

A network-based measure
of the socio-economic roots of the migration flows

Roy Cerqueti^{b*}, Gian Paolo Clemente[‡], Rosanna Grassi[‡]

^b University of Macerata, Department of Economics and Law.

Via Crescimbeni 20, 62100, Macerata, Italy.

Tel.: +39 0733 2583246; fax: +39 0733 2583205. Email: roy.cerqueti@unimc.it

[‡] Catholic University of Milan, Department of Mathematics, Finance and Econometrics

Email: gianpaolo.clemente@unicatt.it

[‡] University of Milano-Bicocca, Department of Statistics and Quantitative Methods

Email: rosanna.grassi@unimib.it

December 2, 2017

*Corresponding author.

A network-based measure of the socio-economic roots of the migration flows

March 1, 2018

Abstract

This paper provides a unified view for defining a measure of the reasons behind migration flows whose nature is of social and economic type. To this aim, worldwide migration flows are here presented in the context of complex network and a related socio-economic indicator is conceptualized. The ingredients of the indicator also include the economic strengths of the countries and how they behave in terms of community structure, where “community” has to be intended in the sense of how countries interact in terms of immigration and emigration. Empirical analyses on a wide set of real data validate the theoretical framework, hence giving a paramount quantitative view of the roots of the worldwide migration flows.

Keywords: Migration flows, socio-economic indicator, networks, clustering coefficient.

JEL Classification: F22, O15, C02, Z13

1 Introduction

The movement of individuals from the origin countries to new ones represents – and has represented in the past – one of the most pervasive demographic phenomenon at international level. A number of reasons can be found behind migrations. Most of them are of socio-economic nature. Often migrants leave their countries to escape from ethnic tensions and conflicts, civil wars or poverty and absence of a satisfactorily labour market.

This paper aims at providing a new perspective for exploring the socio-economic reasons behind the migration flows. Specifically, we propose a novel socio-economic indicator with the purpose of defining a synthetic measure of the incoming and outgoing flows between countries in terms of: *(i)* the economic characteristics of the origin and destination countries; *(ii)* the interconnectedness of the countries in the overall context of migration flows; *(iii)* the entity of the migration flows from the origin country to the destination one.

The issues listed in points (i) – (iii) are reasonably assumed to broadly include the entire set of motivations for explaining migration flows. As a very relevant example, think at the motivations for migrating related to the social connections with people living in a given destination country. In this case, the roots of migration can be found in the economic strength of the hosting country – which indeed gave the possibility to migrants’ connected people to take a position in the socio-economic context – and in how easily the final destination can be reached through other transit countries.

Several papers have dealt with the analysis of the roots of migrations. Migration is a global phenomenon, which is the effect of many different factors. For instance, sociological and economic motivations have driven (and they still drive) the migrations (see [2], [13], [15]). A crucial hypothesis is that the migration flow comes from poor to rich countries. This fact results from disequilibrium in the global distribution of income and wealth, encouraging the movement towards richer countries in the attempt to improve the conditions of life (see [4], [19]). Other theories about migrations are related to spatial distance between countries, as well as the mass of the migrants. A well-explored literature focuses on the “gravity-model”, in which the migration is directly related to the population of the origin and of the destination country, and inversely related to the distance between them ([1], [3], [12]). For instance, in [18] the mass of migrant flows reduces with increasing the spatial distance, due to travel time, monetary costs but also cultural differences, facilitating in this way the back-and-forth migration.

An interesting perspective in the migration roots context is offered by [23]. Such a point of view is rather different from ours, since the authors deal with a concept of migration related to the changing of opinion, while we consider the socio-economic contextualization of the migrants. The physical nature of migration is well explained in [24], where the motion of substance (or of human) is modelled over the nodes of a network. The paper is of particular interest, because it assumes that the nodes can also exhibit an attracting behaviour. In this respect, our paper carries out arguments which are similar to those of [24], where countries are able to polarize migrants. Such a property is here captured by the proposed indicator.

Global migration flows between countries naturally can be well described in terms of complex networks (see [9], [22]). However, a network analysis of internal migration patterns has not been explored enough. A recent contribute to this aspect is given by [11], where the network structure of US migration on a yearly basis is analysed, identifying communities that arise, as well as their trend.

One of the most recent and prominent contributes is due to [25], where the network of global migration is considered and empirically explored over the period 1990-2013. This paper uses methods of temporal and cross-sectional exponential random graph for identifying the determinants of migrations and assessing the presence of noticeable clusters in migration flows.

We feel close to [25]. Indeed, the quoted paper advances a methodological proposal for simulta-

neously combining different aspects of geographic, demographic, economic, religious, linguistic, and historical nature – all related to migration. Also in the present paper a unified framework is carried out and determinants of migration are evaluated jointly. However, to the best of our knowledge, this is the first paper proposing a synthesis of such elements through a novel indicator, which is then able to capture in a unified framework the universe of socio-economic motivations for explaining the immigration phenomenon.

Our approach is based on theory of complex networks, which seems to be particularly appropriate for our purposes. Countries are viewed as nodes of a weighted network, whose directed weighted arcs model both the presence and the entity of the migration flows between a (origin) country to another (destination) one. In so doing, we can imagine that any node has a in-weight and out-weight associated to an adjacent node, to capture both immigration and emigration phenomena.

The interconnected structure is represented by means of the weighted clustering coefficient of the nodes of the network, with a special reference to immigration for destination countries and emigration for origin ones. In particular, we here adopt the definitions proposed by [8] and [20]. Indeed, the interconnection plays a fundamental role in determining the migration patterns. A migration from a country to another could be the result not only of a direct movement but also the effect of the transition through a third country. In other words, the presence of weighted triangles in this network matters.

The economic strength of the countries are identified through their GDPs.

All the considered quantities have been opportunely normalized, in order to derive a relative indicator and be able to make comparisons. Moreover, we propose a global analysis of the immigration and emigration flows related to single countries and sets of countries by means of a local indicator related to the connections between couples of countries. In so doing, we are able to present and discuss a paramount view of reasoning behind the migration flows.

In dealing with the definition of a social indicator in a complex networks environment, we are in line with a wide strand of the literature. The mainstream literature in social network analysis focuses on local network measures, as vertex centralities, in order to assess the power or influence of a node/link in a whole system, as well as other global network parameters, as assortativity, average path length or average clustering coefficient. For instance, in [16] the author provides a composite index that measures the bundles of institutions channelling the positive effect on economic prosperity.

The rest of the paper is organized as follows. Section 2 describes the network of migration flows. This Section also provides the basic notation used throughout the paper. Section 3 contains the construction of the socio-economic indicator, which is the core of the study. Section 4 is devoted to the empirical experiments and to the discussion of the obtained results. Last Section offers some conclusive remarks.

2 The network of migration flows

First, we briefly remind the terminology on networks that will be used in the paper. For a reference about graph and network theory, the reader can refer to [14], [21].

A graph $G = (V, E)$ is a pair of sets V and E , where V is the set of N vertices (or nodes) and E is the set of M pairs (links) of vertices of V ; if (i, j) or $(j, i) \in E$, then vertices i and j are adjacent. A link $(i, i) \in E$ is said a loop. A weight $w_{ij} > 0$ can be associated with each link (i, j) so that a weighted graph is obtained. The adjacency relationships between vertices of G are described by a non-negative, real N -square matrix \mathbf{A} (the adjacency matrix). The degree of a node i , d_i , is the number of links incident with it. In case of weighted graphs, the so-called strength of the node can be also computed. Formally, it is simply the sum of the weights over all links incident with it, $s_i = \sum_{(i,j) \in E} w_{ij}$.

A graph is directed if it is characterized by directed links (or arcs) between nodes. In this case $(i, j) \in E$ does not imply $(j, i) \in E$. The in-degree d_i^{in} of a node i is the number of links pointing towards i , the out-degree d_i^{out} of i is accordingly defined as the number of edges originating from i . The total degree of a node is the sum of in-degree and out-degree, $d_i^{tot} = d_i^{in} + d_i^{out}$. In-strength and out-strength are defined analogously as:

$$s_i^{in} = \sum_{(j,i) \in E} w_{ji}, \quad (1)$$

$$s_i^{out} = \sum_{(i,j) \in E} w_{ij}. \quad (2)$$

The total strength is the sum of in-and out-strength.

We model the migrant's flow through a network, where nodes are countries and links identify the migrants moving from a country to another. Since a migration between two countries i and j can occur everywhere and in both directions, the network G is complete and all nodes are connected through bilateral directed links. Given the nature of the migration phenomenon, in the network loops are not present, therefore the total number of arcs is $M = 2N(N - 1)$.

The analysis will be performed in one period, taking a "picture" of one reference year. For the analysis, we will consider the entire set of countries, according to the United Nations (UN) classification¹, extended in order to take into account also of dependencies², areas of Special Sovereignty³

¹The standard classification of UN comprises 193 member states and 2 countries that are non-member observes states (Holy See and State of Palestine). Therefore, this classification considers the 195 independent countries in the world.

²A "dependent territory" is a territory that is not fully independent or sovereign. Dependent territories belong to a sovereign state, and are dependent on that state to some degree. There are varying degrees and forms of such a dependence. Several islands and territories depending on Australia, Denmark, France, Netherland, United Kingdom, United States of America, New Zealand are here considered separately.

³Hong Kong and Macau are examples of special sovereignty.

and state in free associations⁴. In particular, we use for our analysis 231 distinct countries.

Links are opportunely weighted by relative amounts of migration flows. Specifically, we denote by $MIGR_{ij} \in \mathbb{N}$ the number of people migrating from the origin country i to a destination country j , for each $i, j \in V$. Observe that, in general $MIGR_{ij} \neq MIGR_{ji}$.

Given $i, j \in V$, the weight of the link (i, j) is defined as:

$$w_{ij} = \frac{MIGR_{ij}}{\sum_{h,k \in V; h \neq k} MIGR_{hk}}. \quad (3)$$

Note that, being the network complete, the absence of a migration flow from i to j is not associated to the missing link (i, j) but, rather, to an existing link with $w_{ij} = 0$. This allows us to catch in our analysis also the effect of relations between countries where no migrants' flows are observed.

We also assign to each node an attribute in order to quantify a specific characteristic of the country. More precisely, for each $i \in V$, we denote by $\alpha_i \in [0, 1]$ the relative weight of the node i in the overall system. In our context, α_i represents the relative richness of country i , which is measured through the ratio between the GDP (Gross Domestic Product) of the country i and the total GDP of all the countries populating the network. Formally:

$$\alpha_i = \frac{GDP_i}{\sum_{k \in V} GDP_k}, \quad (4)$$

where GDP_i is the GDP of country i at the reference year. It's evident that $\sum_{k \in V} \alpha_k = 1$, then α_i represents the proportion of richness of the country i over the totality of the countries.

3 Construction of the socio-economic indicator

In this Section we propose a new socio-economic indicator of the flow of migrants from a country i to a country j . This indicator aims at providing a measure of the economic and social reasons behind the migration flows.

Indeed, there is a wide evidence that migrants generally move from poor countries to rich ones, preferring locations highly connected with the others to feel free to move again in the future from the hosting country to a new one. A socio-economic indicator related to the migrant flows should reflect both aspects, modelling both the macroeconomic status of the considered countries and the neighbourhood structure associated to each country.

To this end, we include the GDP of the country/node i (i.e. the parameter α_i) as well as the local clustering coefficient of i in the definition. Furthermore, the definition of local clustering coefficient will take into account of both the bidirectional links and the weights.

⁴Examples are self-governing state in free association, as Niue and Cook Island that are in free association with New Zealand

Being the distribution of weights really skewed, because of significant differences in terms of migrant stock between countries, the well-known clustering coefficients proposed in [8] may be useless to depict heterogeneity of countries providing values close to zero. Furthermore, as shown in [20], the coefficient proposed in [8] may fail when weighted complete graphs are considered⁵. Hence, we follow the idea proposed in [20] to consider weights in a binary clustering and we also adapt the coefficient to our framework in order to catch the effects of links with $w_{i,j} = 0$. In particular, as in [20], we modify the adjacency matrix fixing a threshold $s \in [0, \max_{i,j}(w_{i,j})]$ and defining \mathbf{A}_s , whose elements are:

$$a_{ij}^s = \begin{cases} 1 & \text{if } w_{ij} \geq s \\ 0 & \text{otherwise} \end{cases}. \quad (5)$$

Hence, \mathbf{A}_s is the adjacency matrix describing the existing links in the network having weight w_{ij} at or above the threshold s .

The idea is to capture the mean cluster prevalence of the network looking at a zoom-out level where only the strongest edges (i.e. edges with a weight greater than a given threshold) are visible. To this end, we compute the following local clustering coefficient for directed networks:

$$C_i(\mathbf{A}_s) = \frac{\frac{1}{2}[\mathbf{A}_s + \mathbf{A}_s^T]_{ii}^3}{(2N-2)(2N-3) - 2(N-1)} = \frac{\frac{1}{2}[\mathbf{A}_s + \mathbf{A}_s^T]_{ii}^3}{2(N-1)(2N-4)}. \quad (6)$$

where $[\mathbf{A}_s + \mathbf{A}_s^T]_{ii}^3$ is the i -th element of the main diagonal of $[\mathbf{A}_s + \mathbf{A}_s^T]^3$, being this matrix the 3-power of the matrix $[\mathbf{A}_s + \mathbf{A}_s^T]$.

The coefficient in the previous formula has been proposed in [8] for binary and directed case, and it is conveniently adjusted to our framework. In particular, the denominator is properly arranged in order to also include in the analysis those links having weight equal to zero. In other words, in this way we take into accounts that migrants from an origin country can hypothetically move to every other destination country.

Therefore, the total degree, that is considered at the denominator of the original formula in [8] is here replaced by the total degree of a complete graph (i.e. $2(N-1)$). As in [8], bilateral arcs (whose number is $2(N-1)$ in a complete graph) have to be removed by the formula, as they represent “false” triangles, being formed by i and by a pair of directed arcs pointing to the same node, e.g., (i, j) and (j, i) .

Since for a country the migrants’ flow is incoming as well as outgoing, we are able to take into account different patterns among neighbouring countries, reflected by specific clustering coefficients ($C_i^{in}(\mathbf{A}_s)$ and $C_i^{out}(\mathbf{A}_s)$).

Also in this particular case we refer to in- and out-clustering coefficients proposed in [8], adapting them to our specific context:

⁵The alternative clustering coefficient proposed in [5] cannot be used for our purpose being equal to one for complete graphs.

$$C_i^{in}(\mathbf{A}_s) = \frac{[\mathbf{A}_s^T \mathbf{A}_s^2]_{ii}}{(N-1)(N-2)}. \quad (7)$$

$$C_i^{out}(\mathbf{A}_s) = \frac{[\mathbf{A}_s^2 \mathbf{A}_s^T]_{ii}}{(N-1)(N-2)}. \quad (8)$$

In-clustering considers triangles such that there are two arcs incoming into i ($j \rightarrow i, k \rightarrow i, j \rightarrow k \vee k \rightarrow j$) (i.e. “in” triangles). Hence, it could give an idea of the movement of incoming migrants, also passing through neighbouring countries. Analogously, out-clustering focuses on triangles such that there are two arcs coming out of i ($i \rightarrow j, i \rightarrow k, j \rightarrow k \vee k \rightarrow j$) (i.e. “out” triangles), giving an idea of the emigration towards neighbouring countries.

The process is repeated varying the threshold s and computing the clustering coefficient C_i for a node i as the average of $C_i(\mathbf{A}_s)$ overall $s \in [0, \max_{i,j}(w_{i,j})]$:

$$C_i = \int_0^{\max_{i,j}(w_{i,j})} C_i(\mathbf{A}_s) ds \quad (9)$$

In a similar way, we can also obtain C_i^{in} and C_i^{out} by averaging overall values of $C_i^{in}(\mathbf{A}_s)$ and $C_i^{out}(\mathbf{A}_s)$ respectively.

Notice that both local in- and out-clustering coefficients are in the range $[0, 1]$, for each $i \in V$. As already mentioned above, such a quantity provides information on the sense of neighbourhood structure associated to a country i in terms of migration flows: a high value of C_i^{in} (or C_i^{out}) is associated to a country i with a large quantity of incoming (or outgoing) flows. We study them in a separate way in order to test possible different patterns in terms of incoming and outgoing flows.

Now, we provide the definition of the indicator at the level of a single directed migration flow between two countries, and then we define how to combine them at a country, macro area or global level.

We introduce the *local* economic indicator associated to the migration flows from country i to country j – and denote it as I_{ij} – as follows:

$$I_{ij} = \begin{cases} \alpha_i(1 + C_i^{out}) - \alpha_j(1 + C_j^{in}) & \text{if } w_{ij} \geq 0 \\ 0 & \text{if } w_{ij} = 0 \end{cases} \quad (10)$$

Definition (10) offers a measure of the economic and social ground of the migration flows from country i to country j . In particular, it takes into account the difference between the economic strengths – measured through the α 's – of the involved countries, which is remarkably amplified if the considered country forms a strict community with the adjacent countries (high value of the clustering coefficient). In particular, the clustering coefficient of a country i acts in formula (10) as a uniperiodal rate of return for the economic parameter α_i . In this respect, the term $\alpha_i(1 + C_i^{out})$ (or $\alpha_j(1 + C_j^{in})$) can be viewed as an indicator which provides a synthesis of the economic strength of i and of the level of the “out” (“in”) community around it.

It is worth listing the properties of the quantity defined in (10).

- $I_{ij} \in [-2, 2]$, for each $i, j \in V$. This is a very important property, allowing to state a consistent comparison among migration flows in terms of the economic wealth of the involved countries.
- $I_{ij} > 0$ ($I_{ij} < 0$) when $\alpha_i(1 + C_i^{out}) > \alpha_j(1 + C_j^{in})$ ($\alpha_i(1 + C_i^{out}) < \alpha_j(1 + C_j^{in})$). This means that the origin country can be briefly seen as richer and more (poorer and less) interconnected – in terms of migration flows – with the surrounding environment than the destination one. The situation $I_{ij} > 0$ can be characterized with more detailed remarks. In this case, one has that $\alpha_i > \alpha_j \cdot \frac{1+C_j^{in}}{1+C_i^{out}}$ or, alternatively, $C_i^{out} > (1 + C_j^{in}) \cdot \frac{\alpha_j}{\alpha_i} - 1$. Thus, the socio-economic indicator I_{ij} is positive when the origin country i has economic strength or community level above some thresholds also depending on the parameters related to the destination country. The same arguments can be opportunely reverted for explaining the situation of $I_{ij} < 0$. One can expect that the case $I_{ij} > 0$ will be the exception rather than the rule. Empirical data will support this intuition (see the next Section).
- $I_{ij} = -2$ is associated to the case in which $\alpha_i = 0, \alpha_j = 1, C_j^{in} = 1$ and $w_{ij} = 1$. Thus, this extreme case captures the situation in which the total amount of migrants flows from the poorest country – even, with null GDP – to the richest one, and the richest country has the largest possible sense of community with its neighborhood (i.e., the maximal level of clustering coefficient). The community level of the origin country does not play any role in this specific situation.
- $I_{ij} = 2$ is associated to the case in which $\alpha_i = 1, \alpha_j = 0, C_i^{out} = 1$ and $w_{ij} = 1$. The total amount of migrants flows from the richest country to the poorest one – also in this case the poorest country is the one with null GDP and the richest country has the maximal clustering coefficient. Symmetrically with the previous case, the community level of the destination country does not matter.
- $I_{ij} = 0$ when $\alpha_i(1 + C_i) = \alpha_j(1 + C_j)$ and/or $w_{ij} = 0$. This neutral case is then related to migration flow between countries sharing an identical quantitative synthesis of economic and community levels and/or absence of migrants from i to j . Thus, this situation of fairness might depend on migration flows between identical countries – to be intended under a joint economic and community perspective.

It is clear that there is no any country with null GDP . Thus, the bounds ± 2 are extremal cases and cannot be achieved in practical analysis.

Moving from the indicators computed for each couple of countries, we can provide a measure of the migration network with the aim of analyzing the socio-economic reasons of such a phenomenon at either country, macro area or global level.

We define the *country* (“in” and “out”) socio-economic indicators of the migration flows as a weighted mean of the local socio-economic indicators, as follows:

$$I_i^{in} = -\frac{\sum_{j \neq i \in V} I_{ji} w_{ji}}{\sum_{j \neq i \in V} w_{ji}}, \quad (11)$$

$$I_i^{out} = \frac{\sum_{j \neq i \in V} I_{ij} w_{ij}}{\sum_{j \neq i \in V} w_{ij}}. \quad (12)$$

It is worth pointing out that I_i^{out} is lower than zero when emigrants of country i move on average to a destination country with a higher economic strength or community level. In a opposite way, I_i^{in} is lower than zero when immigrants of country i comes from origin countries with higher economic strength or community level.

The approach can be extended by aggregating values at different levels. For instance we can derive both indicators for a set K (of size k) of countries $i \in K \subset V$ (for instance all countries of the same macro area or the same continent)

$$I_K^{in} = -\frac{\sum_{i \in K} \sum_{j \neq i \in V} I_{ji} w_{ji}}{\sum_{i \in K} \sum_{j \neq i \in V} w_{ji}} = \frac{\sum_{i \in K} I_i^{in} s_i^{in}}{\sum_{i \in K} s_i^{in}} \quad (13)$$

$$I_K^{out} = \frac{\sum_{i \in K} \sum_{j \neq i \in V} I_{ij} w_{ij}}{\sum_{i \in K} \sum_{j \neq i \in V} w_{ij}} = \frac{\sum_{i \in K} I_i^{out} s_i^{out}}{\sum_{i \in K} s_i^{out}} \quad (14)$$

It is noteworthy that the aggregate indicators can be derived as the weighted average of country indicators with weights equal to in (out) strength (i.e. s_i^{in} and s_i^{out}).

For $K = V$, we have *global* (“in” and “out”) socio-economic indicators of the migration flows as a weighted mean of the local socio-economic indicators. By construction, all the aggregated indicators are in the range $[-2, 2]$. Therefore, it is possible to derive insights from the comparison among the local indicators and the global ones.

4 Empirical experiments

In order to test the performance of the indicator proposed in the previous Section, we consider data based on international migrant stock in 2015. In particular, the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat provides the international community with timely and accessible population data and analysis of population trends and development outcomes for all countries and areas of the world. As regard to the data we deal with, the dataset presents the estimates of total migrant stock at mid-year 2015 by origin and destination. The estimates are available for all countries and areas of the world, based on official statistics on the foreign-born or the foreign population (mainly based on population censuses but also on population registers and nationally representative surveys). Specific adjustments are also made by the Department in order to exclude both foreign-born persons naturalized in their country of residence

and children born to international migrants that live in countries where citizenship is granted upon birth (i.e a birthright citizenship based on jus soli). Finally, also the refugees' statistics reported by international agencies have been used by the Department in order to catch the number of person recognized as refugees or find themselves in refugee-like situations (see [6] for details).

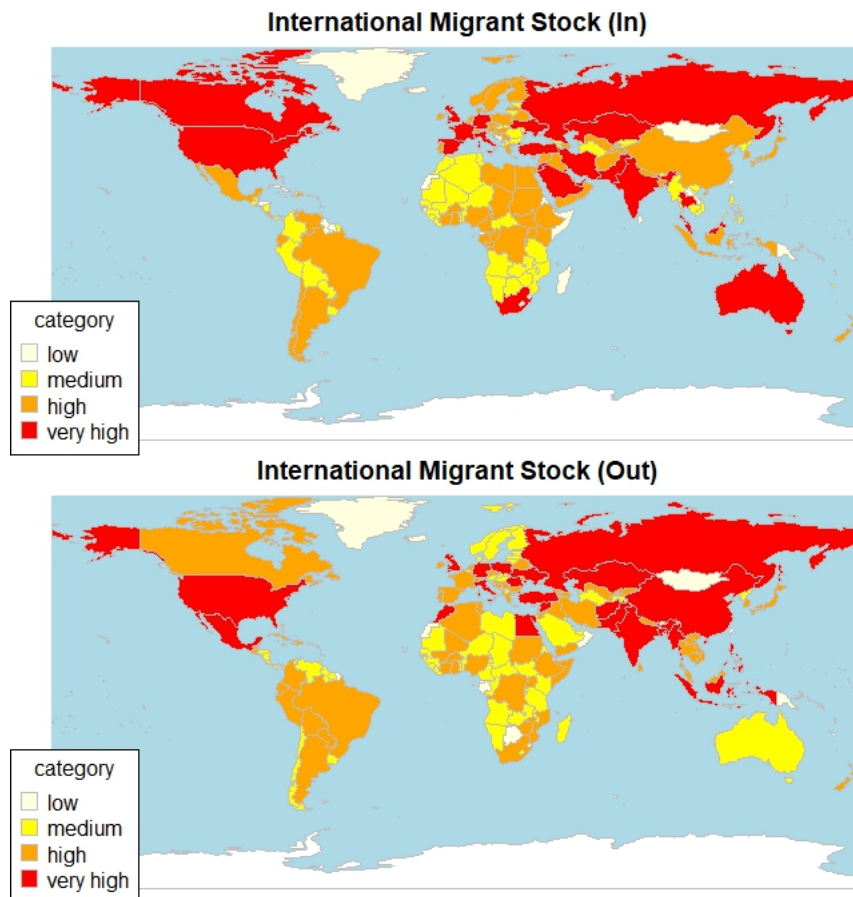


Figure 1: Total Stock of migrants (2015 data) according to different destination (in) and origin countries (out). Categories are based on the following classes $[0 - q_{30}]$, $(q_{30} - q_{60}]$, $(q_{60} - q_{90}]$, $(q_{90} - q_{100}]$ where q_p is the p -quantile of the distribution. The top figure displays the number of immigrants observed into different countries. Northern America, Central and Western Europe, India, Russia and richest Arabian countries are top of the list. Bottom figure shows instead countries with the most emigrants living abroad. India, Russia and Mexico are close contenders in this case.

We have at disposal data of 231 countries also including dependencies and areas with special Sovereignty. To give a brief idea of the heterogeneity between different countries, we display in Figure 1 (top) the number of immigrants observed in 2015 into different destination countries (“in” values). In a similar way, we report at the bottom in Figure 1 the number of emigrants according to the origin country. Data have been plotted by dividing the distribution into four different classes (from “low” to “very high”) where bins are based on the 30th, 60th and 90th quantiles of the

distribution⁶. For instance, origin (or destination) countries, that have a number of migrants lower or equal than the 30th quantile, belong to the “low” category.

In 2015, two thirds of all international migrants were living in only 20 countries, starting with the United States of America, which hosted 18.6% of all migrants, with an immigrant population that climbed in 2015 at 43.6 million⁷. Germany, Russian Federation, Saudi Arabia, United Kingdom, United Arab Emirates, Canada, France, Australia, Italy and India follow in decreasing order with a number of immigrants greater than 5 million.

On the other hand, India has the largest diaspora in the world with 16 million people living outside of their country, followed by 12 million from Mexico. Other countries with larger number of emigrants included the Russian Federation, China, Bangladesh, Pakistan and Ukraine. Of the fifteen countries with the largest number of international migrants living abroad, 9 are in Asia and 5 in Europe.

As described in Section 2, a network is used to describe migrants’ flow where nodes are countries (i.e. 231 nodes) and weighted directed links, computed as in (3), consider migration between origin and destination countries. The network is plotted in Figure 2.

We can compute some classical network indicators. Values of degree and node strength provide meaningful insights. Ranking of top ten countries according to these indicators are reported in Table 1. As regard to destination countries, we observe the case of Chile, characterized by a low number of immigrants (in-strength is approximately 0.27% at the 67th position) but also by a significant heterogeneity of people that comes from 209 different origin countries⁸. Furthermore, of the twenty countries with the largest in-degree, fifteen are in Europe. This result is also an effect of million of migrants and refugees that crossed into Europe over the last years.

According to “out” indicators, we observe that India has the highest out-strength but also a high out-degree, meaning that Indian people are widely distributed in different countries of world. A different situation is instead observed for Mexico that, although a high diaspora (i.e. a high out-strength), has the 98% of its emigrants settled in USA.

⁶We did not use the classical grouping based on quartiles in order to emphasize differences on the right tail of the distribution.

⁷According to the U.S Census Bureau, an increase of 12.6 million has been observed since the turn of the century. Immigrants now comprise 13.5% of the U.S. population, roughly one out of eight residents, the highest share in 106 years.

⁸Immigration has indeed contributed to the demographics and the history of this South American nation. This country has been characterized by different flows of immigration over time from Europe, South America, North and Central America, North Africa and Asia and Sub-Saharan Africa. Today, continuing economic growth and reconsolidated political stability have positioned Chile as an emerging country of destination.

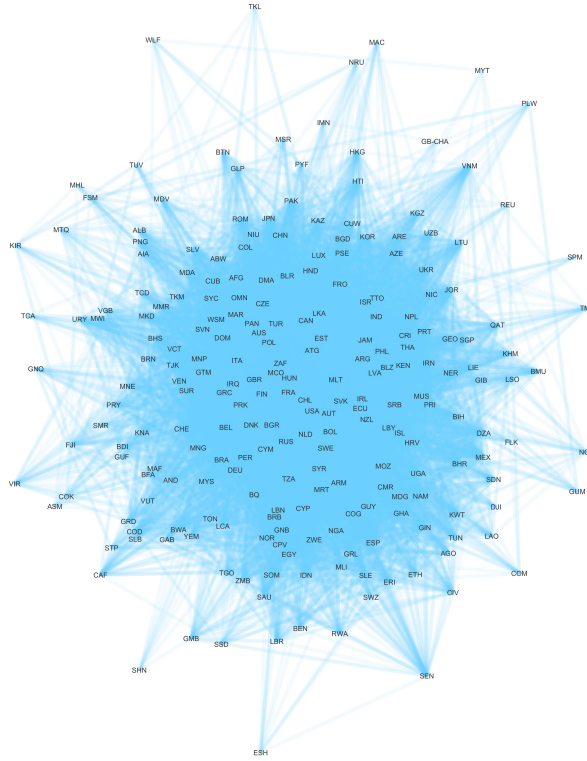


Figure 2: Network of migrants. Each country is a node and a weighted directed edge measures total migrants' flow from a country to another one. Edges opacity is proportional to weights (i.e. number of migrants).

Countries	in-degree	Countries	in-strength	Countries	out-degree	Countries	out-strength
Chile	209	USA	18.62%	USA	161	India	6.66%
France	206	Germany	5.13%	UK	144	Mexico	5.27%
Australia	206	Russian Federation	4.98%	China	143	Russian Federation	4.52%
UK	205	Saudi Arabia	4.21%	France	137	China	4.08%
Canada	197	UK	3.63%	India	128	Bangladesh	3.08%
Ireland	195	UAE	3.42%	Canada	127	Pakistan	2.54%
Greece	194	Canada	3.35%	Germany	125	Ukraine	2.49%
Italy	192	France	3.33%	Italy	110	Philippines	2.27%
Austria	192	Australia	2.86%	Australia	107	Syrian Arab Republic	2.14%
Denmark	186	Spain	2.50%	Japan	102	UK	2.10%

Table 1: Top 10 countries ranked in terms of in/out-degree and in/out-strength.

In order to catch the neighbourhood structure of each destination and origin country in terms of migrant, we compute directed clustering coefficients as described in Section 2. Through the clustering coefficient, we want to catch both the number of relations between different countries and the number of migrants involved.

It is noticeable that main patterns are caught by the coefficient but also some specific behaviours (see Figure 3). For instance, we observe on average a very low in-clustering for the African continent and part of the Western Asia. Migrants in these countries comes in general from countries that are not related each other leading to a low coefficient. An exception is represented by South Africa, that

is characterized by a high clustering because of different kind of migrants. In particular over the last years, neighbouring people from different nationalities that have fled the political and economic instability at home (Nigerians, Zimbabweans, Congolese and Somalis for instance), go in South Africa for the hope of employment and political protection (see [10] for an analysis of intra-continental migration within Africa).

As regard to the “out” values, it is noteworthy that Mexico does not belong to the very high category, despite a huge number of people living abroad. As said before, this is a consequence of almost unidirectional diaspora towards USA. Furthermore we observe that North America, Europe, China, India and Australia are characterized by very high value of clustering. Finally, it is interesting the case of Brazil characterized by both in and out-clustering. On one hand this country is characterized by a huge variety of immigrants, where current trends show the presence of people of different nationalities⁹. On the other hand, Brazilian people are leaving the country in the search of employment, as shown on a research made by International Centre for Migration Policy Development (ICMPD). There is a growing emigration towards Europe, but United States, Japan and Paraguay are at moment the most popular destinations.

The indicator proposed in previous Section has been computed at country level in order to emphasize differences between countries in terms of immigrants and emigrants (see Figure 4). As regard to the origin country, we observe a prevalence of negative values. It means that on average origin countries has economic strength and community structure lower than the destination countries. In particular, the movement of migrants is often towards countries with a greater GDP. Exception are represented by countries with a very high level of GDP and clustering. United States, China, Russian Federation are the only three countries with an average index significantly greater than zero. Other countries with a high GDP level (as Japan and UK) have a weighted average (see formula (12)) lower than zero because of the significant number of migrants towards countries with a higher GDP (as USA and China). For instance, Japanese people live abroad in 102 different countries, but almost half of them are in USA and China. In a similar way, United Kingdom has a positive index $I_{i,j}$ with respect to 140 different countries, but the weights of negative indices (based on the relations with USA, China, Japan and Germany) lead to a total negative indicator.

Strong negative values of I_i^{out} are instead observed for Central America and Caribbean countries. For instance, Mexico and Puerto Rico rank at the bottom, because, as well-known, these countries are characterized by a significant emigration towards USA.

According to the countries where migrants settle, several destination countries have a community and GDP level higher than the country of origin of migrants. Other than well-known results (as for USA, China and Russian Federation), it is noticeable the case of Europe where on average immigrants come from “weaker” countries. Effects of recent migration from Africa, Western Asia

⁹For instance, since 2010 Brazil is home to 4,251 refugees from 76 different nationalities.

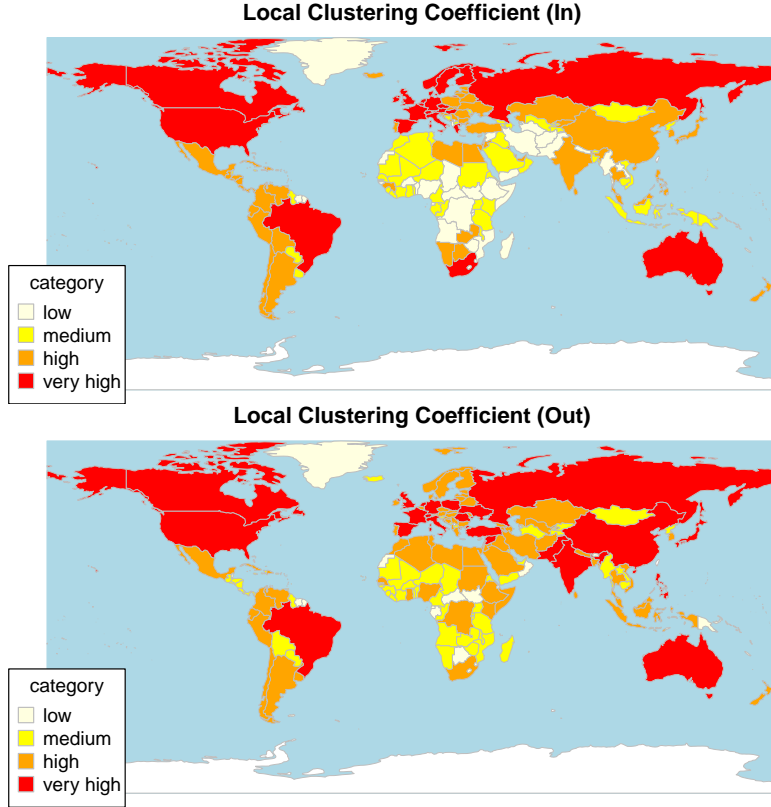


Figure 3: In and out local clustering coefficients of different countries. Data have been plotted by dividing the distribution into four different classes (from “low” to “very high”) where bins are based on the 30th, 60th and 90th quantiles of clustering distribution. We depict the neighborhood structure of different countries (i.e. a red colour means a high neighbourhood structure in terms of in-flows or out-flows). Main patterns are caught, but also some specific behaviours (as high immigration in South Africa from neighbouring countries and the emigration of Mexicans towards USA)

and East Europe is here caught.

It is also very singular the case of Nigeria. On one hand, as Africa’s most populous country, Nigeria, with an estimated population of 150 million and over 250 ethnic groups, deals with a range of migration issues, from massive internal and regional migration to brain drain and a large, well-educated diaspora in the West (mainly towards the United States and the United Kingdom) that it sees as key to future development (indeed $I_i^{out} = -6.3\%$). On the other hand, country is characterized by an immigration from other African countries (as Benin, Chad, Liberia, Mali) with a lower level of GDP and community (indeed $I_i^{in} = 0.6\%$). In particular, this is also a consequence of the development of ECOWAS¹⁰ treaty, between countries of Western Africa in order to strengthen regional economic

¹⁰The Economic Community of West African States (ECOWAS) Treaty is a multilateral agreement signed by the member states that made up the Economic Community of West African States. Benin, Burkina Faso, Cape Verde, Cote d’Ivoire, Gambia, Guinea, Guinea Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leone and

integration through progressively free movement of goods, capital, and people and to consolidate states' efforts to maintain peace, stability, and security (see [7]). At the bottom of the ranking (with $I_i^{in} \leq -10\%$), we have mainly countries of Eastern Asia (as Hong Kong, Macao, Mongolia, Myanmar, Philippines) characterized by an immigration from China and also from other countries with higher GDP (for instance from Russia for Mongolia, from USA for Hong Kong and Philippines, from Portugal for Macao).

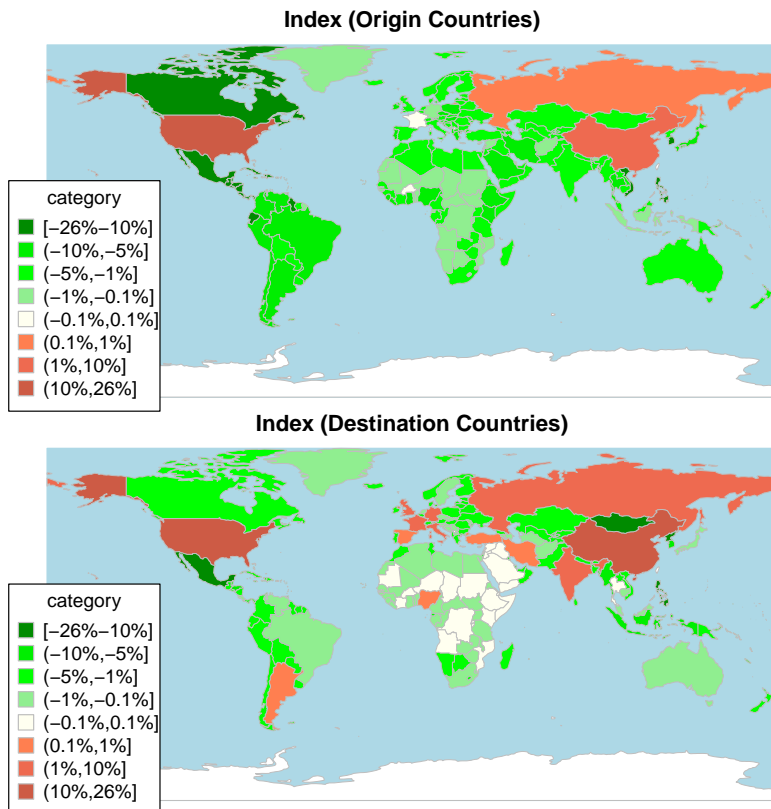


Figure 4: Figure reports indicators for different origin and destination countries (see formulas (11) and (12)). In top figure, a negative index (green colour) means that people of that country migrated on average to countries with a greater economic strength and community level. In general, results confirm that migrants move toward destination countries with a higher community and GDP level. As regard to the destination countries (bottom figure), we observe a negative index when, on average, immigrants come from richer or more interconnected countries.

Indicators have been also computed by grouping countries according to a macro area classification. We followed the composition of geographical regions used by the Statistics Division of United Nations in its publications and databases. Each country or area is shown in one region only. These geographic regions are based on continental regions, further subdivided into sub-regions and inter-

Togo have accordingly agreed to a Revised Treaty of 24th July, 1993

mediary regions drawn as to obtain greater homogeneity in terms of population size and accuracy of demographic statistics. Results are displayed in Table 2.

It is noteworthy that emigrants move on average towards countries and regions with a higher GDP or community structure. Exception are represented by Northern America and Eastern Asia. However, Eastern Asia is characterized by a significant heterogeneity with a positive index for China and very negative values for Republic of Korea and Hong Kong. As already stressed before, at the bottom we have Caribbean and Central America with a significant emigration towards other countries of the North and South American continents.

Looking instead at values of I_K^{in} , Northern America and all European sub-regions show positive indices. The effect of recent migration flows to EU ([17]) explain these values. Also Eastern Europe seems more affected than in the past by immigration. How many people working in this sector are suggesting for some time now: Europe's restrictive policies and tightened borders do not deter overall migration but instead merely divert flows towards different routes.

UN geographic regions	I_K^{out}	UN geographic regions	I_K^{in}
Northern America	12.0%	Northern America	20.5%
Eastern Asia	2.5%	Western Europe	2.0%
Middle Africa	-1.2%	Northern Europe	1.0%
Western Europe	-1.4%	Eastern Europe	0.5%
Southern Asia	-1.9%	Southern Europe	0.3%
Northern Africa	-2.1%	Southern Asia	0.2%
Central Asia	-2.2%	Western Africa	0.0%
Western Asia	-2.3%	Eastern Africa	-0.1%
Eastern Europe	-2.3%	Middle Africa	-0.2%
Northern Europe	-2.4%	Northern Africa	-0.4%
Western Africa	-2.5%	Western Asia	-0.6%
Southern Africa	-2.6%	Southern Africa	-0.6%
Eastern Africa	-2.6%	South America	-0.7%
Australia and New Zealand	-3.3%	Australia and New Zealand	-1.3%
Southern Europe	-3.6%	Central Asia	-1.4%
Polynesia	-4.8%	South-Eastern Asia	-1.4%
South-Eastern Asia	-5.8%	Melanesia	-2.3%
Melanesia	-5.9%	Polynesia	-3.1%
South America	-7.0%	Caribbean	-5.5%
Micronesia	-8.9%	Eastern Asia	-6.1%
Caribbean	-19.3%	Micronesia	-6.5%
Central America	-23.2%	Central America	-11.5%

Table 2: Out and in indicators for each macro area based on UN classification (see formulas (14) and (13) respectively). Indicators are ranked in decreasing order.

5 Conclusions

Immigration is now a prominent feature in the economic, social, and political landscape of many countries. The reception and assimilation of immigrants is indeed a significant economic and social phenomenon in many previous emigration countries. This paper provides a unified view of the analysis of migration whose nature is of social and economic type. To this aim, worldwide migration is here presented as a network and a related socio-economic indicator is proposed. Both the macroeconomic status and the community structure of origin and destination countries are taken into account. Indeed, the interconnection plays a fundamental role in determining the migration patterns. The interconnectedness of the network is here represented by means of the weighted clustering coefficient, referring to both immigration for destination countries and emigration for origin ones.

Empirical analysis confirms that migrants move, on average, toward destination countries with a higher community and GDP level. Furthermore, the indicator proves effective also in capturing patterns of migration flows showing at different “observation scales” peculiarity of countries or macro-areas of the world. Hence, we are able to catch distinctive behaviours related to specific contexts.

References

- [1] M. Beine. The role of networks for migration flows: an update. *International Journal of Manpower*, 37(7):1154–1171, 2016.
- [2] M. Beine, S. Bertoli, and J. Fernández-Huertas Moraga. A practitioners guide to gravity models of international migration. *The World Economy*, 39(4):496–512, 2016.
- [3] P. Boyle, K. H. Halfacree, and V. Robinson. *Exploring contemporary migration*. Routledge, 2014.
- [4] S. Castles, H. De Haas, and M. J. Miller. *The age of migration: International population movements in the modern world*. Palgrave Macmillan, 2013.
- [5] G.P. Clemente, R. Grassi. Directed Clustering in weighted networks: a new perspective. *Chaos, Solitons & Fractals*, 107:26–38, 2018
- [6] Department of Economic and Social Affairs. Trends in international migrant stock: the 2015 revision. Technical report, United Nations, 2015.
- [7] Economic Community of West African States. Revised treaty of the economic community of West African states. Technical report, 1993.
- [8] G. Fagiolo. Clustering in complex directed networks. *Physical Review E*, 76(2), 2007.

- [9] G. Fagiolo and M. Mastrorillo. The international-migration network: Topology and modeling. *Physica Review E*, 88, 2013.
- [10] M.L. Flahaux and H. De Haas. African migration: trends, patterns, drivers. *Comparative Migration Studies*, 4(1), 2016.
- [11] T. Goldade, B. Charyyev, and M. H. Gunes. Network analysis of migration patterns in the united states. In Chantal Cherifi, Hocine Cherifi, Márton Karsai, and Mirco Musolesi, editors, *Complex Networks & Their Applications VI: Proceedings of Complex Networks 2017 (The Sixth International Conference on Complex Networks and Their Applications)*, pages 770–783, Cham, 2018. Springer International Publishing.
- [12] M. J. Greenwood. Modeling migration. *Encyclopedia of social measurement*, 2:725–734, 2005.
- [13] J. Grogger and G. H. Hanson. Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, 95(1):42–57, 2011.
- [14] F. Harary. *Graph theory*. Addison-Wesley, Reading, MA, 1969.
- [15] S. Haug. Migration networks and migration decision-making. *Journal of Ethnic and Migration Studies*, 34(4):585–605, 2008.
- [16] H. Helfer. Prosperity-enhancing institutions: Towards a comprehensive composite index. *Social Indicators Research*, 134(3):805–845, 2017.
- [17] International Organization for Migration. Migration flows to Europe. Technical report, United Nations, 2017.
- [18] P. Levitt and B. N. Jaworsky. Transnational migration studies: Past developments and future trends. *Annu. Rev. Sociol.*, 33:129–156, 2007.
- [19] D. S. Massey, J. Arango, G. Hugo, A. Kouaouci, and A. Pellegrino. *Worlds in Motion: Understanding International Migration at the End of the Millennium: Understanding International Migration at the End of the Millennium*. Clarendon Press, 1999.
- [20] M. P. McAssey and F. Bijima. A clustering coefficient for complete weighted networks. *Network Science*, 82(6):061110, December 2015.
- [21] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 6:28384, June 2003.
- [22] E. Tranos, M. Gheasi, and P. Nijkamp. International migration: a global complex network. *Environment and Planning B: Planning and Design*, 42(1):4–22, 2015.

- [23] N. K. Vitanov, M. Ausloos, G. Rotundo. Discrete model of ideological struggle accounting for migration. *Advances in Complex Systems*, 15(1):1250049, 2012.
- [24] N. K. Vitanov, K.N. Vitanov. On the motion of substance in a channel of a network and human migration. *Physica A* 490:1277-1294, 2018.
- [25] M. Windzio. The network of global migration 1990–2013: Using ERGMs to test theories of migration between countries. *Social Networks*, 2017.