Environment and Planning B: Planning and Design 2015, volume 42, pages 4 – 22

doi:10.1068/b39042

International migration: a global complex network

Emmanouil Tranos¶
School of Geography, Earth and Environmental Sciences, University of Birmingham B15 2TT, England; e-mail: e.tranos@bham.ac.uk

Masood Gheasi, Peter Nijkamp¶
Department of Spatial Economics, VU University, De Boelelaan 1105, 1081 HV, Amsterdam; email: m.a.g.gheasi@vu.nl, pnijkamp@feweb.vu.nl

Received 2 April 2012; in revised form 15 November 2012; published online 7 March 2014

Abstract. Migration has become a prominent research theme in geography and regional science and it has been approached from various methodological angles. Nonetheless, a common missing element in most migration studies is the lack of awareness of the overall network topology, which characterizes migration flows. Although gravity models focus on spatial interaction—in this case migration—between pairs of origins and destinations, they do not provide insights into the topology of a migration network. We employ network analysis to address such systemic research questions, in particular: how centralized or dispersed are migration flows and how does this structure evolve over time? And, how is migration activity clustered between specific countries, and if it is clustered, do such patterns change over time? Going a step further than exploratory network analysis, in this paper we estimate international migration models for OECD countries based on a dual approach: gravity models estimated using conventional econometric approaches such as panel data regressions and network-based regression techniques such as multivariate regression quadratic assignment procedures. The empirical results reveal not only the determinants of international migration among OECD countries, but also the value of blending network analysis with more conventional analytic methods.

Keywords: immigration, gravity model, complex networks, community detection, MRQAP

1 Introduction

International migration is becoming an important feature of the global economy. Declines in transportation and communication costs, as well as developments towards fewer movement restrictions, have encouraged the circulation of people across national and international borders. Our main objective here is to study international migration from a network standpoint. Although research interest on migration flows has been growing, the focus of most studies is on the level of the country flow or, in the best case, on the level of country-to-country flows (dyadic level). There are very few exceptions to the above statement, with the work of Maier and Vyborny (2008) being one of them. Our starting point is that international migration flows form a network of connected countries and this could provide the basis of empirical analysis.

To provide a brief introduction, the ideas which underpin this paper derive from the so-called *new science of networks* (Barabási, 2002; Buchanan, 2002; Watts, 2003; 2004), an analytical field of complexity science which has expanded rapidly over the last ten to fifteen years, the main focus of which is large-scale real-world networks and their universal, structural, and statistical properties (Newman, 2003). While the starting point of network science lies in statistical physics and graph theory, strong parallels exist between network analysis and regional science, as traditionally the latter has a strong interest in networks and interregional

¶ Also at Tinbergen Institute, University of Amsterdam.

systems [for a review on spatial complex networks see Barthelemy (2011); and for a discussion on networks and regional science see Reggiani (2009)].

Two different though complementary streams of network analysis have been developed over time. Most network studies are based on stochastic approaches, which assume an underlying probability model, usually following a power law as the main mechanism for the network creation. The main objective in this strand of research is to identify the underlying mechanisms using constructive modeling and simulation techniques. However, this approach includes the risk that the probability model and the underlying statistical mechanisms do not depict precisely the actual world network (Li et al, 2005).

The second strand of research adopts a 'softer' approach and focuses on ex post empirical tests for identifying characteristics of theoretical network models in real-world networks. Such analysis enables researchers to understand the network attributes of the system and then to model those using network or more conventional modeling techniques. The main drawback to this approach is the rather descriptive nature of the analysis.

In this paper we attempt to bridge these two different approaches using a panel dataset on bilateral international migration flows. The data come from the online database for the International Migration Statistics for OECD countries, which contains information on immigrant flows by country of origin and destination, based on the OECD's continuous reporting system on migration (OECD, 2011). In particular, we use here data on yearly immigration flows between thirty-two OECD countries for the period 2000–09.[1] On the basis of these data, we are able to create ten migration networks for each individual year. It is important to indicate that the OECD online database does not report flows below 1000 observations, therefore our analyses are based on flows above this figure.

The first two steps of our analysis fit into the second strand of network analysis, as described above. We initially study the different centrality measures and the derived attributes for the countries in our sample. Then, we explore further the topology of this network by identifying the communities formed by the intensity of the migration flows between the member countries. Finally, on the basis of knowledge gained from the above investigation, we proceed with the modeling exercise, which is based on a dual approach. Firstly, we use standard econometrics, such as panel data regressions, to estimate the determinants of international migration flows among OECD countries. Then, we validate these results with cross-sectional multivariate regression quadratic assignment procedures (MRQAP) models, which utilize the network structure by addressing potential network-dependency issues.

The novelty of this paper lies in the adoption of a network perspective. Migration is a network phenomenon and is characterized by network dependencies (see subsection 5.2). However, the latter are not usually captured by mainstream statistical analysis. This gap in migration analysis is the focal point of this.

The structure of the paper is as follows. In the next section we present some insights from the relevant literature on international migration. Next, in section 3 the different network attributes are explored, and then, in section 4 the different network communities are highlighted. In section 5 we present the applied modeling exercises, and conclude with some remarks and directions for future research.

2 Literature review on international migration

Migration movements can be studied from the perspective of push and pull factors. Push factors, such as poverty, unemployment, conflict, and natural disasters, and pull factors, including employment opportunities, wealth, favorable climate, political stability, and low risk from natural hazards, made millions of people move from their country of origin to other countries

[1] Korea and the Slovak Republic were excluded from the analysis due to missing data.

and even to different continents. In addition, globalization and developments in transport had a great impact on short- and long-range mobility[2] of people (Nijkamp et al, 2011). Long-range mobility can be temporary, or it can lead to permanent settlement. Over the past few decades cross-border migration has become a megatrend of the globalizing economy, to the extent that some people even speak of the 'age of migration' (see Goldin et al, 2011; Nijkamp et al, 2012).

Nowadays, around 3% (more than 200 million people) of the world population were not born in the country in which they live (Özden, 2005). Empirical data show that the majority of OECD countries are final destinations for the largest part of international migration (Gheasi et al, 2011). The foreign-born population in 2006 accounted for about 11.7% of the total population in OECD countries, and this shows a drastic increase in comparison with previous years (OECD, 2011).

Migrants may be considered as a bridge of information between the host country and the country of origin. Therefore, there is a growing body of literature on migration and its related economic impacts. Studies have found a close relationship between immigration and international trade (Girma and Yu, 2002; Gould, 1994; Head and Ries, 1998; Lewer and Van den Berg, 2008; Rauch and Trindade, 2002), migration and international tourism (Fischer, 2007; Gheasi et al, 2011; Williams and Hall, 2002), and migration and foreign direct investment (Aroca and Maloney, 2005; Bhattacharya and Groznik, 2008; Gheasi et al, 2011; Javorcik et al, 2011; Kugler and Rapoport, 2007).

Gravitational models have a long and established tradition in the estimation of migration flows. The use of the gravity model has grown considerably since Tinbergen (1962) and Pöyhönen (1963) were the first to use this model to explain international trade patterns. The gravity model has been recognized for a long time for its consistent empirical success in explaining different types of flow, such as migration, commuting, shopping trips, tourism, and trade. Migration, like other types of flow, can also be driven by attraction forces between the country of origin and the country of destination, which decrease by the cost of the distance between them (Lewer and Van den Berg, 2008). Such a model suggests that the attraction force between two countries depends on labor-income and population-size differences between them.

Studies indicate that demography plays a major role in explaining international migration. The younger the population of a country, the bigger the share of the population that is most likely to emigrate. Various studies (Hatton and Williamson, 1998; 2003; Mayda, 2007) suggest that the share of the origin country's population aged 15–29 years has a significant positive impact on outmigration. Moreover, regarding other covariates used in migration gravity models, a common language and cultural ties between origin and destination can facilitate migrants' integration into the host society. Adsera and Chiswick (2007) found that there is a 9% earnings premium for immigrant men if they come from a country where the language spoken belongs to the same language family group as the destination country. A recent study by Belot and Ederveen (2012) shows that cultural barriers may explain patterns of migration flow between developed countries better than traditional economic variables.

The knowledge gained from the above review will support our modeling endeavor later in this paper. Before that, the next sections will shed light on the structure of the international migration network. Such structural characteristics will also influence our modeling strategy.

3 Network attributes of international migration

The first step of the analysis focuses on the different centrality measures based on international migration flows. Table 1 presents these elements for the year 2000. Firstly, the topology of the migration network is analyzed by using only binary links. Such a binary network is

[2] For example, short-range mobility refers to commuting between work and home, and social visits, while long-range mobility to international migration, and international tourism.

Table 1. Degree centrality measures for migration in 2000**.** Absolute degree centralities are presented here along with the relevant rankings in italics. Normalization occurs by dividing centralities by the population of the destination country for in-degree in column (7) and origin country for out-degree in column (8). Balance, column (9), is weighted in-degree−weighted out-degree.

Country	In-degree (1)		Out-degree (2)		Degree (3)		Weighted in-degree (4)		Weighted out-degree (5)	
Germany	30	*8*	23	*3*	53	*3*	318.649	*1*	77.936	*6*
United States	31	*1*	26	*1*	57	*1*	256.272	*2*	98.430	*4*
Switzerland	25	*15*	19	*13*	44	*15*	50.822	*6*	10.264	*25*
United Kingdom	23	*16*	24	*2*	47	*11*	130.931	*3*	98.693	*3*
Spain	31	*1*	20	*9*	51	*5*	53.189	*4*	24.163	*17*
Belgium	16	*19*	19	*13*	35	*19*	36.943	*8*	12.556	*22*
Australia	30	*8*	15	*25*	45	*14*	53.154	*5*	30.728	*12*
Austria	31	*1*	19	*13*	50	*7*	30.760	*13*	18.859	*18*
Luxembourg	30	*8*	13	*32*	43	*16*	9.154	*18*	0.920	*32*
Netherlands	30	*8*	21	*7*	51	*4*	36.257	*9*	28.240	*15*
Ireland	2	*26*	18	*17*	20	*28*	10.900	*17*	8.997	*26*
Sweden	31	*1*	18	*17*	49	*8*	18.607	*14*	17.064	*19*
Norway	30	*8*	16	*23*	46	*13*	12.465	*16*	11.531	*23*
Japan	2	*26*	20	*9*	22	*24*	30.984	*12*	30.440	*13*
Canada	31	*1*	22	*4*	53	*2*	31.807	*11*	31.580	*11*
Estonia	0	*28*	15	*25*	15	*30*	0	*32*	1.626	*31*
Iceland	0	*28*	14	*29*	14	*32*	0	*31*	2.261	*30*
Slovenia	14	*21*	14	*29*	28	*22*	0.255	*27*	2.734	*29*
Denmark	30	*8*	17	*20*	47	*10*	8.914	*19*	13.240	*21*
Israel	18	*18*	17	*20*	35	*20*	3.793	*23*	8.515	*27*
Chile	0	*28*	14	*29*	14	*31*	0	*30*	6.212	*28*
Finland	31	*1*	15	*25*	46	*12*	3.383	*24*	11.106	*24*
Czech Republic	3	*25*	17	*20*	20	*26*	0.318	*26*	15.164	*20*
Hungary	23	*16*	19	*13*	42	*17*	3.058	*25*	24.254	*16*
Portugal	8	*23*	18	*17*	26	*23*	4.438	*21*	28.528	*14*
Italy	11	*22*	21	*7*	32	*21*	37.091	*7*	62.554	*8*
Greece	0	*28*	20	*9*	20	*27*	0	*29*	35.939	*10*
New Zealand	6	*24*	15	*25*	21	*25*	7.254	*20*	46.307	*9*
Turkey	31	*1*	20	*9*	51	*6*	34.983	*10*	84.794	*5*
France	15	*20*	22	*4*	37	*18*	15.166	*15*	73.561	*7*
Poland	26	*14*	22	*4*	48	*9*	4.423	*22*	106.521	*2*
Mexico	0	*28*	16	*23*	16	*29*	0	*28*	180.253	*1*

represented by an adjacency matrix, the (i, j) element of which is 1 if there is a migration flow from country i to country j in 2000, and 0 otherwise. The in-degree centrality denotes the number of different origin countries for every destination. According to table 1 column (1), a quite diverse group of countries is at the top of this hierarchy: on the one hand North American (US and Canada) and European countries (Austria, Finland, Spain, and Sweden) can be identified and on the other hand Turkey. At the other end of the spectrum, Chile, Estonia, Greece, Iceland, and Mexico do not receive any migration flows. The hierarchy is different, when we focus on the out-degree centrality, table 1 column (2), which represents the number of different destination countries for each origin. At the top of the hierarchy wealthy

Table 1 (continued).

Country		Weighted degree (6)		Normalized weighted in-degree (7)		Normalized weighted out-degree (8)		Balance (9)	
Germany	*6*	396.585	*1*	0.004	*3*	0.001	*28*	240.713	*1*
United States	*4*	354.702	*2*	0.001	*16*	0	*31*	157.842	*2*
Switzerland	*25*	61.086	*14*	0.007	*2*	0.001	*19*	40.558	*3*
United Kingdom	*3*	229.624	*3*	0.002	*10*	0.002	*16*	32.238	*4*
Spain	*17*	77.352	*10*	0.001	*14*	0.001	*29*	29.026	*5*
Belgium	*22*	49.499	*17*	0.004	*5*	0.001	*24*	24.387	*6*
Australia	*12*	83.882	*9*	0.003	*8*	0.002	*17*	22.426	*7*
Austria	*18*	49.619	*16*	0.004	*4*	0.002	*10*	11.901	*8*
Luxembourg	*32*	10.074	*28*	0.021	*1*	0.002	*12*	8.234	*9*
Netherlands	*15*	64.497	*11*	0.002	*9*	0.002	*15*	8.017	*10*
Ireland	*26*	19.897	*24*	0.003	*6*	0.002	*9*	1.903	*11*
Sweden	*19*	35.671	*19*	0.002	*11*	0.002	*13*	1.543	*12*
Norway	*23*	23.996	*22*	0.003	*7*	0.003	*6*	0.934	*13*
Japan	*13*	61.424	*13*	0	*24*	0	*32*	0.544	*14*
Canada	*11*	63.387	*12*	0.001	*15*	0.001	*27*	0.227	*15*
Estonia	*31*	1.626	*32*	0	*32*	0.001	*25*	−1.626	*16*
Iceland	*30*	2.261	*31*	0	*31*	0.008	*2*	−2.261	*17*
Slovenia	*29*	2.989	*30*	0	*25*	0.001	*20*	−2.479	*18*
Denmark	*21*	22.154	*23*	0.002	*13*	0.002	*7*	−4.326	*19*
Israel	*27*	12.308	*27*	0.001	*19*	0.001	*21*	−4.722	*20*
Chile	*28*	6.212	*29*	0	*30*	0	*30*	−6.212	*21*
Finland	*24*	14.489	*26*	0.001	*17*	0.002	*11*	−7.723	*22*
Czech Republic	*20*	15.482	*25*	0	*27*	0.001	*18*	−14.846	*23*
Hungary	*16*	27.312	*21*	0	*22*	0.002	*8*	−21.196	*24*
Portugal	*14*	32.966	*20*	0	*21*	0.003	*4*	−24.09	*25*
Italy	*8*	99.645	*7*	0.001	*18*	0.001	*26*	−25.463	*26*
Greece	*10*	35.939	*18*	0	*29*	0.003	*3*	−35.939	*27*
New Zealand	*9*	53.561	*15*	0.002	*12*	0.012	*1*	−39.053	*28*
Turkey	*5*	119.777	*5*	0.001	*20*	0.001	*22*	−49.811	*29*
France	*7*	88.727	*8*	0	*23*	0.001	*23*	−58.395	*30*
Poland	*2*	110.944	*6*	0	*26*	0.003	*5*	−102.098	*31*
Mexico	*1*	180.253	*4*	0	*28*	0.002	*14*	−180.253	*32*

countries such as the UK, the US, and Canada can be found along with Poland. This measure can be approached as a population mobility indication. For instance, the various locations of the British diaspora (Bridge and Fedorowich, 2003) become apparent as well as the number of different destinations to which Americans migrate. The latter, though, might indicate a scalar issue related to the emigration volume because of the origin country's population size. In total, degree centrality, table 1 column (3), which is the sum of the in-degree and out-degree centrality measures, can be understood as an indication of a country's cosmopolitan and extroversive character.

The picture is somewhat different, when migration flows are introduced. In this case, the (i, j) element of the adjacency matrix represent the number of migrants migrated from country i to country j during one year. The distribution of migrants across the OECD countries is very unequal: almost 60% of all the migration flows from OECD countries end up in Germany, the US and the UK, resulting in a Gini coefficient of 0.71. Regarding the weighted out-degree centrality, table 1 column (5), 24% of all migration flows originates from Mexico and Poland. Concerning the former, the vast majority of Mexican emigrants targets the US (96% of all Mexican emigrants in 2000), followed by Spain, Germany, and Canada. Poland, on the other hand, has a more balanced profile of emigrant destinations, with the neighboring Germany being the main destination (70% of all Polish emmigrants). Furthermore, countries such as the US and the UK are also at the top of this hierarchy, indicating their extroversive and mobile character, but also Turkey, which is a well-known emigration country (Gibney and Hansen, 2005). 59% of Turkish emigrants ended up in Germany in 2000. In total, according to the weighted degree centrality (sum of weighted in-degree and out-degree), an interesting division can be observed in the first six places of the most central countries: Germany, the US, and the UK are the most central ones, mostly due to their attraction as destinations, followed by Mexico, Turkey, and Poland, the high weighted degree centrality of which is caused by their intensive emigration.

The normalization of the above centralities by the population of the destination country, in-degree in column 7, and origin country, out-degree in column 8, reveals new results. As can be seen from table 1, smaller countries such as Luxemburg and Switzerland are at the top of the hierarchy as they receive significant migration inflows relative to their population. The only country which is in the highest tier using both the absolute and the relative weighted in-degree centrality is Germany. Despite its large population, Germany still receives a great inflow of migrants even in relative terms, while this is not the case for the UK and the US. Similarly, countries such as New Zealand, Iceland, Greece, and Portugal are characterized by high migration outflows compared with their resident populations. These countries lose a significant part of their work force through emigration, and this is not replaced by in-migration. This can be seen in the last column of table 1, where the balance (difference) between in-degree and out-degree centrality is presented. Indeed, countries such as Mexico, Poland, France, and Turkey have a negative balance because of migration, as they lost 49 000 to 180 000 people in 2000. At the other end of the spectrum, Germany and the US are by far the net gainers in terms of in/out-migration.

Analyzing the same metrics for 2009, some interesting changes can be observed.(3) In total, the 2009 migration network is denser than the 2000 one (from 0.594 to 0.700).(4) Also, interesting realignments are observed in the rankings, such as the fall in the UK in-degree centrality, as in 2009 immigrants from only ten countries entered the UK compared with twenty-three in 2000. This might reflect the migration policy change in the UK during this period, which probably caused lower flows of immigrants.(5) Regarding changes in out-degree centrality, Ireland is the main example as it is placed fifth in the standardized out-degree centrality ranking in 2009, four positions higher than in 2000. This increase in the Irish out-migration could be explained by the financial crisis and its impact on the Irish economy.

4 International migration communities

In this section we focus on uncovering the communities that different OECD countries form in the migration network. Before presenting the results of the analysis, the distinction between community detection and cluster analysis needs to be highlighted. The latter refers to multivariate methods aiming to reorganize observations into homogeneous groups known

(3) The centrality measures for 2009 can be provided upon request.

(4) Network density refers to the number of edges in a network divided by the number of all possible edges. The case of nonplanar networks is defined as $\tilde{a} = E/0.5V(V-1)$ (Taaffe et al, 1996, page 254), where E denotes the number of edges present in a network and V the number of nodes.

(5) OECD dataset does not report below 1000 observations. Lack of data can also cause the disappearance of countries in the UK in-degree centrality.

as clusters (Aldenderfer and Blashfield, 1984). Nonetheless, such methods have been developed targeting conventional datasets and not network structures, since the emphasis in the latter is not on the behaviour of single observations but on the information about who is connected to whom (Latora et al, 2003). In such a context, the creation of homogeneous clusters takes a different meaning. Instead of creating groups of observations which share the same characteristics, clustering in a network context, which is known as *community detection*, considers the network structure. The main idea is to identify clusters of nodes with dense connections inside the clusters, but not between the clusters (Blondel et al, 2008). While the focus of the community detection lies on the ties between nodes, a conventional cluster analysis method would focus on the nodes attributes neglecting the network topology.

Such an exercise can provide useful insights into the complex structure of the international migration network (figure 1). Community detection will highlight these clusters of countries, which are characterized by strong bilateral ties. Such knowledge can be used as a first step towards understanding and explaining the push and pull factors behind international migration.

Various methods have been suggested for community detection, including, for example, the work of Newman and Girvan (2004), Pons and Latapy (2006), and Clauset et al (2004), with the algorithm developed by Blondel et al (2008) being the most widely used. This algorithm, which is known as the *Louvain method*, aims to maximize *modularity* in a network. This is an indication of the quality of the derived communities, measuring the density of the links inside the community compared with those outside the community (Blondel et al, 2008). This algorithm is also able to cope with weighted networks. For the implementation of the Louvain method, the Pajek(6) software was utilized.

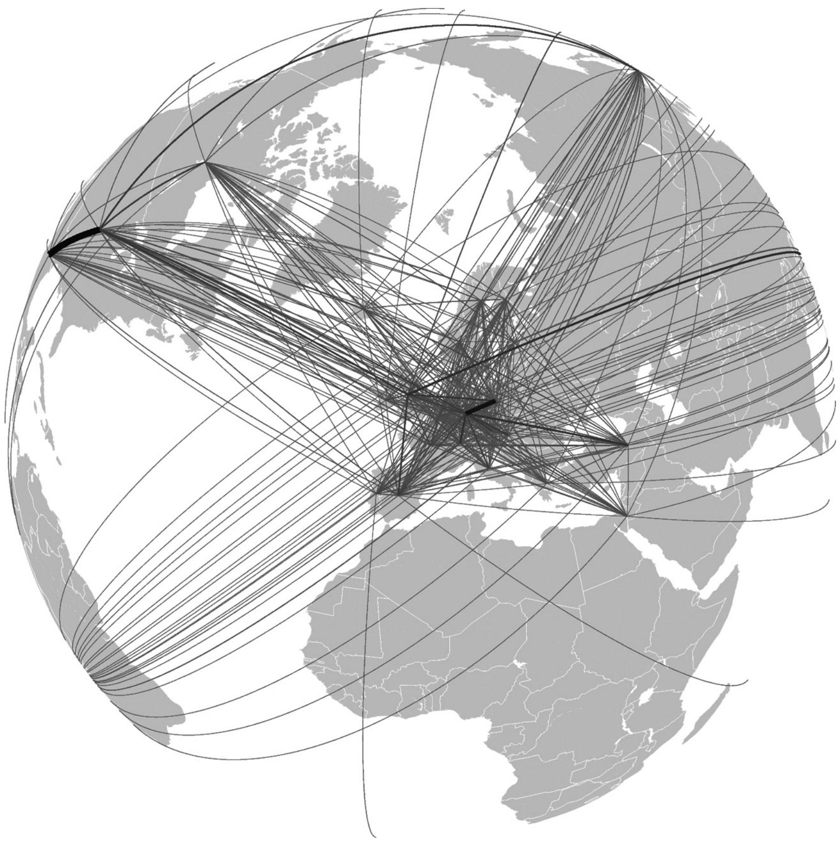

Figure 1. International migration network, 2009.

(6) For more details for the software see http://pajek.imfm.si/doku.php?id=start

The outcome of this analysis reveals familiar structures.[(7)] Firstly, the most robust community over time is the US and countries tightly related with the US, including Canada, Mexico, Japan, and Israel. From these four countries, only Japan has more migrant inflows from the US than outflows to the US, something which might be an indication of return migration. The other three countries have substantially more outflows to the US than inflows from the US, on a yearly basis. Canada and Mexico are adjacent to the US and, in addition, the US hosts the biggest Jewish population reflecting the cultural ties with Israel.

Secondly, mobility among Scandinavian countries is intensive enough to result in another stable community over time. For most of the years of the study period, Denmark, Finland, Sweden, Estonia, Norway, and Iceland are clustered together denoting the strong cultural ties between Northern European countries. Thirdly, former members of the British Empire form a community, the configuration of which does not remain constant over time, as other countries such as Spain, Chile, and also France join it. Nonetheless, the clustering of the UK, Australia, Ireland, and New Zealand highlights the effect of postcolonial and Commonwealth ties in the formation of the migration network. Finally, for some years of the study period, central European countries such as Belgium, France, Luxemburg, the Netherlands, and Switzerland are also clustered together with Portugal, Spain, and some times France and Italy, highlighting the ease of migration inside the European Union. Although modularity for these communities is relatively low and lies between 0.28 (2007) and 0.34 (2001),[(8)] the stability of these communities over time increases the importance of our findings.

In total, the above analysis reveals interesting clustering patterns in the migration network. Countries form fairly robust communities over time, revealing the impact of various factors in the formation of migration networks.

5 Modeling migration flows

Using the knowledge gained from the above analysis, we aim to build models explaining international migration flows among OECD countries. Our starting point is the generalized gravity model:

$$M_{ij} = \frac{A(m_i m_j)}{D_{ij}^b} . \tag{1}$$

Following the Newtonian equation, migration flows (M_{ij}) originating in country i and ending in country j are related to the size m of countries i and j and the distance D between countries i and j. A is a proportionality constant. Following Zhou's (2011) approach, a two-level research methodology is adopted. Firstly, panel data specifications are adopted for conventional econometric analysis to take advantage of the temporal dimension of the dataset. Then the focus turns on the network structure of international migration with the use of repeated MRQAP regressions over time.

5.1 Panel data approach

After the relevant log–log transformations, equation (1) can be represented as a linear model. More specifically, the empirical model we are estimating is the following:

$$\begin{aligned} \ln(M_{ijt}) = \beta_0 &+ \beta_1 \ln(\text{GDPpc}_{it}) + \beta_2 \ln(\text{GDPpc}_{jt}) + \beta_3 \ln(\text{edu}_{it}) + \beta_4 \ln(\text{edu}_{jt}) \\ &+ \beta_5 \ln(\text{pop1529\_share}_{it}) + \beta_6 \ln(\text{D}_{ij}) + \beta_7 (\text{border}_{ij}) + \beta_8 (\text{colony}_{ij}) \\ &+ \beta_9 (\text{language}_{ij}) + \beta_{10} (\text{region}_{ij}) + \varepsilon_{ijt} , \end{aligned} \tag{2}$$

(7) A detailed table of the detected communities can be provided upon request.

(8) A discussion on the quantification of the community variation can be found in Expert et al (2011).

where M_{ijt} represents the migration flows from country *i* to country *j* in year *t*. GDPpc denotes gross domestic product (GDP) per capita in countries *i* and *j* in year *t* as purchasing power parities at constant prices; edu_{it} and edu_{jt} represent the number of graduates from tertiary institutions in countries *i* and *j*, respectively, in year *t*; $\text{pop1529\_share}_{it}$ denotes the share of population aged between 15–29 years in origin country *i* in year *t* following Mayda (2007); *border* is a binary variable which takes the value of 1 when countries *i* and *j* share a common border; *colony* is also a binary variable which takes the value of 1 when countries *i* and *j* were part of the same empire; *language* takes the value of 1 when countries *i* and *j* have the same official language; and *region* denotes that countries *i* and *j* are part of the same geographic region (eg, The Americas, Asia and Pacific, Scandinavia, and rest of Europe).

The main characteristic of the above is the panel structure. Apart from the cross-sectional dimension, the temporal dimension *t*, which represents the ten-year study period, is also addressed here. Panel data specifications come with advantages. Firstly, panel data improve researchers' ability to control for missing or unobserved variables (Hsiao, 2003). Such an omitted-variable bias as a result of unobserved heterogeneity is a common problem in cross-section models. In addition, potential selection bias in migration flows because of missing data can be addressed more efficiently with panel data. In a nutshell, a panel data specification reduces the risk of obtaining biased estimators (Baltagi, 2001).

While panel data introduce methodological gains, there are also shortcomings that need to be addressed. According to literature (Wooldridge, 2003), the most widely used panel data models are based on either *fixed effects* (FE) or *random effects* (RE). As our main aim is to estimate the impact of the different variables on migration flows, it is preferable to use an RE model rather than an FE model, as, because of the inherent first differentiation process, the latter will result in the elimination of the time-invariant explanatory variables which are vital in our analysis (eg, Brun et al, 2005; Etzo, 2011).

Different specifications are tested here in order to estimate equation (2) and are presented in table 2. Firstly, the RE model is estimated without and with country origin and destination effect effects (regressions 1 and 2, respectively, in table 2). The latter can be useful to address unobserved country-specific effects such as the different migration policies among countries. The results in both cases are similar. Distance has a significant negative impact on the intensity of the migration flows, and the existence of a common border between origin and destination countries also has a positive impact. The above reflects the inherent cost in migrating between remote countries and on the other hand the easiness in migrating between adjacent countries. In addition, cultural proximity in terms of common language and post-colonial ties also has a positive impact. The ability to speak the same mother language is an asset for potential immigrants and the same applies to cultural similarity, both of which come as a consequence of a common colonial past. In regards to the pull and push factors, which represent the masses of the Newtonian formula, interesting impacts can be identified. The GDP per capita of the origin country does not have significant impact as a push factor. Of course, it needs to be highlighted here that our analysis focuses on the OECD countries, so countries with very low GDP per capita are excluded from the analysis. However, GDP per capita appears to be a significant pull factor as the GDP per capita of the destination country has a significant positive impact on migration flows. A significant push factor is the share of the young population (15–29 years) in the origin country. This part of a population represents the pool of potential emigrants from the origin countries. In addition, the effect of the education level is also tested here. As expected, when the origin and destination effects are not included in the model, the education level has a positive pull and push effect. However, when the origin and destination effects are included in the analysis, the impact is negative for both cases, indicating a nonstable effect.

Table 2. Panel data regressions on migration flows (ln). Standard errors are shown in parentheses. See text for descriptions of regressions and variables.

Variable	1	2	3	4
D_{ij} (ln)	−0.427	−0.649	−0.439	−0.648
	(0.044)***	(0.059)***	(0.042)***	(0.056)***
D_{ijt} (ln)				
$border_{ij}$	1.082	0.445	1.044	0.447
	(0.190)***	(0.153)***	(0.183)***	(0.145)***
$border_{ijt}$				
$language_{ij}$	0.749	0.631	0.711	0.625
	(0.172)***	(0.142)***	(0.166)***	(0.135)***
$language_{ijt}$				
$colony_{ij}$	1.251	0.947	1.230	0.950
	(0.273)***	(0.215)***	(0.263)***	(0.203)***
$colony_{ijt}$				
$region_{ij}$	−0.004	−0.012	−0.002	−0.008
	(0.015)	(0.014)	(0.013)	(0.013)
$region_{ijt}$				
$GDPpc_i$ (origin, ln)	−0.068	0.150	−0.082	−0.164
	(0.105)	(0.170)	(0.108)	(0.205)
$GDPpc_{it}$ (origin, ln)				
$GDPpc_j$ (destination, ln)	1.301	0.893	1.436	1.152
	(0.093)***	(0.180)***	(0.094)***	(0.221)***
$GDPpc_{jt}$ (destination, ln)				
$pop1529share_i$ (origin)	1.143	3.428	0.313	0.831
	(0.642)*	(1.266)***	(0.657)	(1.642)
$pop1529share_{it}$ (origin)				
edu_i (origin, ln)	0.471	−0.180	0.511	−0.090
	(0.025)***	(0.053)***	(0.024)***	(0.056)
edu_{it} (origin, ln)				
edu_j (destination, ln)	0.396	−0.092	0.431	0.012
	(0.025)***	(0.052)*	(0.024)***	(0.052)
edu_{jt} (destination, ln)				
yearly effect	−0.016	0.051	−0.023	0.038
	(0.004)***	(0.006)***	(0.004)***	(0.006)***
constant	−20.571	−7.470	−22.327	−12.331
	(1.549)***	(2.321)***	(1.588)***	(2.618)***
Origin and destination effects		yes		yes
Wooldridge serial correlation test	17.685***	17.685***		
Observations	5046	5046	5046	5046
Number of groups	762	762	762	762

* significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2 (continued).

Variable	5	6	7
D_{ij} (ln)			
D_{ijt} (ln)		−0.016	−0.030
		(0.003)***	(0.003)***
$border_{ij}$			
$border_{ijt}$		−0.021	−0.005
		(0.010)**	−0.013
$language_{ij}$			
$language_{ijt}$		−0.009	0.005
		(0.009)	(0.012)
$colony_{ij}$			
$colony_{ijt}$		0.037	0.064
		(0.014)***	(0.019)***
$region_{ij}$			
$region_{ijt}$		−0.003	−0.002
		(0.002)	(0.002)
$GDPpc_{i}$ (origin, ln)	−0.140		
	(0.204)		
$GDPpc_{it}$ (origin, ln)		−0.038	−0.004
		(0.008)***	(0.010)
$GDPpc_{j}$ (destination, ln)	1.170		
	(0.220)***		
$GDPpc_{jt}$ (destination, ln)		0.029	0.088
		(0.007)***	(0.008)***
$pop1529share_{i}$ (origin)	0.461		
	(1.638)		
$pop1529share_{it}$ (origin)		−0.166	−0.059
		(0.043)***	−0.057
edu_{i} (origin, ln)	−0.095		
	(0.056)*		
edu_{it} (origin, ln)		0.018	0.034
		(0.002)***	(0.002)***
edu_{j} (destination, ln)	0.010		
	(0.051)		
edu_{jt} (destination, ln)		0.001	0.019
		(0.002)	(0.002)***
yearly effect	0.037	0.088	−1.197
	(0.006)***	(0.128)	(0.157)***
constant	−16.428	−1.471	−1.274
	(2.437)***	(0.053)***	(0.046)***
Origin anddestination effects	yes		
Wooldridge serial correlation test			
Observations	5046	5046	5046
Number of groups	762	762	762

The results shown in table 2 provide a good overall explanation of international migration. However, going a step further, we should also explore the possibility that repeated observations over time violate the assumption of independent errors. We know that serial correlation in panel models results in biased standard errors and less-efficient results. Using the Wooldridge (2002) test for autocorrelation in panel data, implemented by Drucker (2003), first-order autocorrelation is indeed an issue in our data. In order to address it, we follow Zhou (2011) and use the methods derived by Baltagi and Wu (1999). The RE model is estimated with AR(1) correction using the xtregar module in Stata (Stata Corporation, 2007). The results of this model are presented in regressions 3 and 4 in table 2. The difference between the last two specifications is that the latter also includes origin and destination effects. Both these regressions validate, in general, the previous results. The only exemptions are the pull effect of the share of young population, which stops being significant, and the impact of education, which is only positive and significant when the origin and destination effect are excluded from the model.

The next column in table 2 (regression 5), provides another robustness test for the impact of the push and pull effects. In this case, instead of using country-specific effects, country-pair effects are introduced and consequently the variables reflecting bilateral predictors (distance, language, border, colony, and region) are excluded from the analysis as their impact is already included in the country-pair effects (Mayda, 2007). This specification highlights the importance of the push effect of the GDP per capita, which is highly significant and positive. Other than this, only the education level of the origin country is marginally significant, with a negative sign.

In order to examine how the effects change over time, a year-by-covariate interaction term for each covariate is introduced (regressions 6 and 7 in table 2) (Zhou, 2011). Together with the linear year term, these variables are used to estimate migration flows using both RE and the xtregar model. Firstly, we can see that the negative effect of distance becomes more important over time. However, this is not the case for the border effect as its interaction term appears to be negative and significant according to the RE model. However, this stops happening when we correct for serial autocorrelation as in this case the interaction term has no significant impact. The same nonsignificant impact is detected for the common language between origin and destination country. This can be justified by the increasing use of English due to the globalization process. However, the importance of cultural similarities because of a common colonial past between origin and destination country increases over time. Not surprisingly, no significant coefficients were estimated for the common geographic variable region. The pull effect of the GDP per capita of the destination country increases over time, but this is not the case for the push effect of the share of young population in the origin country, which decreases during our study period. Finally, the positive signs of the education interaction terms are difficult to interpret as the impact of the education level was not clear in the previous regressions.

Next, potential endogeneity issues have to be addressed. Endogeneity might arise in our case from reverse causality: although we test the impact of GDP per capita in the destination country as a pull factor, prosperity level might also be affected by the inflows of immigrants. In order to—at least indirectly—address this issue, we present in table 3 the basic models (regressions 3 and 4 in table 2) using lagged regressors. This ‘poor man’s exogeneity’ approach implies that the past years prosperity level of the destination country or the past year’s share of young people in the origin country are not affected by the subsequent year’s migration flows. Although it would be interesting to test this at a later stage with an IV (instrumental variable) approach, the results of this exercise presented in regressions 1 and 2 of table 3 are almost identical with the previous specification without the lagged variables, verifying the previous discussion.

Furthermore, in table 3 we introduce a dynamic dimension to our analysis as the stock of immigrants ($stock95_j$) in destination country in 1995 is used as an explanatory variable. The underlying assumption is that countries which are well-known destinations will continue to attract migration flows. Migrants mostly migrate to countries where their compatriots have already established a network 'beaten path'. Through the established network, previous migrants transfer their knowledge and experience to the newcomers and make it easier for them to find jobs, accommodation, and even to deal with bureaucratic obstacles. This is verified in regressions 3 and 4 of table 3 which display the significant and positive impact of the stock of migrants even after the inclusion of the country-specific effects.

Thus, the above analysis signifies the importance of geographical and cultural proximity, but also the importance of push and pull factors in the formation of migration flows among OECD countries.

Table 3. Panel data regressions (lagged) on migration flows (ln). Standard errors are shown in parentheses. See text for descriptions of regressions and variables.

Variable	Regression			
	1	2	3	4
$GDPpc_i$ (origin, ln, $t-1$)	−0.051	0.28	0.026	1.042
	(0.46)	(1.16)	(0.22)	(3.36)***
$GDPpc_j$ (destination, ln, $t-1$)	1.054	0.942	0.293	1.937
	(11.76)***	(3.59)***	(2.10)**	(4.64)***
$pop1529share_i$ (origin, $t-1$)	0.790	1.450	1.552	8.696
	(1.26)	(0.84)	(2.36)**	(3.97)***
edu_i (origin, ln, $t-1$)	0.029	0.004	0.035	0.004
	(13.86)***	(1.86)*	(14.24)***	(1.66)*
edu_j (destination, ln, $t-1$)	0.011	−0.010	0.001	−0.004
	(5.34)***	(5.40)***	(0.23)	(1.74)*
$stock95_j$ (destination, ln, $t-1$)			0.919	1.715
			(25.48)***	(2.30)**
D_{ij} (ln)	−0.252	−0.639	−0.338	−0.586
	(6.60)***	(11.31)***	(8.27)***	(6.44)***
$border_{ij}$	1.217	0.472	0.831	0.766
	(7.28)***	(3.23)***	(5.12)***	(4.50)***
$language_{ij}$	0.926	0.627	0.457	0.214
	(6.10)***	(4.62)***	(2.42)**	−1.12
$colony_{ij}$	1.369	0.899	0.585	0.619
	(5.72)***	(4.40)***	(2.22)**	(2.42)**
$region_{ij}$	−0.006	−0.002	0	0.016
	(0.39)	(0.16)	(0.02)	(0.86)
yearly effect	−0.443	0.096	−0.37	0.011
	(12.74)***	(2.79)***	(9.08)***	(0.26)
constant	−9.422	−4.868	−7.594	−38.455
	(6.01)***	(1.36)	(3.88)***	(4.67)***
Origin and destination effects		yes		yes
Observations	4637	4637	2615	2615
Number of groups	757	757	425	425

* significant at 10%; ** significant at 5%; *** significant at 1%.

5.2 Network modeling approach

For further validation of the results of the econometric analysis, a second modeling exercise is introduced. This is a cross-sectional analysis, which incorporates the network structure of international migration. The main modeling tool here is the MRQAP regression, which is primarily used to model social interactions and is widely used in social network analysis research. The main advantage of this method is its ability to address potential interdependence across observations, a phenomenon which is quite often in dyadic data known as dyadic autocorrelation [for a discussion see Agaian et al (1995)]. Such a problem might arise here as the same country appears in various origin–destination country pairs and this might violate the OLS (ordinary least squares) assumption of observation independence (Zhou, 2011). MRQAP is widely used in order to tackle this issue when the level of analysis is the dyadic ties (Krackhardt, 1987; 1988; Mizruchi, 1993).

Just like with any other methods based on quadratic assignment procedure, both dependent and independent variables are $N \times N$ matrices. The algorithm works in two steps. Firstly, a standard multiple regression across the corresponding cells of the matrices representing the dependent and the independent variables takes place. Then, the algorithm randomly permutes rows and columns of the dependent variable matrix and reestimates the regression. This step is repeated hundreds of times for the estimation of standard errors. The coefficients derived from these iterations are compared with those from the original OLS model and the percentage of these coefficients surpassing the original coefficients indicates the statistical reliability of the original outcome (Zhou, 2011). The Ucinet software (see Borgatti et al, 2002) is used for this analysis.

Table 4. Multivariate regression quadratic assignment procedures on migration flows (ln). Standard errors are shown in parentheses.

Variable	2000	2001	2002	2003	2004
D_{ij} (ln)	−0.704	−0.699	−0.925	−0.507	−0.549
	(0.288)***	(0.326)***	(0.385)***	(0.310)**	(0.276)**
border_{ij}	0.928	0.603	0.781	1.406	1.426
	(0.575)*	(0.690)	(0.757)	(0.662)**	(0.568)***
colony_{ij}	1.235	1.494	1.837	1.204	1.251
	(0.778)*	(0.899)**	(1.136)*	(0.883)*	(0.868)*
language_{ij}	1.355	1.441	0.335	1.274	1.261
	(0.624)**	(0.685)**	(0.896)	(0.697)**	(0.573)**
pop1529share_i (origin)	3.134	3.386	3.597	4.001	2.524
	(1.548)**	(1.887)**	(2.355)**	(2.118)**	(1.766)*
edu_i (origin, ln)	0.566	0.672	0.550	0.402	0.782
	(0.272)**	(0.363)**	(0.461)*	(0.366)	(0.360)***
edu_j (destination, ln)	0.722	0.710	0.669	0.737	0.728
	(0.162)***	(0.183)***	(0.202)***	(0.183)***	(0.173)***
GDPpc_i (origin, ln)	2.383	1.864	1.213	1.439	1.971
	(1.189)**	(1.275)*	(1.546)	(1.342)	(1.204)**
GDPpc_i (destination, ln)	0.839	0.757	0.547	0.958	0.521
	(0.366)***	(0.408)**	(0.506)	(0.435)***	(0.331)**
Constant	−45.098	−40.366	−28.401	−37.510	−41.814
R^2	0.340	0.286	0.245	0.220	0.300
Observations	756	600	420	650	756

* significant at 10%; ** significant at 5%; *** significant at 1%.

After the relevant log–log transformations, equation (1) can be transformed to a linear model. More specifically, the empirical model we are estimating is the following:

$$\ln(M_{ijt}) = \beta_0 + \beta_1 \ln(\mathrm{GDPpc}_i) + \beta_2 \ln(\mathrm{GDPpc}_j) + \beta_3 \ln(\mathrm{edu}_i) + \beta_4 \ln(\mathrm{edu}_j)$$

$$+ \beta_5 \ln(\mathrm{pop1529\_share}_i) + \beta_6 \ln(\mathrm{D}_{ij}) + \beta_7(\mathrm{border}_{ij}) + \beta_8(\mathrm{colony}_{ij})$$

$$+ \beta_9(\mathrm{language}_{ij}) + \varepsilon_{ij} \,. \qquad (3)$$

Apart from the quadratic assignment procedures estimation approach, the main difference between equations (2) and (3) is that the latter is a cross section. In addition, the variable representing geographic regions (region_{ij}) has been excluded from the analysis as it failed to provide any insights above.

In table 4 we present the MRQAP results for the ten-year study period. The first observation is the consistency of the impact of cultural proximity through the study period, as is reflected in colonial ties and common language between origin and destination. Secondly, geographic proximity is represented here by two different variables: physical distance and adjacency. Distance has a significant negative impact for six years and border effect is positive and significant for eight years. The above indicates the cost that distance imposes in the migration process. Regarding the pull and push effects, the picture is more complicated. Education in both origin and destination countries has a significant positive impact on migration flows. GDP per capita of the destination country has a positive impact for eight years, while the GDP per capita of the origin country does not have a stable impact. Moreover, the magnitude of the

Table 4 (continued).

Variable	2005	2006	2007	2008	2009
D_{ij} (ln)	−0.383	−0.147	0.016	−0.188	−0.173
	(0.267)*	(0.231)	(0.137)	(0.249)	(0.258)
border_{ij}	1.383	1.245	1.193	1.435	1.133
	(0.567)***	(0.510)***	(0.520)***	(0.512)***	(0.491)**
colony_{ij}	0.903	0.999	2.097	1.294	1.004
	(0.810)	(0.754)*	(0.826)***	(0.713)**	(0.908)
$\mathrm{language}_{ij}$	1.544	1.186	1.214	1.312	1.138
	(0.612)***	(0.544)**	(0.557)**	(0.585)***	(0.630)**
$\mathrm{pop1529share}_i$ (origin)	1.724	1.172	2.144	−0.197	−1.527
	(1.630)	(1.632)	(1.585)*	(1.364)	(1.569)
edu_i (origin, ln)	0.547	0.777	0.911	0.504	0.850
	(0.316)**	(0.308)***	(0.167)***	(0.264)**	(0.330)***
edu_j (destination, ln)	0.694	0.738	0.677	0.618	0.735
	(0.151)***	(0.165)***	(0.131)***	(0.135)***	(0.179)***
GDPpc_i (origin, ln)	1.443	2.091	1.346	−0.021	−0.004
	(1.209)	(1.133)**	(0.684)**	(1.090)	(1.063)
GDPpc_i (destination, ln)	0.242	0.294	0.293	−0.498	−0.535
	(0.298)	(0.309)	(0.288)	(0.233)***	(0.260)***
Constant	−31.947	−43.874	−38.993	−8.728	−13.953
R^2	0.254	0.327	0.336	0.246	0.321
Observations	930	900	540	992	650

coefficient of the latter is much lower than the former. Finally, the share of young population in the origin country also has a positive impact for six cross-sections.

In summary, MRQAP validates the previous results on the impact of physical and cultural proximity on international migration, the pull effects of the GDP per capita, and the push effect of the young population. The two models disagree on the impact of education, but MRQAP results are in accordance with the lagged regressions.

6 Concluding remarks

The novelty of our approach is the adoption of a network perspective in the study of international migration. Such migration flows create a dense, complex, and dynamic network constellation. The topology and the structure of this network change over time, clearly reflecting current economic, social, and political conditions. Although the growing attention that research on migration has attracted, and in light of the recent major developments in network science, the interaction between these two fields has been very limited.

Using methods widely used in migration studies and network analysis, much research effort is spent here to bridge this gap. Exploratory network analysis using centrality measures as well as community detection provides the fundamentals in order to approach international migration as a global network. The outcomes of this exercise support the modeling part of our analysis. In this stage standard econometric techniques are blended with network approaches to gain a better understanding of the determinants of international migration. Both methodological approaches converge on the importance of physical and cultural proximity in the formation of migration flows. Physical distance and border effects are significant predictors of migration flows among OECD countries. Moreover, postcolonial ties and a common language between origin and destination countries have a positive impact on migration flows. In addition, both modeling approaches agree on the pull effect of the prosperity level of the destination country as well as the push effect of the existence of a pool of young people in the origin country. The value added by this dual approach is the emergence of education as a significant predictor of migration flows. Indeed, according to the MRQAP models, a higher level education can generate both pull and push effects in migration among OECD countries. Nonetheless, traditional regression techniques did not capture the impact of education and it was the network-level modeling approach adopted here that highlighted this impact.

In our 'age of migration' countries are increasingly tied together through human-capital flows. In a globally ageing world these forces tend to become even stronger. In this new international playing field, education and skills will become increasingly important forces that drive global connectedness. It is certain that the future of many welfare states will be determined by cross-border migration, not only in terms of volumes, but also in terms of quality. Due policy attention and advanced quantitative research is needed to achieve a balance in global human capital flows.

To conclude, we have highlighted the need for borrowing methods and techniques from the field of network analysis in order to better understand complex spatial phenomena such as international migration. Although effort was spent to understand international migration as a network phenomenon and to bridge the gap between descriptive and modeling network approaches, there is still much ground to be covered in this respect, with estimation of constructive stochastic migration models being one of the most important elements in this novel approach.

References

Adsera A, Chiswick B R, 2007, “Are there gender and country of origin differences in immigrant labor market outcomes across European destinations?” *Journal of Population Economics* **20** 495–526

Agaian S, Astola J, Egiazarian K, 1995 *Binary Polynomial Transformation and Nonlinear Digital Filters* (Marcel Dekker, New York)

Aldenderfer M S, Blashfield R K, 1984 *Cluster Analysis* (SAGE, Newbury Park, CA)

Aroca P, Maloney W F, 2005, “Migration, trade, and foreign direct investment in Mexico” *The World Bank Economic Review* **19** 449–472

Baltagi B H, 2001 *Econometric Analysis of Panel Data* (John Wiley, Chichester, Sussex)

Baltagi B H, Wu P X, 1999, “Unequally spaced panel data regressions with AR (1) disturbances” *Econometric Theory* **15** 814–823

Barabási A-L, 2002 *Linked: The New Science of Networks* (Perseus, Cambridge, MA)

Barthelemy M, 2011, “Spatial networks” *Physics Reports* **499** 1–101

Belot M, Ederveen S, 2012, “Cultural and institutional barriers in migration between OECD countries” *Journal Population Economics* **25** 1077–1105

Bhattacharya U, Groznik P, 2008, “Melting pot or salad bowl: some evidence from U.S investment abroad” *Journal of Financial Markets* **11** 228–258

Blondel V D, Guillaume J-L, Lambiotte R, Lefebre E, 2008, “Fast unfolding of community hierarchies in large networks” *Journal of Statistical Mechanics: Theory and Experiment* **10** P1000

Borgatti S P, Everett M G, Freeman L C, 2002, “Ucinet for Windows: Software for Social Network Analysis”, Analytic Technologies, Harvard, MA, http://pages.uoregon.edu/vburris/hc431/Ucnet_Guide.pdf

Bridge C, Fedorowich K, 2003 *The British World: Diaspora, Culture and Identity* (Frank Cass, London)

Brun J-F, Carrere C, Guillaumont P, Melo J D, 2005, “Has distance died? Evidence from a panel gravity model” *The World Bank Economic Review* **19** 99–120

Buchanan M, 2002 *Nexus: Small Worlds and the Groundbreaking Science of Networks* (Norton, New York)

Clauset A, Newman M E J, Moore C, 2004, “Finding community structure in very large networks” *Physical Review E* **70** 1–6

Drucker D M, 2003, “Testing for serial correlation in linear panel-data models” *The Stata Journal* **3** 168–177

Etzo I, 2011, “The determinants of the recent interregional migration flows in Italy: a panel data analysis” *Journal of Regional Science* **51** 948–966

Expert P, Evans T S, Blondel V D, Lambiotte R, 2011, “Uncovering space-independent communities in spatial networks” *Proceedings of the National Academy of Sciences* **108** 7663–7668

Fischer C, 2007, “The influence of immigration and international tourism on the import demand for consumer goods—a theoretical model”, in *Advances in Modern Tourism Research: Economic Perspectives* Eds D Álvaro, P Nijkamp, P Neto (Physica, Heidelberg) pp 37–50

Gheasi M, Nijkamp P, Rietveld P, 2011, “Migration and tourist flows”, in *Advances in Tourism Economics* Eds A Matias, P Nijkamp (Physica, Heidelberg) pp 111–126

Gibney M J, Hansen R, 2005 *Immigration and Asylum: From 1900 to The Present* (ABC-CLIO, Santa Barbara, CA)

Girma S, Yu Z, 2002, “The link between immigration and trade: evidence from the United Kingdom” *Review of World Economics* **138** 115–130

Goldin I, Cameron G, Balarajan M, 2011 *Exceptional People* (Princeton University Press, Princeton, NJ)

Gould D M, 1994, “Immigrant links to home country: empirical implications for US bilateral trade flows” *Review of Economics and Statistics* **76** 302–316

Hatton T, Williamson J, 1998 *The Age of Mass Migration: Causes and Economic Impact* (Oxford University Press, New York)

Hatton T, Williamson J, 2003, “Demographic and economic pressure on emigration out of Africa” *Scandinavian Journal of Economics* **105** 465–486

Head K, Ries J, 1998, "Immigration and trade creation: econometric evidence from Canada" *Canadian Journal of Economics* **31** 47–62
Hsiao C, 2003 *Analysis of Panel Data* (Cambridge University Press, Cambridge)
Javorcik B S, Özden Ç, Spatareanu M, Neagu C, 2011, "Migrant networks and foreign direct investment" *Journal of Development Economics* **94** 231–241
Krackhardt D, 1987, "QAP partialling as a test of spuriousness" *Social Networks* **9** 171–186
Krackhardt D, 1988, "Predicting with networks: nonparametric multiple regression analysis of dyadic data" *Social Networks* **10** 359–381
Kugler M, Rapoport H, 2007, "International labor and capital flows: complements or substitutes?" *Journal of Economic Letters* **94** 155–162
Latora V, Crucitti P, Marchiori M, Rapisarda A, 2003, "Complex systems: analysis and models of real-world networks", in *Energy and Information Transfer in Biological Systems: How Physics Could Enrich Biological Understanding* Eds F Musumeci, L S Brizhik, M-W Ho (World Scientific Publishing, Singapore) pp 188–204
Lewer J, Van den Berg H, 2008, "A gravity model of immigration" *Journal of Economic Letters* **99** 164–167
Li L, Alderson D, Tanaka R, Doyle J C, Willinger W, 2005, "Towards a theory of scale-free graphs: definition, properties, and implications (Extended version)" *Internet Mathematics* **2** 431–523
Maier G, Vyborny M, 2008, "Internal migration between US states: a social network analysis", in *Migration And Human Capital* Eds J Poot, B Waldorf, L v Wissen (Edward Elgar, Cheltenham, Glos) pp 75–95
Mayda A M, 2007, "International migration: a panel data analysis of the determinants of bilateral flows" *Journal of Populations Economics* **23** 1249–1274
Mizruchi M S, 1993, "Cohesion, equivalence, and similarity of behavior: a theoretical and empirical assessment" *Social Networks* **15** 257–307
Newman M E J, 2003, "The structure and function of complex networks" *SIAM Review* **45** 167–256
Newman M E J, Girvan M, 2004, "Finding and evaluating community structure in networks" *Physical Review E* **69** 1–15
Nijkamp P, Gheasi M, Rietveld P, 2011, "Migrants and international economic linkages: a meta-overview" *Spatial Economic Analysis* **6** 359–376
Nijkamp P, Poot J, Sahin M, 2012 *Migration Impact Assessment: New Horizons* (Edward Elgar, Cheltenham, Glos)
OECD, 2011, "OECD.StatExtracts", http://stats.oecd.org/Index.aspx
Özden Ç, 2005, "Brain drain in Latin America", in *Expert Group Meeting on International Migration and Development in Latin America and the Caribbean* (UN), http://www.un.org/esa/population/meetings/IttMigLAC/P10_WB-DECRG.pdf
Pons P, Latapy M, 2006, "Computing communities in large networks using random walks" *Journal of Graph Algorithms Applications* **10** 191–218
Pöyhönen P, 1963, "A tentative model for the volume of trade between countries" *Weltwirtschaftliches Archiv* **90** 93–99
Rauch J E, Trindade V, 2002, "Ethnic Chinese networks in international trade" *Review of Economics and Statistics* **84** 116–130
Reggiani A, 2009, "From complexity to simplicity", in *Complexity and Spatial Networks* Eds A Reggiani, P Nijkamp (Springer, Berlin) pp 275–284
Stata Corporation, 2007, "Stata: Release 10.0", Stata Corporation, College Station, TX
Taaffe E J, Gauthier E L, O'Kelly M E, 1996 *Geography of Transportation* (Prentice-Hall, Englewood Cliffs, NJ)
Tinbergen J, 1962 *Shaping the World Economy: Suggestions for an International Economic Policy* (The Twentieth Century Fund, New York)
Watts D J, 2003 *Six Degrees: The Science of a Connected Age* (Norton, New York)
Watts D J, 2004, "The 'New' Science of Networks" *Annual Review of Sociology* **30** 243–270
Williams A M, Hall C M, 2002 *Tourism and Migration: New Relationship Between Production and Consumption* (Kluwer Academic, Dordrecht)

Wooldridge J M, 2002 *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Cambridge, MA)
Wooldridge J M, 2003 *Introductory Econometrics: A Modern Approach* (South-Western, Mason, OH)
Zhou M, 2011, "Intensification of geo-cultural homophily in global trade: evidence form the gravity model" *Social Science Research* **40** 193–209