WP 24
23 April 2010

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint Eurostat/UNECE Work Session on Demographic Projections
(28-30 April 2010, Lisbon, Portugal)

Item 9 – Forecasting demographic components: migration

The role of social networks in the projection of international migration flows: an agent-based approach

Carla Anjos, University of Aveiro, and Pedro Campos[1], Statistics Portugal and University of Porto, Portugal

ABSTRACT

In recent years, migration has been modelled within the perspective of social networks. Models are based on the idea that migration flows are influenced by the social networks where the agents operate. In this work we use a Multi-Agent System to simulate social networks of migrants and analyse the impact of the structure of these networks in the flow of migrants. The model we propose uses information of a IPUMS database of immigrants in the United States. We focused in four different countries of origin: Germany, Mexico, Portugal and China and in six variables: age, educational level, income, number of people in the household, labour status and number of individuals in the social network of the agent. We have analysed four important measures of network structure: density and three measures of degree centralization: input, output and general. Results indicate that Mexicans have higher input and general degree centralization, meaning that their networks have higher levels of influence of the agents. We concluded that the agents that stay in the U.S. (and do not go away to their country of origin), have network connections that are weaker than those of other agents.

1. INTRODUCTION

Social network analysis has had a great development in recent times, although the main concepts were proposed between 1960 and 1970. According to Mitchell (1969), "a social network is a specific set of links among a set of persons", with the additional property that the characteristics of these links as a whole may be used to interpret the social behaviour of the persons involved. Likewise the network structure reflects the pattern of relationships between individuals (Newman & Girvan, (2004), Wasserman & Faust, (1994)). Migration is often modelled within the perspective of social networks (McKenzie & Rapoport (2007), Woodruff & Zenteno (2007), Hussey (2007)). In general it appears that migration flows towards a specific country B with origin in country A induce further migration of other individuals from A to B (Helmenstein & Yegorov, 2000).

In this work we use a Multi-Agent System to model the flow of migrants in social networks. Agent-Based Computational Demography (ABCD) is a computational approach to Demography, with emphasis placed on simulation based on agents (Billari and Prskawetz, 2005). In particular, in ABCD models, it is relatively easy to integrate micro-based demographic behavioural theories with aggregate-level demographic outcomes. In such models, space and networks can be formalized as additional entities in which the agents will interact (Billari, Ongaro, Prskawetz, 2003). Our model is based on the idea that migration flows are influenced by the structure of the social network where the agents operate. The proposed model of migration uses four countries (Germany, Mexico, Portugal and China) and six variables: age (y), educational level, (e), household income (r), number of people in the household, (p), labour status (working/not working), (w) and number of individuals in the social network of the agent (s).

[1] E-mail: pedro.campos@ine.pt

The social network of the individual is simulated, based on a correlation with the number of persons in the household. We have used information from the IPUMS (Integrated Public Use Microdata Series, Ruggles et al, (2009)), containing data of migration flows to the United States between 2001 and 2008. We decided to consider four communities in the U.S. with origin in four different countries (Portugal, Mexico, China and Germany), and identified the force that these communities have to attract new migrants from the country of origin, as well as to repel those already leaving in the U.S. We used a gravitational model incorporating the mass of the agent (M_a) that includes the household income (r), the number of people of their household (p), the agent's age (y) and the labour status of the agent, or working situation, (w). Similarly, the mass of the corresponding social network of the agent (M_N) was computed. To determine the gravitational force of migration, a distance d between the agent and their social network was defined. The Euclidean distance was chosen as the distance measure. Migration costs have also been considered. We analysed the structural properties of the generated networks and compared the impact of these structures on the magnitude of the migration flows. Empirical validation was applied in order to evaluate the quality of the model. The simulation was initialized with real data of 2000, based on the clustering of the main properties of the individuals coming from the four different countries of origin. After running the model and iterating it for nine years, we compared the simulated migration flows with the real data observed in the US in the period of analysis.

Measures of network structure were used: density degree centralization (input, output and general). The latter constitute signs of agents' influence in network since the higher these degree centralization measures, the more central agents are in the social network. We concluded that the agents that stay in the U.S. (and do not go away to their country of origin), have higher values of individual mass, Ma, and are associated with less average distance with the members of their social networks. Networks where these agents belong are more cohesive. According to the results of the simulation, in 2008, Germans, Chinese and Portuguese have similar values of the degree centralization measures, meaning that agents occupy central positions in networks. Simultaneously, the networks keep growing during the period. Centrality is therefore an important property for the growth of networks. We have concluded that Mexicans have higher input and general degree centralization, meaning that their networks have higher levels of influence of the agents.

The paper is structured as follows: in Section 2, an overview of the concepts of Migration and Social Networks is made. In Section 3, we introduce the approach of Agent-Based Computational Demography (ABCD) and the methodology we have used in this study. Results are discussed in Section 4. In this section we also analyse the stability and the validation of the model. Conclusions are in section 5, where we propose topics for further work.

2. MIGRATION AND SOCIAL NETWORKS

2.1. Migration studies

Studies on migratory flows are essential to the achievement of more accurate demographic forecasts. Migration flow is to be viewed as the movement of a person through a limited space, with the intention of changing residence temporarily or permanently. International migration (migration between countries) and internal migration (migration within a country or within a particular region – within a country or an aggregation of countries) constitute different perspectives approaching the problem. Internal flows of migrants associated with the movements of persons within a country are difficult to register. In some situations like in European Union, country borders have lost some of their previous economical sense, and internal flows of persons, goods and services can now circulate with fewer restraints in the whole territory. Therefore it is not easy to correlate these flows with the specific borders of each country and alternative measures are needed to calculate migration flows.

There are several methods for the development of studies of migratory flows: registers of the entries and exits can be produced in most regions where the possibility of having such control in the borders exists. In other situations, national and international sampling surveys produce information on the flow of migrants: different countries may combine statistics obtained from these surveys and estimate their own flows in the perspective either of the exits (emigration) or the entry (immigrants). For example, Portuguese emigrants in France are considered immigrants in that country. There are other different approaches to the prediction of migration flows, such as modelling migration flows using stochastic and deterministic models, as in Maier and Weiss (1991) where a random utility, model (based in the regional utility of migration) is presented. Stillwell and Congdon (1991) make a very complete approach of deterministic and stochastic migration models.

When the results of migratory flows obtained with these models are included in population projections, the corresponding values are therefore considered in the equations for estimating or projecting the population totals, according to the most appropriate methods (Shyrock et al. (1976)): (i) for computing post-censitary estimates associated to current and past flows,

after a census, taking into account all the census until (and including) a particular one, not including further census; (ii) Projections, associated to periods after the last census, for which there is not available data.

2.2. The social network perspective

In recent years, migration has been modelled within the perspective of social networks. The works of McKenzie and Rapoport (2007), Woodruff & Zenteno (2007) and Hussey (2007) are examples of such applications. In general it appears that migration flows towards a specific country B with origin in country A induce further migration of other individuals from A to B (Helmenstein & Yegorov, 2000). Social network analysis has had a great development in last decades. According to Mitchell (1969), "a social network is a specific set of links among a set of persons", with the additional property that the characteristics of these links as a whole may be used to interpret the social behaviour of the persons involved. Likewise the network structure reflects the pattern of relationships between individuals (Newman & Girvan, (2004), Wasserman & Faust, (1994)). In the literature of social networks, relationships are explored in network construction and shaped by the analyst in the most appropriate format. Wasserman and Faust (1994), for instance, describe a full comprehensive approach of the role of social networks in the Social Sciences.

In the work of Woodruff and Zenteno (2007), migration induces the reduction of inequality between individuals of the same community. The model takes into account the relationship between the wealth of individuals and migration. Hussey (2007) reports a study on migration of medical doctors in order to identify the countries with the highest rate of migration in the United States. Neto and Mullet (1998) study a group of 40 Portuguese adolescents according to 20 different variables. The group is also quantitatively divided into descendants of the working class and middle class. The authors identified several interesting features in this study, as the intention to migrate (which is much higher when family or friends exist in the host country), the effect of the wage gap (which is greater when the job opportunity in question is good, as it has greater influence when there is already a social network in the host country. The effect of employment opportunities has a greater influence when there is a network and the network effect is greater when the job opportunity is relatively good. On the other hand, the way how employment opportunities affect the wage gap varies depending on the presence or absence of a social network.

2.3. Some concepts of social networks

Social Network analysis examines the relationships between individuals (actors or agents) and is based on graph theory. A graph is a set of nodes (agents, or vertices) and lines: lines represent links, meaning relationship between the vertices (Lemieux and Ouimet (2008)). It is assumed therefore that society is an organized structure of agents and not just an aggregation of agents. Each agent is an individual, (a person), or a set of social, economic or cultural units. The main purpose of social network analysis is not only to analyse the population, but rather to detect and interpret social standards (Nooy et al. (2005)).

In social network analysis the relationship between two nodes or agents may or may not have an orientation. If the relation between two agents A and B is oriented, then the link between them is named "arc", defining the targeted transmission of a flow (that represents information, goods, etc.) and can be of two types: from A to B ($A \rightarrow B$), or, instead, from B to A ($A \leftarrow B$). If, on the other hand, the relation between two agents has not a specific orientation, but it only means that two agents A and B are connected (the orientation being represented as $A \leftrightarrow B$ or simply A-B), then the relation is called "edge" or "link "(Lemieux and Ouimet (2008); Nooy et al. (2005)).

The graphical representation of the network structure is an important step for understanding and interpreting the social network. For that purpose, specific programs exist, such as Pajek (Pajek (2010)), where it is possible to display the structure and calculate the parameters that characterize the social network. In this work we are interested in determining the importance of each agent in the network. Therefore, a set of measures of node centrality are used. In each vertex it is possible to measure a level (degree) of centrality, corresponding to the number of lines involving the vertex. If the network is not oriented, the degree of each vertex is equal to the number of its neighbours (the adjacent vertices). However the calculation is different when networks are oriented. In this case it is necessary to distinguish the links leaving the vertex (outdegree) from those that are coming to the vertices (indegree) (Nooy et al. (2005)). Centrality is a measure aiming at comparing the position (more or less central) of an agent, and is given by the ratio between the number of connections of agents and the total number of connections. There are at least three measures of centrality most frequently cited (Freeman (1977; (1979)):

- Centrality degree (degree centrality) - measures the number of direct connections of each agent in a graph
- Centrality of proximity (closeness centrality) - measures the length of the shortest path connecting two agents.
- Betweenness centrality - measure the importance of a member in the network.

In addition to these measures of centrality, there are several other possible measures for network analysis (Campos (2008)), such as:

- Clustering (transitivity) - measures the connectivity within the network, and it is expressed by the probability that two neighbours of a given vertex are connected;
- Density - defined as the ratio between the number of relationships and the number of possible relationships. If the network is oriented, the number of possible relationships is equal to the number of vertices N multiplied by N-1. If the network is not oriented, then the number of possible relationships is given by N (N-1) / 2 (Lemieux and Ouimet (2008)).
- The average path length - measures the length of the network, and is given by the average number of links in the shortest path between any two pairs of vertices;
- Diameter - measures the length of the network, and is the result of the maximum number of links in the shortest path between any two vertices;
- Degree of centralization (degree centralization) – it is determined based on the centrality of the degree of network agents, and measures the centrality that exists on the network.

In the present study, we analyse the centrality and density of networks. Our aim is to verify to what extent the existence of networks with major centrality contributes to favour the increase the number of migrants. We believe that in cases of higher centrality, the information flows easily through the links and nodes of the network, but the center of the network is critical for the transmission of information. The degree of centralization of a network is the ratio between the variation in degrees of the vertices and the maximum degree of variation, which is possible in a networks having the same size (Nooy et al. (2005)). The centralization degree may be analysed in three different perspectives in oriented networks: input centralization degree (taking into account the arcs of entrance towards an agent), output centralization degree (taking into account the arcs of exit from an agent) and general centralization degree (taking into account both input and output degrees).

These network measures are analysed later in Section 4. In the next section, we introduce the model within the Agent-Based model perspective.

3. ABCD, MAS, AND THE MODELLING OF MIGRATION NETWORKS

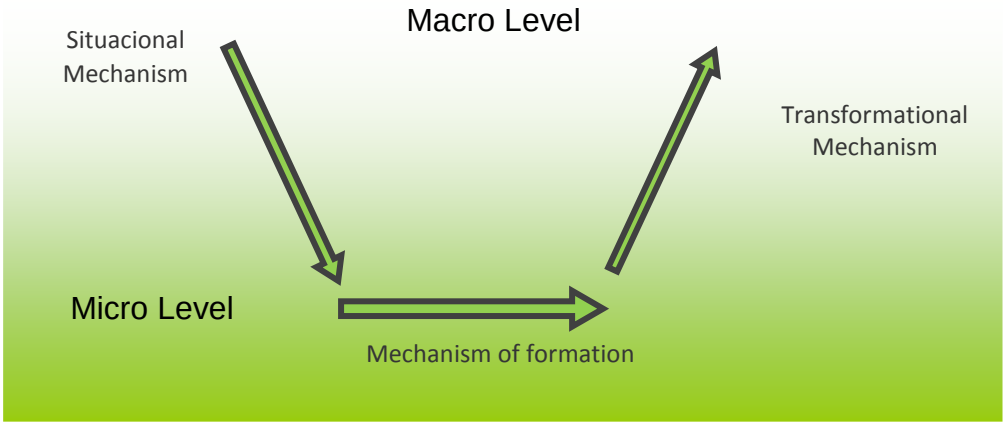
3.1. Agent-Based Computational Demography (ABCD)

In this section, we introduce the ABCD perspective of modelling demographic issues and present the Agent-Based approach to model the role of Social Networks in the projection of international migration flows.

In the last decades, simulation has been a useful tool in Demography. Recently, the area of Agent-Based Models was applied to Demography. In fact, Agent-Based Models may be seen as containing (or being contained in) the wide scope of simulation techniques. Agent-Based Computational Demography (ABCD), suggested in Billari et al. (2003a) and Prskawetzy and Billari (2005), focuses more on the explanation of the behaviour of agents in Demography than on the usual demographic forecast. Thus, individual agents are modelled and the overall result of their interaction is studied in a "bottom-up" perspective.

According to the social model of Coleman, cited by Billari and Prskawetzy (2005), there are three transitions that explain the phenomenon at the macro level (Figure 1). The first transition occurs from the macro to the micro and concerns the influence of a social macro affecting each individual. In the second transition, the interactions of individuals are explored. Finally the last transition is the influence of interactions, at the individual level, on the macro level.

Figure 1 - Diagram of the interaction of social mechanisms in Agent-Based Computational Demography adapted from Billari and Prskawetzy (2005)



One of the goals in applying ABCD is to understand the transition from the micro to the macro level. In fact, simulation using Multi-Agent Systems has turned out to be a very useful technique and is rapidly becoming important in every scientific field because of its simplicity and efficiency when run on existing computers. The fact that it is not required to use fully rational agents is a challenge for the classical mathematical models which defines formal equations to guide agents' behaviours. Another advantage of ABCD is the capacity to build models that provide answers to problems that do not have analytical solutions (Billari et al. (2003a); Billari and Prskawetzy (2005)). In addition, agent-based models are used in situations that can be solved by other models, but models with agents are more "visual" and easy to understand (Axtell (1999)).

Agent software has been much influenced by the work of Artificial Intelligence, (AI) especially the subfield called Distributed Artificial Intelligence (DAI) (Bond and Grasser, 1988). DAI is very important to social simulation, because it pays attention to building networks of intelligent agents and investigating their properties.

In the literature of Multi-Agent Systems (MAS), an agent is usually defined as an entity that lives in a particular environment and has the ability to interact with other agents. According to Ferber (1999) an agent has the following characteristics:

- Action and interaction - agents interact with other individuals and the environment. Actions modify the environment of the agents, and hence future decisions to be taken;
- Communication with other agents - the main form of communication with other agents;
- Aims and individual autonomy - agents are not controlled by commands from the user or another agent, but by a set of "trends" that may take the form of individual objectives or survival functions that agents try to maximize;
- Perception - the agents have only a limited or partial perception of the environment in which they live. They have a global perception of everything that happens around them. Often it is assumed that agents have a "bounded rationality" in the sense that use a limited computing resources to extract the consequences of what was seized.

Wooldridge (2002) and Gilbert and Troitzsch (1999) present a complementary definition of the aspects described above. According to these authors, in order to characterize an agent it is necessary to include other aspects such as reactivity, pro-activity, social skills and autonomy. These aspects are not characterized in the present study. In Section 3.3., the social skills of agents are defined with more detail, as it is important to have a good adherence between models and reality.

3.2. Motivation and hypothesis

The promising developments presented in the studies of social networks and the desire to frame social factors in the migration of individuals constitutes the motivation of this work. The general hypothesis to be tested is that the structure of social network influences the level of migratory flow. To this end, the analysis of social networks resulting from the simulation is very important. This analysis aims at highlighting two important aspects, which can be seen as research hypotheses:

1) The density of the network influences the quantity of the migratory flow;
2) The degree of centralization of the network influences the quantity of the migratory flow.

To calculate the density that is used in Hypothesis 1), both the number of arcs (links) of the network, and the total number of agents (persons) are computed. Higher density means more connections between individuals, reflecting a more intense social network structure. The degree of centralization is used to evaluate the performance of social relations in the network.

3.3. Data and methods

3.3.1. Data and representative groups of agents

As referred above, when simulation is used, then the social skills of agents have to be defined with more detail, according to empirical observations, as it is important to have a good adherence between models and reality. A real database, selected from the IPUMS project has been used, in order to gather information about the actual situation of the population to be modelled in the Multi-Agent System. IPUMS[2] database has been used for this project, containing micro-data (individual registers) gathered from several surveys conducted in the United States of America.

Table 1: variables collected in the IPUMS database

variable	Description	Scope
y	Age of the agent	{1, …, 95}
e	Educational level of the agent	{1, 2, 3}
r	Income of the houseold ($/1000)	[2; +∞[
p	Number of individuals in the household	{1, 2, …, 15}
s	Number of individuals in the agents' social network	{2, …, 20}
w	Labour status (working situation: working/not working)	{0,1}

Six variables were selected to describe the most important aspects that we aim at modelling. The original database of IPUMS contains information about the main variables of migration for several years. It addresses 154 countries of origin with immigrants living in the United States. For this study, we have selected nine years (2000 to 2008) and four countries of origin: Germany, China, Mexico, and Portugal. The reasons for the selection of the countries are the following:

- Countries belong to three different continents, with different territorial dynamics;
- Different development stages: Germany and Portugal, are developed countries, and the other, China and Mexico, are countries in the developing world, China being a country with a strong potential for economical growth.
- The migration of each country and the type of immigrants found in the United States have different characteristics.

The analysis of the individual features was made by grouping the initial dataset in clusters. This way, it is easier to understand the characteristics of the individuals in the database, for each country, and it is possible to institute a set of representative groups. In order to define a set of representative groups of agents for each country, three clusters of individuals have been created, for each country, using *k-means*, a non-hierarchical multivariate clustering method (Hair, 2009).

For the whole period of analysis (2000 to 2008), three natural clusters were found for every country (forming 3X4=12 clusters in total), each of which having distinct characteristics with respect to the variables that have been selected. This information is important to characterize the agents to simulate starting in the year 2000. In Table 2, the main results of the clustering process are described for each country.

[2] IPUMS is the Integrated Public Use Microdata Series of the Minnesota Population Center (Ruggles et al., 2009).

Table 2: Results of the clustering process for the creation of representative agents

	Cluster	Age(y)	HH dimension (p)	HH Income (r)	% workers (w)
GERMANY					
1st and 3th quartiles	1	31 a 50	2 a 4	134 a 174	-
	2	12 a 34	3*a 5	35,2 a 71,4	-
	3	37 a 70	1 a 2	24,3 a 64,5	-
$N(\bar{x};\sigma_x)$	1	N(41;17)	-	N(168;59)	-
	2	N(23;13)	-	N(54,3;27,9)	-
	3	N(54;19)	-	N(45,7;29,0)	-
(%) Representative	1	-	83,7	-	68,9
	2	-	83	-	39,4
	3	-	82,4	-	55
CHINA					
1st and 3th quartiles	1	36 a 55	2*a 4	139 a 188	-
	2	20 a 44	3*a 6	34,3 a 82,7	-
	3	37 a 67	1*a 3	16,2 a 74,9	-
$N(\bar{x};\sigma_x)$	1	N(46;12)	-	N(178;73,1)	-
	2	N(31;17)	-	N(59,8;29,6)	-
	3	N(53;18)	-	N(48,1;32,5)	-
(%) Representative	1	-	76,2	-	85,9
	2	-	82,4	-	51,4
	3	-	87,4	-	56,4
MEXICO					
1st and 3th quartiles	1	27 a 50	1*a 5	128 a 170	-
	2	20 a 37	3*a 7**	19,7 a 50,9	-
	3	32 a 56	1 a 3	16,0 a 43,0	-
$N(\bar{x};\sigma_x)$	1	N(38;15)	-	N(160;50)	-
	2	N(29;13)	-	N(38,2;25,2)	-
	3	N(45;17)	-	N(32,9;22,7)	-
(%) Representative	1	-	78,8	-	73,4
	2	-	86,7	-	53,2
	3	-	84,9	-	61,8
PORTUGAL					
1st and 3th quartiles	1	47 a 57	4 a 5	147 a 289	-
	2	22 a 40	4 a 5**	31,2 a 60,0	-
	3	41 a 63	2 a 4**	17,4 a 54,6	-
$N(\bar{x};\sigma_x)$	1	N(52;8)	-	N(211;73)	-
	2	N(32;11)	-	N(45,6;22,4)	-
	3	N(52;16)	-	N(39,2;26,7)	-
(%) Representative	1	-	76,4	-	64,7
	2	-	82,3	-	69,4
	3	-	80,97	-	54,1

According to Table 2, it is possible to conclude that the educational level in all groups is equal or grater than 2. The dimension of the Chinese community in the U.S. in 2000, is lower than the German community. However these two communities have very similar characteristics. The third cluster is the one with the highest percentage of individuals represented. The second cluster represents the group of younger individuals, probably young couples, households with higher than individuals of other clusters, but with relatively low incomes. The percentage of workers is the lowest compared to the other two clusters.

3.3.2. Gravitational model

In the gravitational model of Newton, every massive particle in the universe attracts every other massive particle with a force which is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. Likewise, the essence of the gravitational model applied to migration flows is that the act of migrating may be influenced by the attraction exerted by the social network. Boyd (1989) stresses the importance of family relationships and friendships within a community and in social networks. In our work, immigrant communities in the U.S. have been created on the basis of the dimension of the social networks. We aim at measuring the force that these communities have to attract new migrants from the country of origin to the U.S., as well as to repel those who are already in the U.S.

In order to adapt the Newton's model to this context, some forces have been defined. Fm in Equation (1) determines the force of migration, i.e., the level of attraction of the social network in the country of destiny to an agent in the country of origin.

$$F_m = G\frac{M_N m_a}{d^2} \tag{1}$$

M_N is mass of the social network, ma is the mass of the agent and d is the average Euclidean distance between the agent and the other agents in the social network. Here, G (the constant measuring the gravitational constant in the gravitational model of Newton) is equal as 1. The masses M_N and m_a involved in the model represented in Equation (1) are now substituted by other values that correspond to more realistic measures in this context: M_N is replaced by a quantity that takes into account the values of the social network: the average of the income (r) of the households, the median of the number of persons in the household (p_N) the average age of the agent (y) and the situation in workforce (w). The mass m_a is replaced by similar measures that are computed at the level of the individual agent (see Equation 2)

$$F_m = G\frac{\left(\frac{\log(\bar{r}_N)\times \text{median}(p_N)}{(\bar{y}_N/10)} + \%w_N\right)\times\left(\frac{\log(r_a)\times p_a}{(y_a/10)} + w_a\right)}{d_a^2} \tag{2}$$

We have computed a level of the propensity to migrate (P_M) that considers the difference between the GDP per capita in the U.S. (f_{EUA}) and the GDP per capita in the country of origin (fo). The value h is the geographical distance between the two countries (see Equation 3).

$$P_M = \frac{f_{EUA} - f_O}{(h/100)} \tag{3}$$

The final model that includes all the previous aspects can be formulated in Equation 3 (in which we have included the cost of migration, C_M) and may be interpreted as follows: the chance of an agent to migrate, defined as M_a, is a function of the propensity to migrate (P_M), associated with each nationality, and the gravitational force of migration (F_M). These forces, P_M and F_M are summed up in this equation, since they can be taught as complementary forces. In this sense, the higher the propensity to migrate, the greater the chance that the agent has to migrate (or to leave the U.S.).

$$M_a = C_M \times (F_m + P_M) \tag{4}$$

3.3.3. The simulation

Some initial considerations have to be made before writing the final simulation algorithm: for the sake of simplification, we have assumed that social networks connecting elements of a particular country can only attract elements of that country. For example, a social network of Mexican persons in the U.S is not able to attract Portuguese persons from their country of origin.

Another important issue is the number of representative agents that were created according to Table 2. A distribution of agents proportional to the number of immigrants was chosen. Some attention was paid to the hardware capacity, as the software that was used to perform the simulation needs a considerable quantity of resources. Therefore, the number of agents created for the starting of the simulation in 2000 for the Germans, Chinese, Mexican and Portuguese was, respectively 459, 449, 404 and 348. Each agent is associated to a identification label, so that it is possible to trace its evolution. At the moment of the creation of an agent, a set of variables is immediately defined according to Table 1: age, educational level, household income, the number of persons in the household and working situation (working/not working). For the year 2001 (and subsequent), new agents were created following a Normal distribution $N(\bar{n}, \sigma_n)$ where $\bar{n}$ is the expected number of agents and σ_n is the corresponding standard deviation. A different Normal distribution has been defined for each country. After the creation of the agents, an evolutionary simulation is performed: in each time step, the age of the agent is updated. In the following lines, the main rules used to update the main variables are dscribed:

a. *Age*(y) – if the age in year *t* (y_t)

i. $y_t \leq 94$ then $y_{t+1} = y_t + 1$;

ii. $y_t = 95$ then the agent die.

b. *Educational level* (e) – depends on variable age:

i. If $e_t = 1$ and $1 \leq y_{t+1} \leq 14$, then $e_t = e_{t+1} = 1$;

ii. If $e_t = 1$ e $15 \leq y_{t+1} \leq 18$, então $e_{t+1} = U(1, \min(2, \max_e))$;

iii. If $e_t = 1$ e $19 \leq y_{t+1} \leq 94$, então $e_{t+1} = U(1, \min(2, \max_e))$

iv. If $e_t = 2$ e $19 \leq y_{t+1} \leq 94$, então $e_{t+1} = U(2, \min(3, \max_e))$;

c. Income (r) varies in [2;+∞[, and depends on the inflation rate of USA (equal to 3 %). In t+1, the value of r is given by: $r_{t+1}=r_t+[U(-1,1)x0,03]$.

d. Labour status (w) depends on variable age:

i. If $1 \leq y_{t+1} \leq 15$ then $w_{t+1} = 0$;

ii. If $16 \leq y_{t+1} \leq 94$ then w_{t+1} = Bernoulli(k), being *k* the fraction w of working people in USA.

e. Number of individuals in the household (p):

i. If $p_t = 1$, then $p_{t+1} = p_t + U(0,1)$;

ii. If $p_t = 15$, then $p_{t+1} = p_t + U(-1, 0)$;

iii. If $2 \leq p_{t+1} \leq 14$ then $p_{t+1} = p_t + U(-1,1)$;

f. The Number of individuals in the agents' social network (s) varies according to the value of M_N in the previous year.

The software REPAST - Recursive Porous Agent Simulation Toolkit, (North et al., 2007), with Java implementations has been used. One of the advantages of this software is the ability to create multi-agent systems coping with social behaviours. In order to represent and analyse the structure of social network structure that emerged from the simulation, we have used Pajek (Pakek, 2010).

4. RESULTS AND DISCUSSION

The main aspects explored in this section are the stability of the results, the possibility of validating the outcomes and the analysis of the social networks. A M_L (Migration Level) was defined exogenously in order to define a scenario of migration. For each agent, the value M_a defined in Equation (4) is computed and compared to M_L. If ML is greater than the value M_a, then the agent remains in the country of origin. Otherwise, the agent will migrate or stay in U.S. We assumed that three different levels of ML may occur (low, medium and high). These values are defined as 1,5, 4,0 and 5,0 respectively, corresponding to the three different scenarios of migration. In Scenario I, M_L=1,5 is low, and agents tend to migrate to the

U.S. In Scenarios, II and III, agents are more predisposed to stay in the country of origin, or else to leave the U.S. in the case they are already out of their country of origin.

4.1. Stability

The stability of the model was analyzed for the three scenarios and for all countries. Each scenario was repeated 15 times. As an example, Table 2 illustrates the results of the main variables for Scenario I (and for Germans). We can state that there is a great stability of the values, confirmed by the variation coefficient (the standard deviation being always smaller than 5% of the mean). The values for the other countries are even more stable than these.

Table 3: Average values and Variability (values in %) of the main variables for German population

Variables	2000	2001	2002	2003	2004	2005	2006	2007	2008
Household Dimension (p)	2,40±0,03 (1,4%)	2,73±0,07 (2,5%)	2,90±0,06 (2,2%)	3,01±0,06 (1,9%)	3,11±0,06 (1,8%)	3,17±0,04 (1,3%)	3,23±0,05 (1,6%)	3,27±0,05 (1,6%)	3,30±0,05 (1,5%)
Age (y)	43,8±0,7 (1,6%)	39,4±1,1 (2,7%)	38,0±0,8 (2,0%)	37,4±0,8 (2,2%)	37,1±0,6 (1,7%)	37,1±0,6 (1,5%)	37,2±0,6 (1,7%)	37,6±0,6 (1,6%)	38,0±0,6 (1,5%)
Social Network	7,85±0,21 (2,7%)	7,31±0,14 (1,9%)	7,39±0,13 (1,8%)	7,57±0,15 (2,0%)	7,79±0,14 (1,8%)	8,02±0,14 (1,7%)	8,22±0,15 (1,8%)	8,39±0,16 (1,9%)	8,53±0,15 (1,8%)
Houseold Income (r)	65,5±1,5 (2,2%)	61,9±1,6 (2,5%)	61,4±1,7 (2,8%)	61,1±1,7 (2,8%)	61,0±1,7 (2,7%)	61,1±1,8 (2,9%)	61,5±1,8 (2,9%)	61,4±1,7 (2,7%)	61,4±1,5 (2,4%)
% Workers (w)	0,476±0,023 (4,9%)	0,552±0,017 (3,1%)	0,504±0,022 (4,4%)	0,473±0,016 (3,3%)	0,465±0,017 (3,7%)	0,460±0,011 (2,3%)	0,455±0,010 (2,3%)	0,457±0,014 (3,1%)	0,460±0,010 (2,2%)

Note: The values in this table may be interpreted as follows: in the year 2001, for example, the mean values of the household dimension for Germans is 2,73 with a variation between -0,07 and +0,07 corresponding to a percentage of standard deviation of 2,5% (variation coefficient).

4.2. Validation

Validation is a very important step in the simulation, since we need to measure the adequacy of the model. Fagiolo et al. (2007) and Windrum et al. (2007) suggests several ways of validating computational models, such as historic friendly validation, calibration, etc.. According to Bianchi et al. (2007), the most intuitive form is made by comparing simulated values with real ones. In our case, a Wilcoxon test (Conover, 1999) was applied. In this test, a set of hypotheses (Ho^{I} to Ho^{iv}) is defined as: "the medians of the variables in the paired populations r and s are equal: Ho^{I}: $y_s=y_r$; Ho^{II}:$r_s=r_r$; Ho^{III}:$p_s=p_r$; Ho^{IV}: $w_s=w_r$" (where r denotes the real values and s denotes the simulated ones and the indexes y, r, p and w stand for the variables of the problem – see Table1). If we reject the null hypotheses, corresponding to a p-value that is lower than 0,05, then it means that there is statistical evidence that differences between real data and simulated data exist (with 5% of significance). Each set of hypotheses is applied to a country. Table 4 contains the results of the Wilcoxon bilateral tests for the variables and countries in which the null hypotheses are not rejected (there is no statistical evidence that differences between real data and simulated data exist (with 5% of significance).

Table 3: Results for the Wilcoxon Bilateral tests for which real data is similar to simulated data

Country of origin	Variable	Scenario	Z*	*p-value*
Germany	Working situation (w)	I	-1,718	0,0858
China	HH Income (r)	I	-1,362	0,1731
	Working situation (w)	I	-0,889	0,3743
Mexico	HH Income (r)	I	-1,362	0,1731
	Hh Income (r)	II	-1,244	0,2135
	Hh Income (r)	III	-1,125	0,2604

4.3. Analysis of social networks

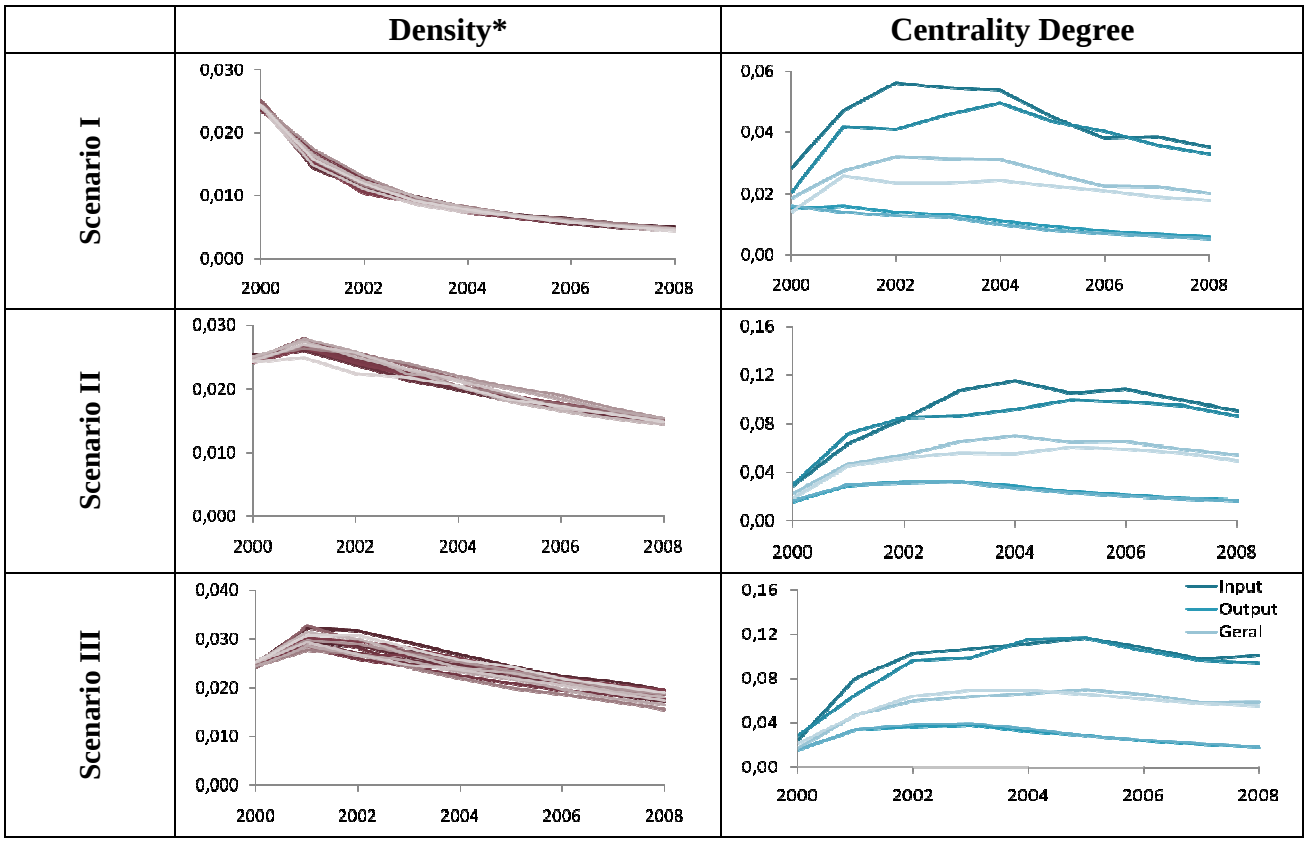
At this point we are interested in analysing the social networks that emerged from the interaction among individuals and to verify in what extent the shape of these networks are important to the international migration flows. In particular, we aim at verifying the impact of network density on the quantity of the migratory flow and to verify if the degree of centralization of the network influences the quantity of the migratory flow.

We focused our attention in specific scenarios: scenarios II and III are more realistic than scenario I. The latter is strongly inductive for immigration since the value M_L=1,5 contributes to more exits from the country of origin. So, we base our further analysis in Scenarios II and III. Mexico is the country with higher growth rate in terms of immigration in USA during the period of analysis. As the size of the social network of Mexicans increases, the density decreases, in general. Germans have the lowest values of density of all four countries.

The centralization degree may be analysed in three different perspectives in oriented networks: input centralization degree (taking into account the arcs of entrance towards an agent), output centralization degree (taking into account the arcs of exit from an agent) and general centralization degree (taking into account both input and output degrees). These measures are signs of network centrality. The higher these degree centralization measures, the more central agents are in the social network and therefore the higher influence the agents have in the network. We concluded that the agents that stay in the U.S. (and do not go away to their country of origin), have higher values of M_a (see Equation 4), and have less average distance among the members of their social networks. Networks of these agents are more cohesive. According to the results of the simulation, in 2008, Germans, Chinese and Portuguese have similar values of the degree centralization measures, meaning that agents occupy central positions in networks. Simultaneously, the dimension of the networks continues to grow during the period. Centrality is therefore an important property for the growth of networks.

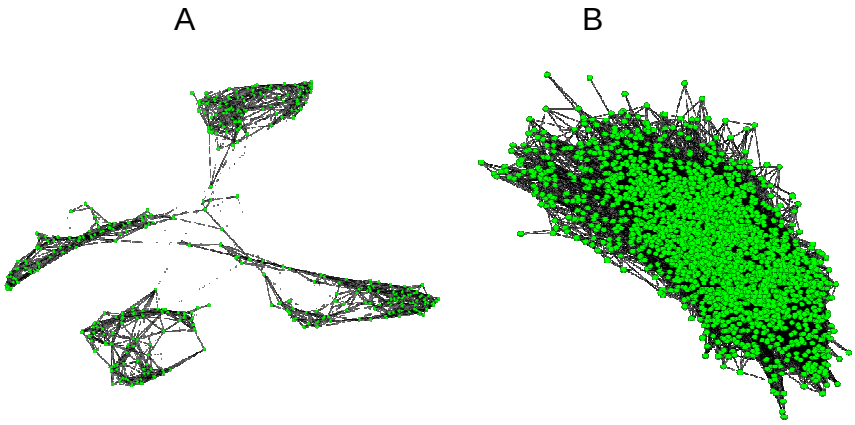
Correlations between the centralizations degrees (input centralization, output centralization, and general centralization) are not consistent according to different variables and it is not possible to take conclusions about a possible statistical association between these measures. We observe that Germans have lower values of the general degree centralization and input degree centralization. Mexicans have lower output degree centralization and higher general degree centralization and input degree centralization. In Fig.3 we observe the evolution of the social network for the Mexicans at the beginning (2000) and at the end (2008) of the simulation. It is possible to verify that the networks are more populated at the final of the simulation and that it is difficult to analyse the groups of individuals that were easier to find at the beginning.

Figure 2: Evolution of two measures of social networks - Density and Centrality Degree - of the Mexicans in the nine-year period of the simulation, according to the three different scenarios



* Each line in the graph of density represents one of the fifteen simulations

Figure 3: Evolution of the social network (Mexicans): 2000 (A) and 2008 (B)



5. CONCLUSIONS AND FURTHER WORK

In this work we use a Multi-Agent System to model the flow of migrants in social networks. Our model is based on the idea that migration flows are influenced by the structure of the social network of the agents. The proposed model of migration uses four countries (Germany, Mexico, Portugal and China) and six variables: age (y), educational level, (e), household income (r), number of people in the household, (p), labour status (working/not working), (w) and number of individuals in the social network of the agent (s). The social network is based on the number of persons in the household. We have used information from the IPUMS, containing data of migration flows to the United States between 2001 and 2008.

Our goal was to verify to what extent the existence of networks with major centrality contributes to favour the increase the number of migrants. We believe that in cases of higher centrality, the information flows easily through the links and nodes of the network, but the center of the network is critical for the transmission of information. This analysis aims at highlighting two important aspects, which can be seen as research Hypotheses: 1) The density of the network influences the quantity of the migratory flow; 2) The degree of centralization of the network influences the quantity of the migratory flow.

A gravitational model has been applied to migration flows in the sense that the act of migrating may be influenced by the attraction exerted by the social network. In our work, immigrant communities in the U.S. have been created on the basis of the dimension of the social networks. We aim at measuring the force that these communities have to attract new migrants from the country of origin to the U.S., as well as to repel those who are already in the U.S.

The stability of the model was analyzed for three different scenarios of migration level (M_L) and for all countries. It is possible to confirm that there is a great stability of the values, confirmed by the variation coefficient. We focused our attention in specific scenarios: scenarios II and III are more realistic than scenario I. The latter is strongly inductive for immigration. We concluded that the agents that stay in the U.S. (and do not go away to their country of origin), have higher values of M_a, and have less average distance among the members of their social networks. Networks of these agents are more cohesive. Centrality is therefore an important property for the growth of networks. In addition, Germans have lower values of the general degree centralization and input degree centralization. Mexicans have lower output degree centralization and higher general degree centralization and input degree centralization.

6. REFERENCES

Axtell, R. (1999), "Why Agents? On the Varied Motivations for Agent Computing in the Social Sciences", Workshop on Agent Simulation: Applications, Models, and Tools, Argonne National Laboratory, Argonne, Illinois, USA.

Bianchi, C., P. Cirillo, et al. (2007), "Validating and Calibrating Agent-Based Models: A Case Study", Computational Economics, Vol. 30, Nº 3, pp. 245-264.

Billari, F. C., & Prskawetzy, A. (2005). Studying population dynamics from the bottom-up: the crucial role of agent-based computational demography. International Union for the Scientific Study of Population - XXV International Population Conference (p. 13). Tours, France: IUSSP

Billari, F. C., Ongaro, F., & Prskawetz, A. (2003). Introduction: Agent-Based Computational Demography. In F. C. Billari, & A. Prskawetz, Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour (p. 210). New York: Physica-Verlag.

Bond, A. H., and L. Grasser (1988), Readings in Distributed Artificial Intelligence, Los Altos, Morgan Kauffmann

Boyd, M. (1989), "Family and Personal Networks in International Migration: Recent Developments and New Agendas", International Migration Review, Vol. 23, Nº 3, pp. 638-670.

Campos, P. (2008), "Organizational Survival and the Emergence of Collaboration Networks: A Multi-Agent Approach", Faculty of Economics, University of Porto, PhD Thesis.

Conover, W. J. (1999), Practical Nonparametric Statistics, 3rd Edition, USA, John Wiley and Sons

Fagiolo, G., A. Moneta, et al. (2007), "A Critical Guide to Empirical Validation of Agent-Based Models in Economics: Methodologies, Procedures, and Open Problems", Computational Economics, Vol. 30, Nº 3, pp. 195-226.

Ferber, J. (1999), Multi-Agent Systems: An Introduction to Artificial Intelligence: Addison-Wesley.

Freeman, L. C. (1977), "Set of Measures of Centrality Based on Betweenness", Sociometry, Vol. 40, Nº 1, pp. 35-41.

Gilbert, N. e K. G. Troitzsch, (1999), Simulation for the Social Scientist: Open University Press.

Hair, J. F., Black., (2009), Multivariate Data Analysis, Prentice Hall

Helmenstein, C., & Yegorov, Y. (2000). The dynamics of migration in the presence of chains. Journal of Economic Dynamics & Control , 24, pp. 307-323.

Hussey, P. S. (2007). International migration patterns of physicians to the United States: A cross-national panel analysis. Health Policy , 84, pp. 298–307.

Lemieux, V. e M. Ouimet (2008), Análise Estrutural Das Redes Sociais, Instituto Piaget, Lisboa

Mckenzie, D., & Rapoport, H. (2007). Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. *Journal of Development Economics , 84*, pp. 1–24.

Mitchell, J. C. (1969), "The Concept and Use of Social Network" in J. C. Mitchell (editor), *Social Networks in Urban Situations*, Manchester, UK, Manchester University Press, pp. 1-50

Neto, F. e E. Mullet (1998), "Decision-Making as Regards Migration: Wage Differential, Job Opportunity, and the Network Effect", Acta Psychologica, Vol. 98, Nº 1, pp. 57-66.

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *PHYSICAL REVIEW E , 69*, p. 026113.

Nooy, W., Mrvar, A., Batagelj, V., (2005) Exploratory Social Network Analysis with Pajek, Cambridge University Press

North, M.J., E. Tatara, N.T. Collier, and J. Ozik, (2007) "Visual Agent-based Model Development with Repast Simphony," Proceedings of the Agent 2007 Conference on Complex Interaction and Social Emergence, Argonne National Laboratory, Argonne, IL USA

Pajek, (2010), "Program for Large Network Analysis", version 1.26", available in http://vlado.fmf.uni-lj.si/pub/networks/pajek/

Ruggles, S., Sobek, M. Alexander, T., Fitch, C., Goeken, R., Kelly Hall, P., King. M., Chad R., (2009) Integrated Public Use Microdata Series: Version 4.0 [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], (IPUMS site: http://usa.ipums.org/usa/)

Shyrock, H. S., J. S. Siegel, et al. (1976), The Methods and Materials of Demography, New York: Academic Press.

Stillwell, J. e P. Congdon (1991), "Migration Modelling:Concepts and Contents", in Migration Models: Macro and Micro Approaches, J. Stillwell e P. Congdon, Vol. 1-16, London and New York: Bellhaven Press.

Wasserman, S. e K. Faust (1994), Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences), Cambridge: Cambridge University Press.

Windrum, P., G. Fagiolo, et al. (2007), "Empirical Validation of Agent-Based Models: Alternatives and Prospects", Jasss-the Journal of Artificial Societies and Social Simulation, Vol. 10, Nº 2

Woodruff, C., & Zenteno, R. (2007). Migration networks and microenterprises in Mexico. Journal of Development Economics , 82, pp. 509–528.

Wooldridge, M. (2002), An Introduction to Multiagent Systems by Michael Wooldridge, Chichester, England: John Wiley & Sons.