade effect of migration is well evidenced in several works where the shape of the trade-migration relationship has been examined empirically; however, they all lack an analytical model that enables them to predict the expected non-linear relationship between migration and trade. Here, using statistical mechanics tools we develop a simple model that demonstrates that there is a positive non-linear relationship between the extensive margin of trade and the proportion of migrants in the total population. Data on Spanish trade and migration provide support for the predictions made by the theoretical model. The model also suggests the need of a critical mass of migrants before their interactions with the natives have any effective impact on trade. The threshold is sensitive to the nationality of the migrants, suggesting that cultural differences between natives and migrants may affect the number of migrants needed to generate a positive impact on trade. Furthermore, we examine the possible relationship between the share of migrants in the total population and the extent of diversification of the portfolio of exported goods, finding evidence of a strong positive correlation. Our approach can be used to examine other related issues such as the impact of formal or informal firm networks on trade.

¹ Dipartimento di Fisica, Sapienza Università di Roma, Rome, Italy ² Dipartimento di Matematica, Sapienza Università di Roma, Rome, Italy ³ Centro Ennio De Giorgi, Scuola Normale Superiore, Pisa, Italy ⁴ Department de Estructura Econòmica, Valencia, Spain

Introduction

Since Gould's (1994) pioneering paper, there has been significant growth in the number of studies examining the economics of the migration-trade nexus (see, for example, Head and Rises, 1998; Partridge and Furtan, 2008; Hunt and Gauthier-Loiselle, 2009; Ozgen *et al.*, 2011; Egger *et al.*, 2012; Rashidi and Pyka, 2013). Most of this research has provided evidence in favour of the pro-trade effect of migration (see Genc *et al.*, 2012 for a meta-analysis). The basic ideas underlying the literature is that operating in international markets is difficult due in part to large transaction costs and that migrants may help to reduce these costs. Local firms may have access to crucial knowledge about business opportunities, demand characteristics and institutional features that exist abroad as a result of interacting with foreign-born people.

Only a few papers have examined empirically the shape of the trade-migration relationship. Using time-series data for the United States, Gould (1994) found that the pro-trade effect of migration disappeared when the number of migrants reached certain level. Egger *et al.* (2012) confirmed for a sample of OECD countries in 2005 that there is a saturation point in the trade-migration relationship. Using regional data for Spain and Italy in 2007, Serrano-Domingo and Requena-Silvente (2013) found that the migration-trade link is U-inverted, that is, when the number of migrants is too small or too large the pro-trade effect is small. The three papers derive the conclusions after searching for non-linearities in the data. However, they all lack an analytical model that enables them to predict the expected non-linear relationship between migration and trade.

Here, we use tools from statistical mechanics to examine the trade-migration link. Starting from a bipartite network formed by two types of population living in the same space (natives and migrants), we examine how the interactions between natives and migrants affect the decisions of natives about trading abroad, and we get a simple theoretical expression that explains how the decision to trade (export) abroad by the natives in a certain territory depends on the share of migrants living there.

Three interesting features emerge from our model. First, the relevant measure of trade to evaluate the migration-trade link is the extensive margin of trade; that is, the number of exporting firms and the number of exported products. While most empirical papers have examined the value of bilateral trade, only a few papers (Peri and Requena-Silvente, 2010; Bastos and Silva, 2012; Bahar and Rapoport, 2014) provide empirical evidence in favour of migrants stimulating trade through more firms trading, more transactions or more exported products.

Second, there is a minimum threshold in the mass of migrants living in a certain territory before any positive pro-trade effect emerges. Previous studies, discussed above, found some evidence of an optimal (maximum) number of migrants but not of a minimum. The existence of a minimum threshold makes sense in our model because the number of interactions between natives and migrants determines the likelihood that migrants may influence in natives' decisions about trading abroad. Third, once the share of migrants reaches a certain level, the migration-trade link has a non-linear relationship; more specifically, it is a square root. As a result, the number of trade relationships abroad grows fast as the share of migrants on the total population exceeds the minimum threshold, while for relatively large values of the share of immigrants it grows slower, eventually reaching a saturation value. This phenomenon ultimately stems from the underlying collective interactions between natives and immigrants. In physics, it is typically evidence of a phase transition, namely, a relatively sharp change in a system's behaviour occurring from a slight perturbation of the system parameters. In this case, the *system* is the ensemble of all natives involved in trade in a given

territory, whose average behaviour is described by our measures of the extensive margin of trade (number of exporting firms or number of exported products), and the *perturbation* is encoded by a variation in the share of migrants in the total population.

Next, we use newly available data for Spain over the period 2000–2012 on the extensive margin of trade (the number of exporting firms and the diversification of export product portfolio) in order to test the predictions of the model. Spain has been the subject of previous studies examining the migration-trade link because of its peculiarities compared with other OECD countries (see, for example, Peri and Requena-Silvente, 2010; Artal-Tur *et al.*, 2012; Serrano-Domingo and Requena-Silvente, 2013). In fact, Spain had one of the lowest immigration rates among OECD countries before 2000 and received massive inflows of migrants in a relatively short period of time (2000–2008). This constitutes an ideal context to investigate the role of immigrants in creating new trade relationships abroad.

Spanish data validates the theoretical model. It is shown that the number of bilateral trade relationships abroad (proxied by the number of firms exporting) is a square root of the percentage of migrants in the host territory beyond a critical value, which captures the minimum share of migrants needed to generate enough interactions between natives and migrants in a defined territory. Such a minimum threshold is found to be positive, though very small. Moreover, it varies with the migrants' countries of origin, suggesting that the pro-trade effect of migration is affected by cultural differences.

We further deepen the analysis by relating the presence of immigrants with the product diversification of the hosting country. Using a Herfindahl index (H-index) to quantify the inverse of product diversification of exports, we find a strong negative connection between the share of immigrants and the degree of export product concentration. We interpret that result as evidence of migrants helping to increase the portfolio of exported products in the host country.

The remainder of the article is organized as follows: first we illustrate the use of tools from statistical-mechanics to investigate the migration-trade link. Next, we present the results concerning the (non-linear) relationship between the share of migrants and the number of exporting firms, as well as between the share of migrants and the extent of product diversification in exports. In the Supplementary Information we explain in further detail the steps used to solve the model and we investigate the topological properties of the network built between natives and migrants.

Methodology

While the overarching theme of this article is in the field of social sciences, the analytical tools employed come from statistical mechanics. The latter aims to bridge two different descriptions of the same perceivable reality that usually lie at two different scales. For example, its original implementation in physics was designed to obtain a unified and consistent picture of a microscopic world (made up of molecules and atoms obeying the time-reversible Newton mechanics) with a macroscopic world (obeying the phenomenological rules of thermodynamics, where time reversibility is lost and whose evolution is described in terms of mean values and fluctuations). Nowadays, statistical mechanics has been used successfully in several fields of research, ranging from chemistry (Agliari *et al.*, 2015) and biology (Buchanan, 2010) to sociology (Durlauf, 1999), finance (Bouchaud and Potters, 2009) and economics (Tacchella, 2002). Here, the microscopic world is the network of connections between natives and migrants and the macroscopic world is the development and diversification of exports in the territory where natives and migrants live.

Our starting point is a bipartite network where each link connects two types of agents: natives and immigrants. The link permits the transmission of information between the agents and accounts for any kind of relationships between natives and migrants (from family or friend links, to strictly professional ones). For simplicity, the links are randomly allocated across pairs of agents. This is a valid assumption because on a large scale (that is, when accounting for relatively wide territorial districts) and with a long time window (that is, with a time record over more than 10 years) possible correlations get smoothed out. In our model, a native is a decision maker who decides whether to trade or not with a foreign country. The decision (captured by σ in our model) can be either positive or negative. The likelihood of a native trading with a foreign country depends on the amount of information that other neighbouring natives and migrants have about the foreign country. In particular, in this work we focus on the information conveyed by the latter (represented by z in our model). Our model ignores other information channels such as own marketing research or news from the media.

In physics, our model is said to be cooperative (or imitative) because the variables σ and z associated with the connected nodes tend to be aligned (namely either both positive or both negative). Clearly, in principle, one cannot exclude the existence of anti-cooperative (or anti-imitative) connections among some pairs of agents but, as long as the prevailing kind of interaction is imitative, the qualitative results of the model are robust.

Finally, a degree of randomness (captured by β in the model) is introduced. The underlying idea is that decision makers do not cooperate in a completely rational way: even if the overall information passed to a native is positive, there is still a possibility that σ is negative. When β tends to zero, decision makers behave randomly; that is, their propensity to trade σ is positive with probability 1/2 and negative with probability 1/2, regardless of the neighbourhood conditions. Conversely, when β becomes very large, decision makers cooperate deterministically, that is, σ is just the sign of the overall information gathered by the native.

Before solving the model, we show that it is possible to move from the original bipartite network to a monopartite network made by natives only and where the reciprocal interactions take into account the information collected by natives in their interplay with migrants. The idea is that some natives may learn from migrants and they inform other natives.

The last step is to solve the model using the statistical mechanics machinery. The system's solution allows a theoretical estimate to be obtained for the average number of decision makers who intend to trade with foreign countries, based on the strength of the interactions between natives and immigrants. In the empirical section, this estimated number is compared with the observed number of bilateral trade relationships that Spain establishes abroad. This comparison connects a microscopic observable with a macroscopic observable. In statistical mechanics such a change of scale is a common route but some checks must be carried out before. In particular, we need to prove that there exists a linear proportionality between the total population and the total number of firms. All the technical details about this passage (along with a check of the underlying assumptions) can be found in the Appendix A of Supplementary Information.

The system's solution provides a very simple relationship between the extensive margin of trade (the decision taken by natives about selling abroad) and the share of migrants in total population. Interestingly, the relationship is positive but not linear (in fact it is a square root), and it only occurs beyond certain cut-off: below a certain share of migrants there is no connection between migration and trade.

It is worth underlining that statistical mechanics provides a description of the system considered just in terms of *average* behaviour. In an ideal experiment one can produce several independent realizations of the same system and their average is expected to converge to the theoretical prediction. As discussed later, for the real system considered here we can reasonably take each of the Spanish provinces as an independent realization and then average the related outcomes. Such an average will be compared with the theoretical outcomes from statistical mechanics.

The model

We assume a population of N agents divided into two groups: N_1 natives and N_2 immigrants, being $N_1 + N_2 = N$ and

$$\gamma \equiv \frac{N_2}{N}, \quad (1)$$

$$1 - \gamma \equiv \frac{N_1}{N}, \quad (2)$$

$$\Gamma \equiv \gamma(1 - \gamma), \quad (3)$$

where γ and $1 - \gamma$ measure the relative size of each group, respectively, while Γ represents the normalized number of cross links between the two communities. For a small γ , $\Gamma \approx \gamma$. Since the share of migrants in a particular territory tends to be small (and certainly it is the case of Spain), we will use Γ to capture γ .

Next, we introduce variables $\{\sigma_i\}_{i=1}^{N_1}$ and $\{z_\mu\}_{\mu=1}^{N_2}$, such that $\sigma_i \in \{-1, +1\}$ represents the propensity of the native agent i to establish ($\sigma_i = +1$) or not establish ($\sigma_i = -1$) a trade, while the variables $z_\mu \in \mathbb{R}$ encode the information that the μ -th immigrant can provide regarding trading toward her country of origin. The sign of z_μ indicates whether the information favors ($z_\mu > 0$) or not ($z_\mu < 0$) the establishment of a trade, while its magnitude indicates the strength of the signal. The ensemble $\{z_\mu\}_{\mu=1}^{N_2}$ represents the *social capital* of the immigrant community and, in the absence of any additional information, in a mean-field approach, it can be thought of as a collection of Gaussian variables identically and independently distributed¹

The decisional mechanism results from the interactions between natives and migrants and it is described by a Hamiltonian (that is, a *cost function* in the economics vocabulary) denoted by $\mathcal{H}(\sigma, z; J, \xi)$. This function is dependent on the native-native interactions and on the native-migrant interactions, encoded by J and ξ , respectively (see Fig. 1, left panel), as well as on the overall configurations of agents themselves (that is, $\{\sigma\}$ for natives and $\{z\}$ for migrants). We now inspect in more detail the interaction patterns and the resulting Hamiltonian.

The interaction between the i th native and the μ th immigrant is encoded by the variable $\xi_i^\mu \in \{0, 1\}$ describing the presence ($\xi_i^\mu = 1$) or the absence ($\xi_i^\mu = 0$) of a connection (for example, friend, work colleague, acquaintance, relative, workmate and client) between i and μ . Since we lack detailed information about individual connections and since migration flows are reasonably uncorrelated (that is, the time-scales considered are long enough and migrants come from a wide range of countries), the most basic assumption one can pose is simply to consider the completely general set of ξ_i^μ as i.i.d. aleatory variables, extracted with probability

$$\mathbb{P}(\xi_i^\mu = 1) = 1 - \mathbb{P}(\xi_i^\mu = 0) = \frac{\xi}{N^\theta}, \quad (4)$$

where $\theta \in (0, 1)$, and $\xi \in \mathbb{R}^+$ are province-dependent parameters. In this way, as developed further in the Appendix C of Supplementary Information, by properly tuning θ and ξ , the network recovers all the standard topological regimes (for

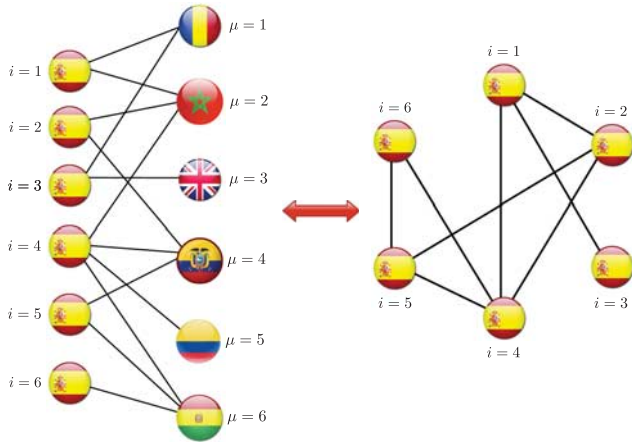


Figure 1 | Representation of the networks describing the mutual interactions among the agents making up the system under study. Left panel: Sketch of the bipartite network modelling mutual interactions between natives (left community) and immigrants (right community). Each node represents an agent, whose flag mirrors her country of origin (for migrants we simply show the six most represented communities in Spain). To simplify the representation of the network, we omit the links representing interactions among Spanish agents. Right panel: Sketch of the equivalent monopartite network built by Spanish decision makers only; this exhibits the same properties (that is, the same first order statistics on all its observables) with the original bipartite one. Basically, in this monopartite network, two agents are connected if they share at least one neighbour in the bipartite network. For instance, nodes labelled as 1 and 2 share one neighbour (corresponding to $\mu = 2$), while nodes labelled as 1 and 6 do not share any neighbour.

example, extreme dilution, finite connectivity and so on) and, accordingly, different degrees of interaction among the two communities are captured.

For the set of variables, J , its entry J_{ij} describes the connection between natives i and j . We assume these couplings to be arbitrary but endowed with a well defined, positive average value \bar{J} , that, in principle, depends on the province p .

In fact, the bulk of the migration-trade relation is known to be an *in-province* phenomenon: exports from a province to a given foreign country do not receive significant stimuli by immigrants coming from that country but living in a different province (Artal-Tur *et al.*, 2012). Therefore, we can fix the *resolution* at the provincial level and treat each province as a different sample over which we will average. Accordingly, the system to be modelled represents any province in Spain.

Therefore, at the provincial level of resolution, the system can be described by the Hamiltonian

$$\mathcal{H}(\sigma, z; J, \xi) = -\frac{1}{N_1} \sum_{(ij)}^{N_1} J_{ij} \sigma_i \sigma_j - \frac{1}{N^{1-\theta}} \sum_{i=1}^{N_1} \sum_{\mu=1}^{N_2} \xi_i^\mu \sigma_i z_\mu, \quad (5)$$

$$\approx -\frac{\bar{J}}{N_1} \sum_{(ij)}^{N_1} \sigma_i \sigma_j - \frac{1}{N^{1-\theta}} \sum_{i=1}^{N_1} \sum_{\mu=1}^{N_2} \xi_i^\mu \sigma_i z_\mu, \quad (6)$$

where the size-dependent factors in front of the sums ensure that the cost function is well defined (that is, linearly extensive with respect to the system size). It should be noted that, because of the minus sign in front of each term in the Hamiltonian above, its minimization tends to *align* the decisions of the agents. Thus, $\mathcal{H}(\sigma, z; J, \xi)$ constitutes the mathematical expression that

describes the imitative behaviour at the level of single agents (Brock and Durlauf, 2001).

Before proceeding, we need to introduce a parameter β to tune the degree of stochasticity in the system, in such a way that for $\beta \rightarrow 0$ the system behaves completely randomly, while as $\beta \rightarrow \infty$ the system deterministically relaxes to the configuration corresponding to the minimum of the cost function. Following the statistical mechanics machinery, we associate with each configuration of the variables $\{\sigma, z\}$ a probability of being observed given by the so called *Boltzmannfaktor* weight $P_\beta(\sigma, z; J, \xi)$, that reads as

$$P_\beta(\sigma, z; J, \xi) = \frac{\exp[-\beta \mathcal{H}(\sigma, z; J, \xi)]}{Z(\beta, J, \xi)}, \quad (7)$$

where, the normalizing denominator $Z(\beta, J, \xi)$ is called *partition function* of the model and is defined as the sum over all the possible 2^{N_1} configurations $\{\sigma\}$ and the Gaussian integral over all the possible configurations $\{z\}$ (hence ensuring $P_\beta(\sigma, z; J, \xi)$ to be a probability), namely

$$Z(\beta, J, \xi) = \sum_{\{\sigma\}} e^{\frac{\beta \bar{J}}{N_1} \sum_{ij} \sigma_i \sigma_j} \int d\mu(z) e^{-\frac{\beta}{N^{1-\theta}} \sum_{i=1}^{N_1} \sum_{\mu=1}^{N_2} \xi_i^\mu \sigma_i z_\mu} \quad (8)$$

$$= \sum_{\{\sigma\}} e^{\frac{\beta \bar{J}}{N_1} \sum_{ij} \sigma_i \sigma_j} e^{-\frac{\beta^2}{N^{2(1-\theta)}} \sum_{(ij)}^{N_1} \sum_{\mu=1}^{N_2} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j}, \quad (9)$$

where we define $d\mu(z)$ the standard Gaussian measure.

Given the exponential form of the Boltzmannfaktor, as β appears coupled to \mathcal{H} in its exponent, one can see that $P_\beta(\sigma, z; J, \xi)$ defined in equation (7) becomes flat (that is, uniform) for $\beta \rightarrow 0$, thus each configuration has the same probability to happen, regardless of its cost, while for $\beta \rightarrow \infty$ only the global minima of the Hamiltonian are selected, thus players are expected to be perfectly rational.

Crucially, by a direct comparison of the arguments in the exponents of equations (8) and (9), respectively, we also learn that the bipartite interactions between natives and immigrants (that is, those $\propto \sum_{i=1}^{N_1} \sum_{\mu=1}^{N_2} \xi_i^\mu \sigma_i z_\mu$ in equation (8)) are stored in an effective coupling, referred to as \tilde{J}_{ij} , between couples of local decision makers alone (that is, those $\propto \sum_{(ij)}^{N_1} \sum_{\mu=1}^{N_2} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j$ in equation (9)). Such a coupling (referred to as *Hebbian-like*; Agliari and Barra, 2011) reads as:

$$\tilde{J}_{ij} = \frac{\sum_{\mu=1}^{N_2} \xi_i^\mu \xi_j^\mu}{N_1^{2(1-\theta)}}. \quad (10)$$

Therefore, the bipartite model described in equation (5) behaves equivalently to a monopartite model with cooperation among natives only and embedded in a random, diluted structure (see Fig. 1, right panel) (Agliari and Barra, 2011):

$$\begin{aligned} \mathcal{H}(\sigma, z; J, \xi) &= -\frac{\bar{J}}{N_1} \sum_{(ij)}^{N_1} \sigma_i \sigma_j \\ &\quad - \frac{1}{N^{1-\theta}} \sum_{i=1}^{N_1} \sum_{\mu=1}^{N_2} \xi_i^\mu \sigma_i z_\mu \Leftrightarrow \hat{\mathcal{H}}(\sigma; J, \xi) \\ &= -\frac{1}{N_1} \sum_{(ij)}^{N_1} (\bar{J} + \tilde{J}_{ij}) \sigma_i \sigma_j. \end{aligned} \quad (11)$$

The system described by $\hat{\mathcal{H}}(\sigma; J, \xi)$ is embedded in a topology ruled by the parameters \bar{J}, θ, ξ , and it is affected by a noise ruled by β . By varying these parameters, this model is known to undergo a *phase transition*, namely, to switch from a (noisy)

behaviour, where its agents choose individually, to a deterministic behaviour, where coordination among decision makers prevails and collective phenomena (those resulting from their interactions) become dominant. The onset of this global change corresponds to a point in the parameter space called *critical point*: in the present context, the latter can be related to the existence of a critical mass of migrants necessary before their interactions become relevant in affecting international trade relationships.

To identify the existence of a critical point, we need to introduce the concept of ‘order parameter’, namely a simple function able to distinguish between a pure random behaviour of the agents and a coordinated one: it is straightforward to check that, for this model, the order parameter is given by $M(\sigma) = 1/N_1 \sum_{i=1}^{N_1} \sigma_i I_{\sigma_i+1}$, namely the fraction of agents inclined to establish trade relationships abroad. This order parameter is equivalent (upon translation) to $m(\sigma) = 1/N_1 \sum_{i=1}^{N_1} \sigma_i$, namely the simpler arithmetic average over the state of all the agents. Thus, in what follows, we focus on m as this is mathematically easier to deal with.

By applying statistical mechanics procedures on $\mathcal{H}(\sigma; J, \xi)$, we attain the following self-consistent equation for m (see Appendix B of Supplementary Information for a detailed derivation)

$$m = \tanh[m(\beta\bar{J} + \beta^2\xi^2\Gamma)]. \tag{12}$$

This is our key finding: exploiting the proportionality between the number of trading agents (total natives) m and the total number of trading companies Y , namely $m \propto Y$ (see Appendix A of Supplementary Information), it relates the number of trade relationships to the share of migrants in the total population (we recall $\Gamma = \gamma(1-\gamma) \approx \gamma$). The agreement between equation (12) and the empirical data concerning Spanish provinces is investigated in the Data Analysis section.

Interestingly, equation (12) contains also information regarding the critical percentage Γ_c of migrants that must be reached before they can start to influence new trade relationships. To extract such information, we exploit the statistical mechanics know-how of phase transitions: when the factor $(\beta\bar{J} + \beta^2\xi^2\Gamma)$ in the argument of the hyperbolic tangent is smaller than one, the only solution for equation (12) is $m = 0$ (that is, migrants do not boost trading). However, as the factor gets larger than one, non-zero solutions appear and the system, in the physical jargon, experiences a phase transition. To find these solutions, we can expand the hyperbolic tangent in small m as

$$m \sim \beta(\bar{J} + \beta\xi^2\Gamma)m - \frac{\beta^3}{3}(\bar{J} + \beta\xi^2\Gamma)^3 m^3 + \mathcal{O}(m^5), \tag{13}$$

and, excluding the unstable solution ($m = 0$), we get

$$m \sim \sqrt{\frac{3}{\beta^3(\bar{J} + \beta\xi^2\Gamma)^3}[\beta(\bar{J} + \beta\xi^2\Gamma) - 1]} = a\sqrt{\Gamma - \Gamma_c}, \tag{14}$$

where we define

$$a = \frac{3\xi^2}{\beta(\bar{J} + \beta\xi^2\Gamma_c)^3}, \tag{15}$$

$$\Gamma_c = \frac{1 - \beta\bar{J}}{(\beta\xi)^2}. \tag{16}$$

From equation (14) we see that a positive m , corresponding to a boost in the number of trade relationships, emerges as soon as the fraction of migrants within a given province is larger than Γ_c . Thus, the latter represents the threshold in the percentage of migrants in the host territory over which export activity in the

host territory is actually stimulated by migrants. As can be seen in equation (14), Γ_c decreases with \bar{J} and with ξ , where, \bar{J} is related to the overall connection of the network of natives alone, and ξ accounts for the connection between the two parties. This suggests that a possible strategy to stimulate more trade relationships at fixed Γ is to increase ξ , namely the number of cross links between the communities of natives and migrants.

Before concluding it is worth summarizing some fundamental features about the migration-trade link, derived from our statistical mechanics investigation. First of all, the relation between Γ and Y is non-linear, as these observables are related via an hyperbolic tangent (see equation (12)); this can be approximated by a simpler square root relation for small migrants percentages (see equation (14)). Second, there is a critical value for the fraction of migrants, depending on the degree of connection among the two communities, beyond which migrants start to have a positive impact on the number of trade relationships of the host territory. As soon as the level of migrants exceeds the critical one, the growth in the number of trade relationships is expected to be very steep since the function $\sqrt{\Gamma - \Gamma_c}$ has an infinite derivative in Γ_c .² Finally, there is a *saturation effect* for large enough Γ as the hyperbolic tangent is a bounded function that eventually reaches a plateau. This is in agreement with exhaustion levels in bilateral exports that have already been linked with migrant saturation effects as, for instance, in the experimental works discussed in (Egger *et al.*, 2012).

Results

In this section we test our theoretical outcomes using Spanish data on migration and bilateral exports over the period 2000–2012. The empirical analysis is split into three sub-sections. First, we assess whether there is a square-root relationship between the total number of trade relationships abroad and the share of migrants in the population, treating the whole set of immigrants as a unique community. Second, we repeat the analysis at a more disaggregated level, distinguishing several sets of immigrants according to their country of origin. Third, we investigate the scaling between product diversification and migrant density.

Data analysis on aggregate exports and migrants

The first dataset is obtained by merging two publicly available sources: trade data come from ADUANAS-AEAT dataset provided by Ministerio de Economía y Hacienda (www.datacomext.comercio.es), and demographic data come from the Spanish Statistical Office (www.ine.es). The unit of reference is the province (Eurostat NUTS III). The trade data comprise the aggregate value of exports and the number of exporting firms. The number of foreign-born population is used to measure immigration stocks. The period of analysis is 1998–2012, with annual frequency.

We consider the time series for the extensive margin of exports $\{Y_{t,p}\}$ and for the share of immigrants $\{\gamma_{t,p}\}$, along the range of years $t = 1998, \dots, 2012$ and for the 50 provinces $p = 1, \dots, 50$ making up the whole country except for two province-cities in Africa, Ceuta and Melilla. Thus, our time range is made of $N_t = 15$ years and our geographic set is made of $N_p = 50$ provinces.

Since we are using historical data, our first task is to check that at least one of the observables Y and γ is monotonically increasing with respect to the years t . Figure 2 shows that $\gamma(t)$ satisfies this request. Thus, we are allowed to invert $\gamma(t) \rightarrow t(\gamma)$ and look at the evolution of Y as a function of γ , so to obtain $Y(\gamma)$, that must then be suitably binned and averaged (see (Barra and *et al.*, 2014a) for details on this procedure).

This operation is performed for each of the N_p Spanish provinces. We consider different provinces as independent realizations (or, otherwise stated, extractions) of the same system. This means that the number of trade relationships of a

given province depends only on the fraction of immigrants within the province itself. While there is general consensus on this (Herander and Saavedra, 2005; Bahar *et al.*, 2014), the empirical consistency of such hypothesis for the Spanish test case is shown in Artal-Tur *et al.* (2012), where the authors prove that the *proximity* (meant as geographical closeness) is key for the diffusion of the social capital and therefore for the expansion in the number of trade relationships abroad.

Therefore, for each province p , we can measure the percentage of immigrants γ_p and plot the related amount of trade relationships Y_p versus $\Gamma_p \approx \gamma_p$; examples for the three largest provinces are shown in Fig. 3 (left panel). In general, the theoretical predictions (see equation (12)) are in remarkable agreement with the empirical behaviour.³ It is worth noting that, in the best fitting procedure, an extra parameter, referred as a , has to be introduced. This is because statistical mechanics usually works with normalized order parameters (that is, $m \leq 1$), but in order to account for the proper range of variability of Y (that is not bounded in principle), a linear scaling must be applied, that is $Y = am$. We notice that this linear proportionality is perfectly consistent with the model assumptions experimentally verified in Appendix A of Supplementary Information.

We performed extensive fits over all the provinces available according to equation (12), which we report hereafter as

$$Y = a \operatorname{atanh} \left[\left(\frac{(1-b)\Gamma}{\Gamma_c} + b \right) \frac{Y}{a} \right], \tag{17}$$

where we highlighted the critical density $\Gamma_c = (1-b)/(\beta \xi)^2$ and we posed $b = \beta \bar{j}$. Also, as mentioned above, the coefficient a accounts for the intrinsic non-normalized nature of Y , in contrast with m . The best-fit coefficients are collected in Fig. 3 (right panels).

In particular, we notice that $\log(a)$ is roughly uniformly distributed along the range (12,19), suggesting that the extent of exports varies over several orders of magnitude, according to the province considered. On the other hand, Γ_c looks Poissonian-like

Figure 2 | The fraction of migrants in each Spanish province grows monotonically with time.

Note: We collect the historical series for the fraction of migrants γ in each Spanish province over the period 1998–2012 and we consider $\Gamma \equiv \gamma(1-\gamma)$ representing the fraction of potential native-immigrant pairs. For small γ (and this is the case for Spain in the time window considered), $\gamma \approx \Gamma$. In this figure we show Γ versus time t for the three largest provinces (from top to bottom: Madrid, Barcelona, Valencia) and for various countries of origin (depicted in different colours as shown by the legend). Notice that, Γ grows monotonically with t , and this is found also for the rest of Spanish provinces. Through the monotonicity of $\Gamma(t)$ we can invert the function to obtain $t(\Gamma)$. The latter can then be plugged into the dynamics of trade $Y(t)$ to obtain the evolution $Y(\Gamma)$, as reported in Figs. 3 and 4.

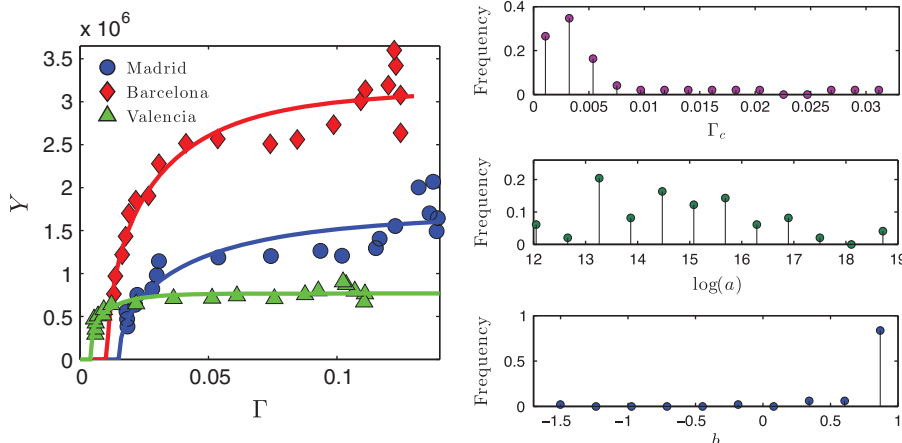
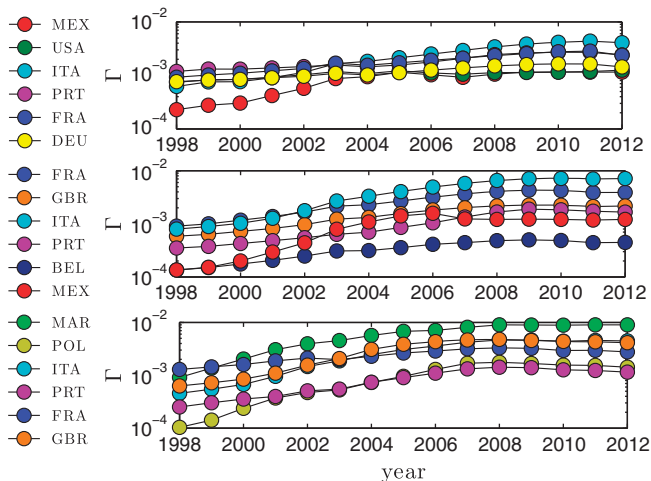


Figure 3 | Fit of experimental data on the total number of exporting firms with our theoretical law and analysis of the fitting coefficients. Left panel: Extent of trade relationships Y versus fraction of immigrants Γ for three different provinces as explained by the legend; we choose the three largest provinces for the sake of readability and for consistency with the analysis of the following sections. However, we checked that analogous plots hold also for the other provinces. In this plot empirical data (symbols) are compared with the theoretical prediction (solid line). More precisely, each data point corresponds to a different year and the solid lines represent the best fit according to equation (17) and the goodness of the fit is $R^2 = 0.94$ (Madrid), $R^2 = 0.97$ (Barcelona), and $R^2 = 0.95$ (Valencia). By repeating the same procedure for all the N_p provinces, we derive for each province p the best-fit coefficients Γ_c , a , and b . The histograms for these coefficients are shown in the right panels. In particular, Γ_c is Poissonian distributed with peak around 0.003 (upper panel), a spans over several orders of magnitude (due to the broad range along which Y_p spans)—and this is why we actually represent the histogram of $\log(a)$ (middle panel), the coefficient b is peaked in agreement with the fact that \bar{j} is a property of the country as a whole and it is quite homogeneous from province to province.

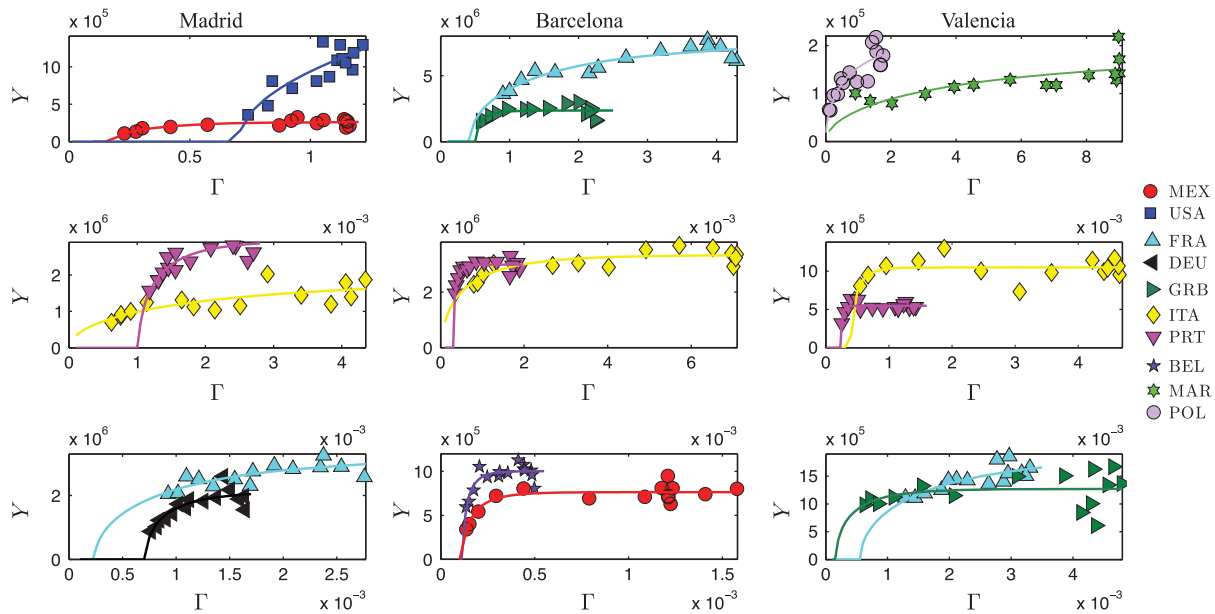


Figure 4 | Fit of experimental data on bilateral trade relationships with our theoretical law.

Note: We select the three largest provinces out of the available N_p and for each of them we consider the data for trade relationships $Y_{p,f}$ performed between the province p and the foreign country f . The data for trade relationships $Y_{p,f}$ are then analysed versus the fraction of migrants $\Gamma_{p,f}$ residing in p and hailing from f and we perform the best fits according to the theoretical law in equation (17). Here we show the comparison between the empirical data (symbols) and best fits (solid lines) for Madrid (left panels), Barcelona (middle panels), Valencia (left panels). Different countries are depicted in different colours and symbols as specified by the legend. The foreign countries considered are those where Γ spans over the largest interval in such a way that fits can be more accurate.

distributed and is peaked around 0.003, suggesting that when the total amount of immigrants is less than 0.3% of the whole population inside the province, their presence is ineffective as facilitator of trade with their country of origin.

Data analysis on bilateral trade relationships (firms) and migrants

In order to get a clearer picture, and to evaluate more precisely the country-dependent critical threshold Γ_c , we consider migrants by nationality and exporting firms from the host province to the country of origin of migrants. As such, we analyse the bilateral trade relationships $Y_{p,f}$ performed between any province p and any foreign country f as a function of the related fraction of immigrants $\Gamma_{p,f}$. Of course, results are expected to be much noisier, as we are dealing with considerable smaller datasets and the intrinsic fluctuations are only partially smoothed by the central limit theorem. Nonetheless, it is worth checking whether the previous results are still valid at this finer level of resolution and inferring the country-dependent critical masses. We focus on the three major Spanish cities, namely Madrid, Barcelona and Valencia (Fig. 4), and on the foreign countries for which the size of immigrant communities are larger and span a wide interval in the time window considered, in order to get more accurate and reliable fits according to equation (17). In general, the agreement between empirical data and theoretical expectations is very good (the coefficient of determination R^2 is close to 1), further corroborating the picture provided by our model.

Moreover, the fitting procedure allows an estimate for Γ_c for each pair (p,f) to be obtained. These values are shown in Fig. 5 restricting to the most solid estimates. Remarkably, for bilateral trades Γ_c follows a distribution peaked around $\bar{\Gamma}_c \approx 10^{-5}$, that is consistent with the previous value $\sim 3 \cdot 10^{-3}$ found for global trades as migrants come from $O(10^2)$ different countries.

Data analysis on export product diversification

Having shown that the amount of trade relationships is positively influenced by migration, in this section we aim to investigate whether the diversification of exports is enhanced by migration too.

In order to keep the analysis as smooth as possible, we follow the simplest possible route (leaving possible improvements based on, for instance, complexity measures (Tacchella, 2002; Caldarelli and *et al.*, 2012; Cristelli and *et al.*, 2013) as options for future additional work): the export portfolio of a province is composed of products and destinations. That is, a province can export several products to a single destination or export the same product to several destinations. Thus, the basic unit in the export portfolio is a product–destination pair. We define K as the total number of product–destination pairs in the export portfolio of a province. Products are defined using the Combined Harmonised System (HS) provided by UN COMTRADE database (we exclude *special* product categories, HS98 and HS99). Destinations are defined as countries with more than 1 million population in 2010. There are 4507 products and 154 countries, so the total number of product–destination K pairs is 694078.

To account for the distribution of export sales across product–destination pairs, we use the export share of each product–destination pair in total export value so to capture the relative importance of each pair for exports. The Herfindahl index N_H (Beck and *et al.*, 2001) is a simple calculation of concentration of exports that uses such export shares: the larger the number N_H , the more concentrated (less diversified) the export portfolio of the province is. Therefore, if migrants do really contribute to diversification of exports, we should evidence a negative correlation between N_H and Γ .

More precisely, the N_H index is calculated as

$$N_H = \sum_{i=1}^K \left(\frac{x_i}{\bar{X}} \right)^2, \tag{18}$$

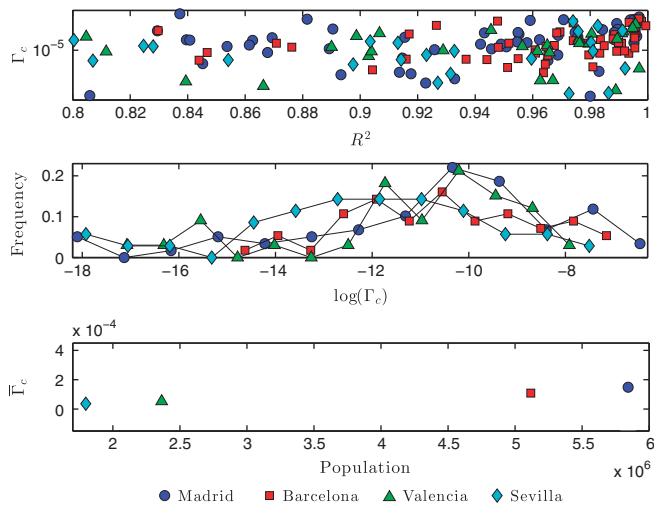


Figure 5 | Estimates for the critical density of immigrants obtained by fitting data on number of exporting firms: their distribution and their average values suggest robust behaviours among provinces.

Note: We select the four largest provinces (Madrid, Barcelona, Valencia and Sevilla) out of the available N_p , and for each of them we consider the data for trades $Y_{p,f}$ performed between the province p and the foreign country f . The data for trade relationships $Y_{p,f}$ are then analysed versus the fraction of migrants $\Gamma_{p,f}$ residing in p and hailing from f and we perform the best fits according to the theoretical law in equation (17). In this way we obtain estimates for the critical density of immigrants Γ_c and the related R^2 , which are plotted in the upper panel. Different symbols and colours refer to different provinces as explained by the legend. Each data point represents a different pair (p,f) . The most reliable fits (that is, R^2 close to 1) suggest that Γ_c is distributed around 10^{-5} , and this holds for all the four provinces analysed. Focusing on estimates corresponding to $R^2 > 0.85$, we build the histogram of Γ_c (shown in the middle panel) and calculate its arithmetic average to get $\bar{\Gamma}_c$, which is plotted in the lower panel as a function of the population of the related province.

where x_i is the value of export in product–destination i and X is the total value of exports. One can further normalize N_H to get an index n_H whose values lie between 0 and 1. Results are shown in Fig. 6, where the negative correlation between n_H and the percentage of migrants Γ within the province is evidenced. More precisely, the two observables scale as $n_H \sim \Gamma^{-1/\delta}$, with $\delta \sim 5$ (see the caption of Fig. 6 for further details).

Thus, at least for small percentages of migrants, that is $\Gamma \approx \gamma$, there is a positive correlation between export portfolio diversification and the density of migrants in a particular province. We can therefore conclude that migrants, in this context, also contribute to diversify the export product portfolio of the host province.

Conclusions and outlooks

Tools from theoretical physics have recently been applied to economics to identify possible mechanisms and relations underlying empirical evidence. For instance, a remarkable result recently obtained is that the composition of exports from the most developed countries tends to be highly diversified rather than concentrated in a few products (Tacchella, 2002; Hidalgo *et al.*, 2007; Hidalgo and Hausmann, 2009). Here, with an analogous intent, we adopt a statistical-mechanics approach and we investigate whether the presence of large number of migrants in a developed country (Spain) contributes to explain the growth in the number of exporting firms and, by extension, to diversify

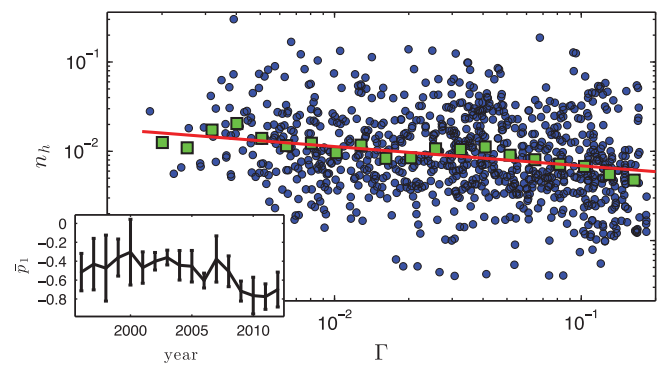


Figure 6 | The Herfindahl index decreases monotonically with the fraction of migrants.

Note: In the main plot bullets represent the value of the normalized diversification index n_H for different province and different years as a function of the fraction of migrants Γ present in the same province in the same year. Green squares represent binned data and the solid red line is the related best fit ($R^2 \approx 0.93$). This is a linear curve (in log-log scale) $y = p_1 x + p_2$, with $p_1 = -0.21 \pm 0.01$ and $p_2 = -5.47 \pm 0.01$. The fitting has also been performed for data of n_H pertaining to any single province and any single year, hence obtaining $p_1(y,p)$. These values have been averaged over the provinces to get $\bar{p}_1(y) = \sum_{p=1}^{N_p} p_1(y,p) / N_p$, which is shown in the inset (the line is a guide for the eye). This plot shows that the monotonicity of n_H with respect to Γ (that is, $\bar{p}_1 < 0$) is robust with respect to the year; the same holds even when we average over the year, namely it is robust with respect to the province.

the product portfolio of exports. Our framework, based on statistical mechanics, allows us to explain how an increasing number of interactions between natives and migrants is positively related to the number of trade relationships between the host country and the country of origin of the migrants.

From an economic perspective, our theoretical results suggest that the scaling between the extensive margins of trade (that is, the number of exporting firms) versus the percentage of migrants living in the host country is non-linear. In particular, for small percentages of migrants in the host country the relation can be well approximated well by a square root, while for larger percentages saturation effects are expected and a plateau can eventually be reached with the scaling of an hyperbolic tangent. Furthermore, we examine the possible relationship between the share of migrants in the total population and the extent of diversification of the portfolio of exported goods. We find evidence of a strong positive correlation.

Unlike previous research (Gould, 1994; Egger *et al.*, 2012; Serrano-Domingo and Requena-Silvente, 2013), we highlight the existence of a minimum threshold of migrants such that, when the percentage of migrants in the host territory is relatively small, migrants do not have any impact on trade. We also find that the threshold is sensitive to the nationality of the migrants, suggesting that cultural differences between natives and migrants may affect the number of migrants needed to generate a positive impact on trade.

Remarkably, the model underlying these findings allows us to relate the importance of the minimum threshold to social features such as the degree of interaction between immigrants and natives. By increasing the number of interactions, the threshold is lowered with a consequent improvements in the number of trade relationships.

Our work also suggests that what matters is the share of migrants as percentage of the native population rather than the total number of migrants. Moreover, our approach can be used

to examine other related issues such as the impact of formal or informal firm networks on trade.

To conclude, we derive a simple model based on agents cooperation that allows us to gain insights into the role of international migration in the evolution and diversification of international trade relationships. We can identify a number of improvements and refinements for future related works. One aspect worthy of consideration is migrants' heterogeneity: even in a community of people from the same country, the social and cultural composition can be very different, from refugees, economic workers and unskilled workers to professionals and entrepreneurs. Furthermore, the adoption of more complex measures of products' diversification to better address the outlined (indirect) influence of migrants on the global market would be advantageous. Finally, turning to the experimental protocol, other hosting countries should be considered beyond Spain to give more ground to the theory as a whole.

Notes

- 1 This assumption is only the simplest working assumption but the scenario we outline in this work is robust under several other choices, that is, $z_{\mu} \in \{+1, -1\}$ as shown in Barra *et al.* (2011); Barra and *et al.* (2014b).
- 2 $\Gamma < \Gamma_c$ is technically possible. However, in that case, the cooperative couplings among agents are ineffective and the solution of the statistical mechanics model is no longer the square-root but a constant. This means that, in this regime, a change in the fraction of migrants is not expected to affect the amount of new trade relationships. In other words, Γ_c is a singular point for the solution $Y(\Gamma)$ and the square-root is solution only beyond Γ_c .
- 3 Note that the error implied by replacing γ with Γ , estimated to be $O(1\%)$, is smaller than intrinsic fluctuations.

References

- Agliari E, Altavilla M, Barra A, Dello Schiavo L and Katz E (2015) Notes on stochastic (bio)-logical gates: Computing with allosteric cooperativity. *Scientific Reports*; 5: 9415.
- Agliari E and Barra A (2011) A Hebbian approach to complex network generation. *Europhysics Letters*; 94 (1): 10002.
- Artal-Tur A, Pallardo-Lopes VJ and Requena-Silvente F (2012) The trade-enhancing effect of immigration networks: New evidence on the role of geographics proximity. *Economics Letters*; 116 (3): 554–557.
- Bahar D, Hausmann R and Hidalgo CA (2014) Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion? *Journal of International Economics*; 92 (1): 111–123.
- Bahar D and Rapoport H (2014) *Migration, Knowledge Diffusion and the Comparative Advantage of Nations*. Tech. rep., Working Paper. Harvard University; Cambridge, MA.
- Barra A, Contucci P, Sandell R and Vernia C (2014a) An analysis of a large dataset on immigration integration in Spain: The statistical mechanics perspective of Social Action. *Science Reports*; 4: 4174.
- Barra A, Galluzzi A, Guerra F, Pizzoferrato A and Tantari D (2014b) Mean field bipartite spin models treated with mechanical techniques. *European Physical Journal B*; 87 (3): 1–13.
- Barra A, Genovese G and Guerra F (2011) Equilibrium statistical mechanics of bipartite spin systems. *Journal of Physics A: Mathematical and Theoretical*; 44 (24): 245002.
- Bastos P and Silva J (2012) Networks, firms and trade. *Journal of International Economics*; 87 (2): 352–364.
- Beck T, Clarke G, Groff A, Keefer P and Walsh P (2001) New tools in comparative political economy: The database of political institutions. *The World Bank Economic Review*; 15 (1): 165–176.
- Bouchaud J-P and Potters M (2009) *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge Press; Cambridge, UK.
- Brock W and Durlauf S (2001) Discrete choice with social interactions. *Review of Economics and Statistics*; 68 (2): 235–260.
- Buchanan M (2010) *Networks in Cell Biology*. Cambridge Press; Cambridge, UK.
- Caldarelli G, Cristelli M, Gabrielli A, Pietronero L, Scala A and Tacchella A (2012) A network analysis of countries export flows: Building blocks of the economy. *PLoS One*; 7 (10): e47278.
- Cristelli M, Gabrielli A, Tacchella A, Caldarelli G and Pietronero L (2013) Measuring the intangibles: A metrics for the economic complexity of countries and products. *PLoS one*; 8 (8): e70726.
- Durlauf SN (1999) How can statistical mechanics contribute to social science? *Proceedings of National Academy of Sciences*; 96 (19): 10582–10584.
- Egger PH, Von Ehrlich M and Nelson DR (2012) Migration and trade. *The World Economy*; 35 (2): 216–241.
- Genc M, Gheasi P, Nijkamp and Poot J (2012) *The Impact of Immigration on International Trade: A Meta-Analysis Meta-Analysis*; Migration Impact Assessment: New Horizons. Edward Elgar Edition.
- Gould D (1994) Immigration links to the home country: Empirical implications for us bilateral trade flows. *Review of Economics and Statistics*; 76: 302–316.
- Head K and Rises J (1998) Immigration and trade creation: Econometric evidence from Canada. *Canadian Journal of Economics*; 31 (1): 47–62.
- Herander M and Saavedra L (2005) Exports and the structure of immigrant-based networks: The role of geographic proximity. *Review of Economics and Statistics*; 87 (2): 323–335.
- Hidalgo CA and Hausmann R (2009) The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*; 106 (26): 10570–10575.
- Hidalgo CA, Klinger B, Barabasi AL and Hausmann R (2007) The product space conditions the development of nations. *Science*; 317 (5837): 482–487.
- Hunt J and Gauthier-Loiselle M (2009) *How Much does Immigration Boost Innovation?*. Institute for the Study of Labor, Bonn.
- Ozgen C, Nijkamp P and Poot J (2011) *Immigration and Innovation in European Regions*. IZA Disc.
- Partridge J and Furtan H (2008) Increasing Canada's international competitiveness: Is there a link between skilled immigrants and innovation? In American Agricultural Economics Meeting.
- Peri G and Requena-Silvente F (2010) The Trade Creation Effect of Immigrants: evidence from the remarkable case of Spain. *Canadian Journal of Economics*; 43 (4): 1433–1459.
- Rashidi S and Pyka A (2013) Migration and innovation. FZID Discussion Papers. 77–2013.
- Serrano-Domingo G and Requena-Silvente F (2013) Re-examining the migration-trade link using province data: An application of the generalized propensity score. *Economic Modelling*; 32: 247–261.
- Tacchella A (2002) A new metrics for countries fitness and products complexity. *Science Reports*; 2: 723.

Data availability

The datasets analysed during the current study are available in the Spanish Ministry of Economy and Competitiveness repository [datacomext.comercio.es] and in the Spanish Statistical Office repository [www.ine.es].

Acknowledgements

The authors are grateful for partial financial support to Sapienza Università di Roma via the grants “Avvio alla ricerca (Tantari2014)” and “Avvio alla ricerca (Galluzzi2015)” and to GNFM-INDAM via the grant “Progetto Giovani (Agliari2014)”. Adriano Barra gratefully acknowledges the financial support of Laboratories for Information Food & Energy (project Match-Innova). Francisco Requena-Silvente gratefully acknowledges the financial support of the Spanish Ministry of Economy and Competitiveness (project ECO2015-68057) and the Valencian Regional Government (PROMETEOII/2014/053).

Additional information

Supplementary Information: accompanies this paper at <http://www.palgrave-journals.com/palcomms>

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available at http://www.palgrave-journals.com/pal/authors/rights_and_permissions.html

How to cite this article: Barra A *et al.* (2016) Assessing the role of migration as trade-facilitator using the statistical mechanics of cooperative systems. *Palgrave Communications*. 2:16021 doi: 10.1057/palcomms.2016.21.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>