



The structure and dynamics of population migration among economic areas in the United States from 1990 to 2011

Zengwang Xu

University of Wisconsin Milwaukee, Department of Geography, P O Box 413, Bolton Hall Room 410, Milwaukee, WI 53201-0413, USA (e-mail: xuz@uwm.edu)

Received: 2 June 2016 / Accepted: 2 December 2016

Abstract. This study constructs the population migration networks among economic areas in the United States for every consecutive year from 1990 to 2011, and examines their structural properties and population migration dynamics. Various aspects of the structural properties of the networks are explored, including the connectivity, clustering, assortativity and centrality. It was found that these structural properties are mediated by migration dynamics and inter-area distance, and the patterns of varying structural properties across areas of different connectivity reveal the hub-and-spoke structure of the networks. It is evident that there exists tremendous complexity in migration connectivity and dynamics in the US internal migration system.

JEL classification: J00

Key words: County-to-county migration, complex networks, migration system, migration network

1 Introduction

Human population migration can be studied as period-specific population flows among places, and the migration flows form complex local and national connections among places due to their specific economic, social, demographic, and cultural characteristics that attract or repulse different kinds/volume of migrants (Greenwood 1985, 1988; Mueser 1989; Rayer and Brown 2001; Plane and Jurjevich 2009). With the rapid development of information and communication technology and the declining cost of the transportation, places are increasingly connected, and the geographic distance becomes a less important factor for migrants when deciding where to move (Pedersen et al. 2008). The network effect or the connectivity among places plays an increasingly significant role in population migration; a typical example is the so-called beaten path effect, in which migrants rely heavily on social networks (such as kinship, friendship, and shared community origin) in their migration decisions and form steady flows among certain localities (Massey et al. 1993; Frey et al. 2005). As a result, migration among places forms population migration networks that are increasingly interconnected and complex. Based on the Internal

Revenue Service (IRS) annual county-to-county migration flow data, this study constructs the population migration networks among the economic areas in the United States for every consecutive year from 1990 to 2011, and investigates the various aspects of the structural properties of the networks and how the connection structure of the networks relates the population migration process taking place on the networks.

Most population migration studies are based on migration measurements aggregated at places and rely on the classic statistical models that are grounded on the independence assumption (Plane 1984; Pandit 1997a, 1997b, 2000). The place-based aggregate migration measurements (e.g., total number of in or out-migrants) reflect the effects of a period of population migration on places but do not contain the information of the population movement processes among places (i.e., migration flows or streams), which are important to address the inadequacy in the assumptions of linear distance and isotropic directionality in migration studies (Wolpert 1967). The places are interconnected to each other in many ways and the interconnections are increasingly complex. Ignoring the connection structures or failing to model the accurate connection relationships can introduce significant bias in modelling the spatial interaction among places (Olsson 1965; Goodchild and Smith 1980; Fotheringham 1984; Mueser 1989). The network approach employed in this study directly addresses the connectivity and population migration process among places.

The area-to-area migration flow represents the migration stream or the volume of the migrants between areas. The area-to-area migration data have been used in studying movement dynamics (Tobler 1987; Holland and Plane 2001; Rae 2009), and in modelling spatial interaction among places (Fotheringham and Webber 1980; Plane 1984; Slater 1985). Areas can be connected by migration streams as population migration networks in which the areas are vertices and migration streams are edges. Nevertheless, the network approach has been an underutilized method in migration study. There is only a paucity of studies that explore the population migration from the network perspective. Slater (1984b, 1985) studied the US inter-county migration (i.e., 5 year inter-county migration flows from 1970 census, e.g., 1965–1970) by applying hierarchical clustering to the standardized inter-county migration table and found US counties form large migration fields. The inter-county migration table is the adjacency matrix representation of population migration network. Plane (1999) recognized that the human population migration among places has great similarity with the networked systems studied in physical geography, and he investigated the temporal migration dynamics of US state-to-state migration flows from 1980 to 1995, and found the most volatile and stable interstate migration streams, and the ‘floods’ and ‘droughts’ of migration streams. In this study, the area-to-area population migration flows are connected as directed and weighted networks of population migration in the United States.

The network approach has been a classic method to study movement and connectivity in networked systems dating back to the geography quantitative revolution (Haggett and Chorley 1969; Haggett et al. 1977; Gastner and Newman 2006). Unlike the classic statistical models that are grounded on the independence assumption, the network approach emphasizes connectivity, interdependence, and evolution (Barabási and Frangos 2014). With the availability of greater and cheaper computing power and large-scale empirical datasets on networked systems, many systems and structures in nature and society have been represented and analysed as networks, and the last two decades have witnessed the explosive growth of studies on a variety of networks and networked systems of much larger size and more complex structure. As the result, a multidisciplinary synthesis of new measures, techniques, and modelling approaches has emerged as a ‘new’ science of networks (Watts 2004; Barabási and Frangos 2014).

The emerging science of networks defines new concepts and measures to characterize various aspects of the structure of networks, discovers the unifying principals and statistical properties common to various empirical networks, and explores how the structure of networks affects

the dynamics taking place on networks (Boccaletti et al. 2006; Newman 2003b). What makes it especially 'new' are the new network models that capture the essential structures of various complex networked systems in nature and society, for example, the small-world network (Watts and Strogatz 1998) and scale-free network models (Barabási and Albert 1999). The small-world network represents a common structure existing in many networks, namely, these networks have both highly clustered and highly connected structures (Watts and Strogatz 1998). Many real-world networks are also found to be scale-free networks of which the distributions of vertex connectivity follow a power-law distribution, with exponents varying in the range 2 to 3 (Barabási and Albert 1999; Barabási 2009). The power law distribution is a statistical manifestation of the facts that the vertex connectivity in many real-world networks, rather than being close to an average number, spans many magnitudes of scales. In fact, it is common to many networks that a few vertices are very highly connected while the majority of vertices have only a few connections. These fundamental natures of the networks have great implications for dynamics occurring on networks, such as the diffusion of epidemics is affected by the structure of social networks (Pastor-Satorras and Vespignani 2002; May and Lloyd 2001; Xu and Sui 2009). In addition to these findings, recent advances in network science have suffused with new measures and methods to characterize various aspects of the structure of networks as they become increasingly large and complex (Boccaletti et al. 2006). Comprehensive reviews on the concepts and methods can be found in (Newman 2003b; Boccaletti et al. 2006).

Recent studies on international migration networks have been evident the usefulness of the network approach in better understanding the structures and migration dynamics of country-to-country population migration network (Davis et al. 2013; Fagiolo and Mastroiello 2013). The population migration networks in the United States from 1990 to 2011 are constructed from the county-to-county migration flows aggregated from the address changes between two consecutive years in IRS income tax returns. The networks contain about 80 per cent of the total US internal migrants with a one-year interval. Based on this valuable dataset, this study takes advantage of the advances in network science and examines various aspects of the structural properties and migration dynamics of the internal migration systems/networks in the United States from 1990 to 2011.

In the remainder of this paper, the population migration networks that are aggregated from the address changes in individuals' income tax returns filed to IRS in every consecutive year from 1990 to 2011 are described. Additionally, the details of the measures and methods of the network approach adopted in this study are articulated and the structural properties of the population migration networks are reported.

2 The population migration networks of counties and economic areas

The US Internal Revenue Service (IRS) provides county-to-county migration flow data based on the address changes reported in the individuals' annual income tax returns. Before 2011, the migration data only include those individuals who filed income tax returns before September, and it is only after 2011 that the IRS migration data are based on full year tax returns (Pierce 2015). The county-to-county migration flow data used here are between two consecutive years from 1990 to 2011, which is a period of consistent data processing in IRS. Similar data should exist for early years as Engels and Healy (1981) and Plane (1982, 1999) studied IRS migration data before 1990, but only the data after 1990 were available for this study. The inter-county migration flows consist of three tax return variables: the number of returns (used as the estimate of the number of migrated households); the number of exemptions (used as the estimates of the number of migrants); and the aggregate adjusted gross income (available only after 1992). Each

county has the same variables for non-migrants whose addresses have not changed (non-movers) or not changed to different county in the consecutive annual tax filing.

Since the IRS migration flow data are based on individuals' income tax records, they are considered inclusive and reliable annual population mobility data with information on the county-to-county migration of the number of households, migrants, and the aggregated change of family income (Gross 2003). However, the IRS migration data have several limitations, most of which as well as their strengths have been well documented (Gross 2003; Manson and Groop 2000). The data are only based on those tax returns filed by the end of September, and they represent 95–98 per cent of total annual filings, and it is only after 2011 that the IRS migration data are based on total annual filings with several other enhancements (Pierce 2015). Since the poor and/or elderly may not file income tax, the IRS migration data only cover about 80 per cent of the total population, and the coverage varies state by state (Plane 1999; Manson and Groop 2000). The IRS migration data do not have other migration relevant attributes, like age, education, race, or ethnicity, and they have suppressed those inter-county migration flows if they have less than 10 returns (Manson and Groop 2000). With these caveats, the IRS migration data however 'represent an extremely large sample compared with other migration sources, such as the Current Population Survey's March mobility question or the decennial census (long-form) question on place of previous residence' (Plane 1999, p. 315), and they provide important migration flow data with a one-year interval for every year through at least last two decades.

Based on the 21-year IRS inter-county migration data from 1990 to 2011, the total number of migrants ranged from 8,658,703 (1990–1991) to 10,729,375 (2005–2006), and the national migration rates ranged from 4.16 per cent (2009–2010) to 5.00 per cent (2005–2006). Figure 1(a) shows the total number of annual inter-county migrants and non-migrants (including non-movers and intra-county movers) from 1990 to 2011. Figure 1(b) shows the total number of the annual inter-county migrants and its ratio to the total population. There were significant increases in total number of migrants during 2004–2007 period (10,243,130 for 2004–2005, 10,729,376 for 2005–2006, 10,157,450 for 2006–2007, and 10,257,449 for 2007–2008, and only these four years have more than 10 million inter-county migrants during the last two decades). During the 2004–2007 period, multiple hurricanes hit the Gulf Coast (in 2004 and 2005), and especially Hurricanes Katrina and Rita caused severe damage and large population displacement in 2005. Economic factors are important to migration; starting in 2008, the recession contributes to the fall off in migration. The national migration rate reached the maximum (5.00%) at 2005–2006 (Figure 1(b)).

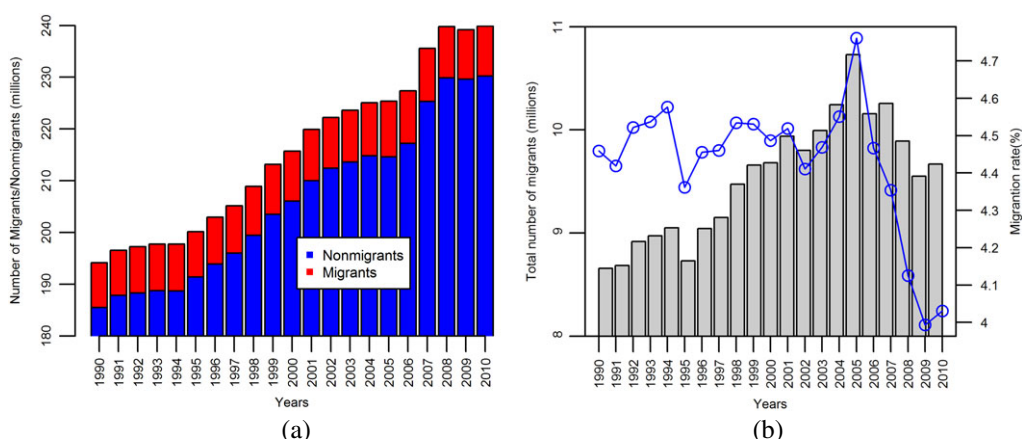


Fig. 1. (a) The number of migrants and non-migrants for every consecutive year from 1990 to 2011, (b) the total number of migrants (bar) and total national migration rate (curve) at every consecutive year from 1990 to 2011

As a basic exploration and verification of the annual county-to-county population migration, we examined the top 10 counties gaining the most population from migration for every annual migration from 1990 to 2011. These counties are the major metropolitan counties in the nation, and their migration effect reflects certain degrees of the most volatile population migration dynamics across the nation. Volatility of the annual migration can be appreciated by comparing the population gain from migration of the county ranked 1st over the years. The county ranked 1st in 2008–2009 (Fort Bend, TX) has only gained 14,516 people from migration, but the county ranked 1st in 2004–2005 (Maricopa, AZ) has gained 56,204 people. Volatility can also be reflected by the change of the rank of top gainers. Maricopa County (AZ) was ranked number nine in 1990–1991, number three in 1991–1992, and stayed in the top three counties until 2006–2007. It then fell to number nine in 2007–2008, and fell out of the top 10 counties afterwards. Clark County (NV) was the number one county gaining population from migration in 1990–1991, stayed in the top three until 2006–2007, and fell out of the top 10 in 2008–2009 and afterwards. These volatile migration dynamics can only be examined with the county-level annual migration data.

The US counties have several features that prevent them from being the ideal unit of analysis in this study. It is well known that the US counties have drastic differences in geographic size and total population. This is the fundamental reason for the modifiable areal unit problem (MAUP) that causes the migration flows or other county-based migration measures not statistically comparable. Moreover, the US counties are not a socially and economically meaningful unit. The nature of migration between two adjacent metropolitan counties can be totally different from migration between two non-metropolitan counties. In the New England region, mainly in Virginia, cities or towns are separated from counties in migration statistics, and they should be merged to be comparable with other counties. In this study, we merge counties and equivalents in the IRS county-to-county migration datasets to the economic areas defined by the Bureau of Economic Analysis (BEA) of USDA in 2004. One hundred and seventy seven economic areas in the continental United States were defined on counties to consider the population density and commuting pattern. These economic areas are also considered better representation of local labour markets.

The population migration networks among the BEA economic areas (area-to-area or inter-area networks thereafter) are still relative large, with 177 BEA Economic Areas as vertices and more than 9,000 inter-area migration flows as edges. In addition to distance, the edges of the networks have three weights: the number of returns, the number of exemptions, and the aggregate adjusted gross income (in thousands). Table 1 summarizes the basic statistics of the 21 inter-area networks.

The number of returns or exemptions, and aggregate adjusted gross income of a migration connection between two areas are affected by or are the artifacts of the total population of the areas. To remove the population effect, the iterative proportional fitting procedure (IPFP) (Fienberg 1970; Slater 1984a) is applied to these three weights as well as distance. Figure 2 maps the large inter-area migration flows (those that have more than 3,000 exemptions) in the population migration networks for the years of 1990–1991 and 2010–2011 as well as the standardized flows of both networks. There were apparent local and national migration flow patterns in the original networks of both years (Figure 2(a)). High volume migration occurs among the nation's major metropolitan areas. In the West Coast, these include metropolitan areas in California, Arizona and Washington. In the East Coast, these mainly include those in Florida and New York. Metropolitan areas in Texas and Illinois also have high volume migration exchange with others. Figure 2(b) maps the same networks in Figure 2(a) but using standardized migration flows by the IPFP. The high volume migration among the major metropolitan areas in the East and West Coast are mainly the artifacts of the large total population of the origin and destination metropolitan areas, and their standardized weights are less than 0.1 and no longer prominent.

Table 1. Summary statistics of the area-to-area population migration networks from 1990 to 2011

Year	Vertices	Edges	Numner of returns			Number of exemptions			Aggregate income (thousands)		
			Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
1990–1991	177	9126	10	215	25197	10	413	47890	N/A	N/A	N/A
1991–1992	177	9245	10	208	23553	10	398	43446	N/A	N/A	N/A
1992–1993	177	9448	10	205	23965	10	398	43959	–2667	6239	757230
1993–1994	177	9352	10	208	23254	10	402	43684	–87	6449	775615
1994–1995	177	9569	10	211	22790	10	400	40998	–975896	6521	804215
1995–1996	177	9423	10	209	22641	10	395	40982	–6566	7272	833992
1996–1997	177	9434	10	216	23096	10	404	40645	–220	8010	952391
1997–1998	177	9505	10	218	21318	10	403	37223	–4362	8767	1006947
1998–1999	177	9525	10	222	20484	10	408	34701	79	9558	993464
1999–2000	177	9495	10	228	21763	10	417	36563	–546	10475	1099392
2000–2001	177	9485	10	234	22936	10	423	44883	97	11611	1439211
2001–2002	177	9437	10	230	24561	10	417	48755	–4409	10420	1309507
2002–2003	177	9197	10	227	22365	10	413	43835	–2754	9805	1140182
2003–2004	177	9091	10	232	21089	10	425	42871	–3433	10240	1272861
2004–2005	177	9317	10	238	20999	10	439	41573	–5224	11139	1547913
2005–2006	177	9445	10	254	21745	10	471	43997	83	12216	1424315
2006–2007	177	9428	10	243	19635	10	447	35844	40	12129	1239259
2007–2008	177	9566	10	246	19768	10	441	33803	–1421	12245	1178699
2008–2009	177	9540	10	234	19216	10	419	32274	–18421	10969	1036778
2009–2010	177	9222	10	224	17538	10	401	32977	–7164	9832	964494
2010–2011	177	9541	10	232	18844	10	412	32105	–1611	10689	1037649

3 Structure and migration dynamics of the networks

The inter-area population migration networks can be represented as $N \times N$ weight matrices, W^t ($t = 1990, 1991, \dots, 2010$), where W^t is the matrix for year t and N is the total number of economic areas in the networks. A cell in the matrix of year (t) , w^t_{ij} , represents the weight (the number of returns, exemptions, or the aggregate adjust gross income) that associate with the migration from area i to j . When only the connection topology of the networks is considered, the weight matrices can be converted to binary matrices, B^t , in which $b^t_{ij} = 1$ if the number of migrants is greater than 0 and $b^t_{ij} = 0$ otherwise. Since the weight of migration from area i to j is very often different from the weight of migration from area j to i , the weight matrices W^t are weighted and asymmetric. This matrix notation provides the basis of the definitions of network metrics in the following sections. These network metrics characterize four essential aspects of network structure: connectivity, clustering, assortativity, and centrality (Fagiolo et al. 2009).

3.1 Persistent heterogeneity in migration connections and weights

A basic migration characteristic of an economic area is how many other economic areas with which it has exchanged migrants. In the inter-area population migration networks, this can be characterized by the migration connectivity of an economic area, that is, the area’s connection degree. Formally, the connection degree of an area i can be represented as $d_i = \sum_{j \neq i} b_{ij}$, where b_{ij} is the binary adjacency matrix, and $b_{ij} = 1$ if the area i and j are connected, $b_{ij} = 0$ otherwise.

The inter-area population migration networks are weighted networks, in which edges represent not only the existence of migration connections but also how much weight the connections carry, or how strong the migration connections are. The weights of the inter-area population

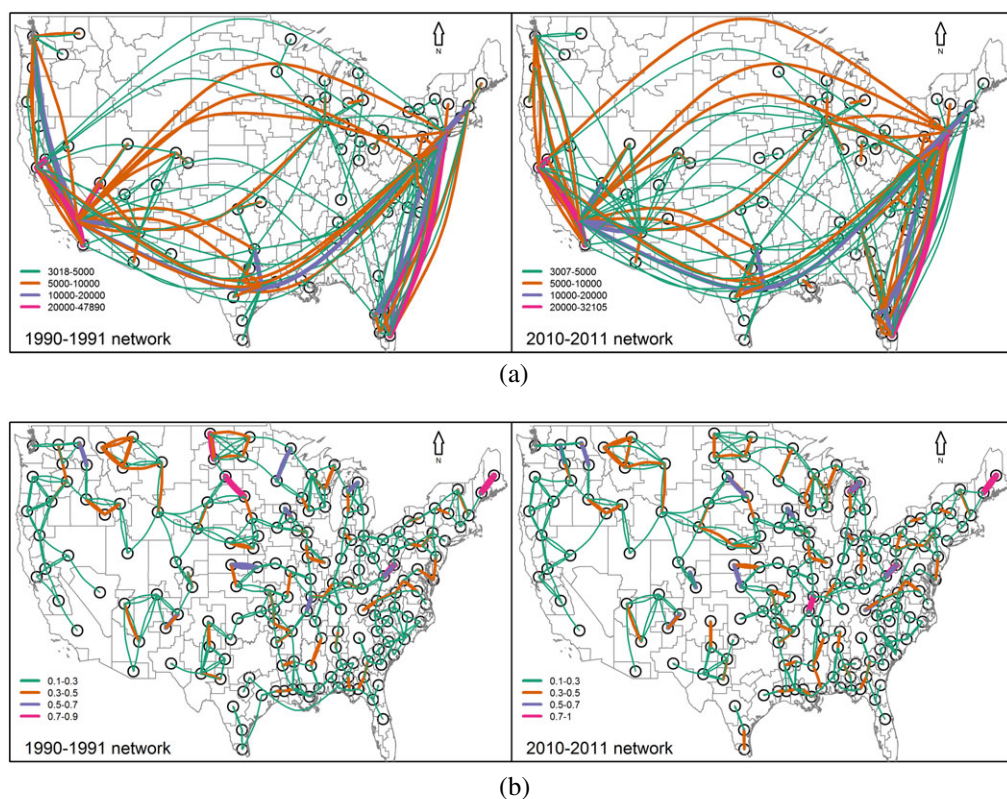


Fig. 2. The 1990–1991 (left) and 2010–2011 (right) population migration networks among economic areas that are defined by BEA in 2004: (a) migration flows of greater than 3000 exemptions; (b) standardized migration flows of greater than 0.1

migration networks include the geographic distance, number of returns or exemptions, and aggregate adjusted gross income.

Across the networks, there exists extraordinary heterogeneity in the areas' migration connectivity and the weights associated with the inter-area migration connections. The heterogeneity is much larger than what the bell-shaped probability distribution can characterize. The probability distributions of these characteristics are right-skewed with heavy tails, and can be approximated by gamma distribution with shape parameters ranging 1.72–1.86 for total degree, 1.59–1.80 for in degree, and 1.76–2.07 for out degree. These heavy tail distributions imply that there are a large number of areas having small number of migration connections to other areas, while only a small number of areas have a large number of migration connections, and this heterogeneity has been persistent in the 21 networks over the last two decades. Figure 3 shows the cumulative probability distribution of in-, out-, and total degree for all the 21 migration networks.

The distributions of standardized weights (i.e., geographic distance, the number of returns and exemptions and the aggregate adjusted gross income) are highly skewed power law distributions (Figure 4). These power law distributions have exponent ranging 1.72–1.75 for the number of returns, 1.73–1.6 for the number of exemptions, 1.74–1.77 for the aggregate adjusted gross income, and 2.38–2.68 for the distance. These highly right skewed distributions imply the existence of large number of small volume inter-area migrations and very small number of large volume inter-area migrations.

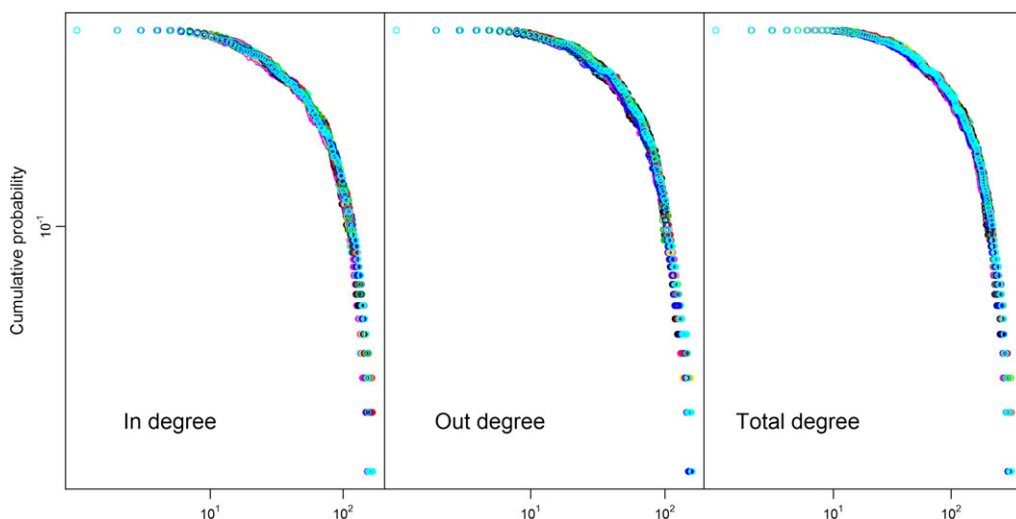


Fig. 3. The cumulative distributions of in-, out-, total degree of the vertices of all networks

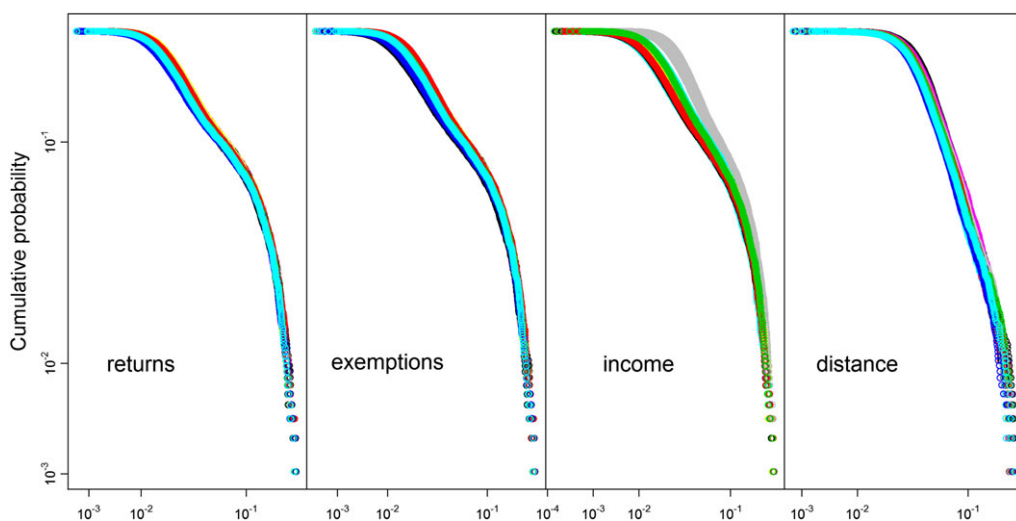


Fig. 4. Cumulative distribution of standardized weights of the edges of all networks (only networks after 1992 have income distribution)

3.2 Persistent high connectivity and relative low clustering

In addition to the connection degree, the connectivity of each vertex in a network can be characterized by its average shortest path length. The connectivity of a network can be quantified by the mean value of its vertices' average shortest path length. High connectivity should be the essential feature of any networks. Real-world networks are usually not as highly connected as fully connected or randomly connected networks, as their vertices have limited capability to connect or more connections might raise costs. However, many networks maintain high connectivity with highly clustered connections, which is the essential feature of the small-world networks (Watts and Strogatz 1998).

Clustering is about the density of the connections. For example, in a hub-and-spoke network, if the spokes do not connect to each other at all, the network is not clustered. If the spokes are fully connected to each other, the hub-and-spoke network is fully clustered. Many real-world networks fall between the not-clustered and fully clustered. In the inter-area population migration networks, this clustering characteristic can be measured by the extent of an area's directly connected neighbours being connected to each other, which is indicative of the cohesiveness of an area's neighbourhoods (i.e., areas have migration connections with the area). The clustering is quantified by the transitivity or clustering coefficient (or weighted clustering coefficient) of an area (Watts and Strogatz 1998; Barrat et al. 2004). The clustering coefficient for an area i can be formulized as $c_i = n_{\text{triplets}}/n_{\text{possible_triplets}} = \sum_{j,k} b_{ij}b_{ik}b_{jk}/d_i(d_i - 1)$, where d_i is the connection degree of area i , b_{ij} (or b_{ik} or b_{jk}) is 1 if area i and j (or i and k , or j and k) are directly connected and is 0 otherwise, and the j and k are all the areas that directly connect to area i . If an area's clustering coefficient is 1, it means that the area's connected neighbouring areas are fully connected to each other. A zero clustering coefficient means that the area's connected neighbouring areas do not connect to each other at all. The clustering coefficient of a network is the mean value of its vertices' clustering coefficients.

The overall connectivity and clustering of the networks are inspected by computing the average shortest path lengths and clustering coefficients of the networks. The average shortest path lengths range from 1.71 to 1.73 for all the 21 networks. This means that two areas in the US inter-area population migration networks on average are a little less than two steps away. A similar sized random network (e.g., a network with 177 vertices and 9,000 edges) has the average shortest path length, 1.42. Therefore, the US inter-area population migration networks have as high connectivity as a similar sized random network.

The clustering coefficients of the inter-area population networks range from 0.26 to 0.28. The clustering coefficient for a similar sized random network (i.e., a network with 177 vertices and 9,000 edges) is 0.58. Therefore, the US inter-area population migration networks have connectivity as high as a similar sized random network, but their clustering are not as high as a similar sized random network. As such, the inter-area population migration networks do not have the connection topology of small-world networks, which have higher clustering than a similar sized random network (Watts and Strogatz 1998).

3.3 Low connectivity areas tend not to connect each other

For weighted networks like the inter-area population migration networks in this study, the clustering coefficient of an area i can be adjusted by considering the weights of edges that connect with the neighbours (the neighbours of an economic area i are all the areas to which the area i directly connect); and the weighted clustering coefficient for area i can be formulized as, $c_i^w = \sum_{j,k} (w_{ij} + w_{ik})b_{ij}b_{ik}b_{jk}/2s_i(d_i - 1)$, where s_i is the strength of the area i . The weighted clustering coefficient (c_i^w) of an area also ranges from 0 to 1. A large value of c_i^w implies that those areas connecting area i with large weight are more likely to connect to each other.

For each area in the inter-area population migration networks, the clustering coefficient is calculated based on both topological connection and the weights of the connections, that is, the number of returns, number of exemptions, aggregate adjusted gross income, and geographic distance between connected areas. The higher value of the weighted clustering coefficient implies not only more triplets are formed among an area's neighbours, but also that the triplets are formed by high weight connections. The average clustering coefficient over the whole network measures the overall network cohesiveness or the average density of interconnected triplets.

The network clustering coefficients range 0.26–0.28 when only the topological connection is considered (i.e., all edges are assumed having no weight or equal weight). When the edge's

weights are considered, the weighted network clustering coefficients for the networks increase to 0.70–0.72 for the number of returns and exemptions, 0.69–0.71 for the aggregate adjusted gross income, and 0.75–0.79 for geographic distance. The distance-weighted clustering coefficients are the highest over all the years, the income-weighted and the return and exemption-weighted are not much different, but they are much larger than the topological clustering coefficients. This implies that many triplets are formed by the inter-area migrations of long geographic distance, high income, and high migration volume (i.e., large number of returns and exemptions).

Areas of different connection degree might have different degree of clustering (or cohesiveness) among areas to which they connect. The relation between the connection degree of areas d_i and their clustering coefficients (or weighted clustering coefficients), $c_i(d)$ (or $c_i^w(d)$), indicates how the areas' connectivity relates the cohesiveness in its connected neighbours throughout the networks (Barrat et al. 2004; Xu and Harriss 2008). A decaying $c_i(d)$ (or $c_i^w(d)$) with increasing connection degree indicates the hub-and-spoke network structure, in which large degree areas serve as hubs and they connect to many low-connectivity areas that do not connect among each other. An increasing $c_i(d)$ (or $c_i^w(d)$) with increasing connection degree indicates that the network is highly connected and clustered.

The clustering coefficients are computed for each area in the networks. The mean clustering coefficients of areas in 50 degree intervals are plotted with the degree. The inter-area population migration networks have demonstrated the declining clustering coefficient (for both weighted clustering and topological clustering) as the connection degree of areas increases, the same pattern found in many real-world networks (Ravasz and Barabasi 2003; Barrat et al. 2004; Xu and Harriss 2008; Lin and Ban 2014). Figure 5 shows the declining clustering coefficients of three selected networks of 1994–1995, 1999–2000 and 2009–2010; and all the 21 networks have the same pattern. As the area's degree increases, the clustering coefficients weighted by migration attributes (number of returns, exemptions, and income) decline, and the clustering coefficient weighted by distance has an initial increase but decline even faster.

The weights make the network more clustered. The declining trend implies that the low connectivity areas have high clustering while high connectivity areas have lower clustering, a

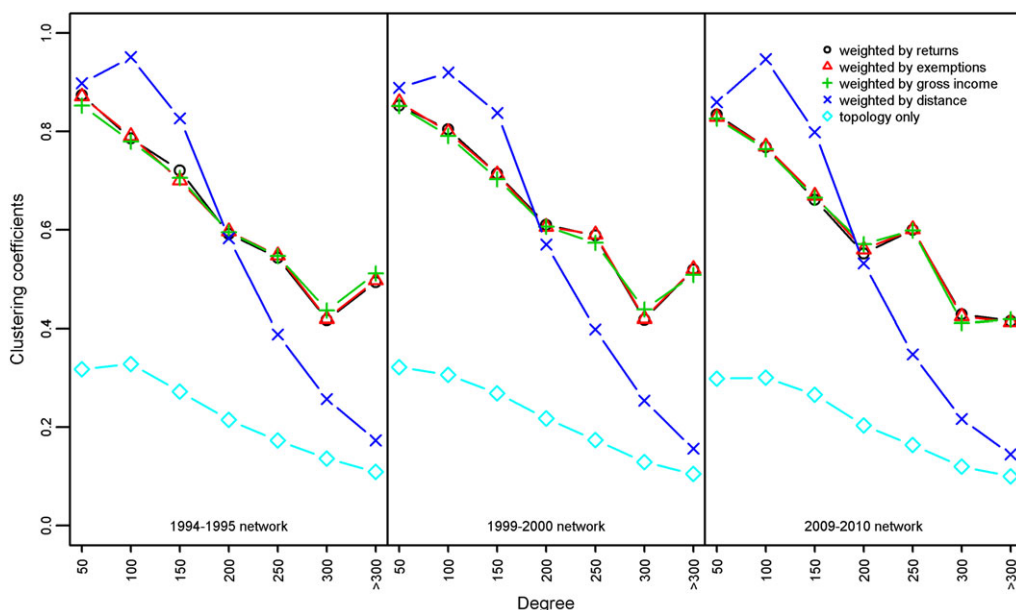


Fig. 5. The declining clustering coefficients as the degree of vertices increases

typical feature of hub-and-spoke network structure. The 200 appears as a bifurcation point in connection degree between decline trend of distance-weighted clustering coefficient and returns- (or exemptions- or income-) weighted clustering coefficient. For areas with less than 200 connection degree, their long-distance-connected areas tend to be connected. For areas with greater connection degree, their high-volume-connected areas tend to be connected.

3.4 Low connectivity areas tend to connect high connectivity areas

As networks are about the connections, it is interesting to know what kinds of vertices connect to what other kinds of vertices. In the inter-area population migration networks, we wonder what areas connect to what other areas in terms of the similarity of their connectivity, which is a characteristic of networks called assortativity (Newman 2002, 2003a). Social networks are often assortative, in which similar people are more likely to connect to each other, a phenomena termed as homophily (Kandel 1978; McPherson et al. 2001), while many technological or biological networks tend to be disassortative, that is, highly connected vertices tend to connect to loosely connected vertices. Newman et al. (2002, 2003a) and Newman and Girvan (2003) define a measure of the network assortativity based on the topological properties of the vertices (e.g., the connection degree, strength, betweenness, etc.) as well as non-topological attributes of the vertices. The measure is a variant of Pearson correlation coefficient, and can be formulized as, $r = \frac{1}{\sigma_q^2} \sum_{jk} jk(e_{jk} - q_j q_k)$, where e_{jk} is the joint excess degree probability for excess degrees of vertex j and k (the excess degree of a vertex is equal to one less than its total degree), $q_k = (k+1) p_{k+1} / \sum_j j p_j$ is the normalized distribution of the excess degree, and σ_q^2 is the variance of the normalized distribution, q_k (Newman 2002; Noldus and Van Mieghem 2015). The assortativity coefficient only provides a global characterization of the networks. Local assortativity (i.e., assortativity of each vertex) can be measured by the average nearest neighbours' degree or the weighted average nearest neighbours' degree. The average nearest neighbours' degree can be represented as, $k_{nn,i} = \sum_{j \in v(i)} d_j / d_i$, where d_j is the degree of vertex j , a vertex in vertex i 's neighbourhood, $v(i)$. For weighted networks, the $k_{nn,i}$ can be weighted by the edge's weights. The weighted average nearest neighbours' degree can be then represented as $k_{nn,i}^w = \sum_{j \in v(i)} w_{ij} d_j / s_i$, where w_{ij} is the weight between vertex i and j and s_i is the vertex i 's strength. A scatterplot between vertices' degree versus their average nearest neighbours' degree can reveal the detailed local degree correlation structure, which is a technique similar to the Moran scatterplot in geography (Anselin 1996).

For every inter-area population migration network, we calculate the average nearest neighbours' degree (and its weighted variants by considering the number of returns, number of exemptions, aggregate adjusted gross income, and geographic distance as weights) for each area in the network, and examine how the average nearest neighbours' degree vary as the area's connection degree increases. To see the general trend, the means of the average nearest neighbours' degrees of areas in every 50-degree interval are plotted over the degree. Figure 6 shows how the average nearest neighbours' degree vary over the connection degree for three selected networks for 1994–1995, 1999–2000, and 2009–2010. The same pattern is common to all the 21 networks. As the area's connection degree increases, both the topological and weighted average nearest neighbours' degrees increase to a certain degree and then decline. This pattern has been found in other networks (Newman 2002, 2003a; Barrat et al. 2004). It implied that the networks show assortative for areas with degree under around 250, and disassortative for areas with degree larger than around 250. However, the networks are essentially disassortative networks where the low-degree areas (lower than around 250) connect with high-degree areas and high-degree areas tend to connect with low-degree areas.

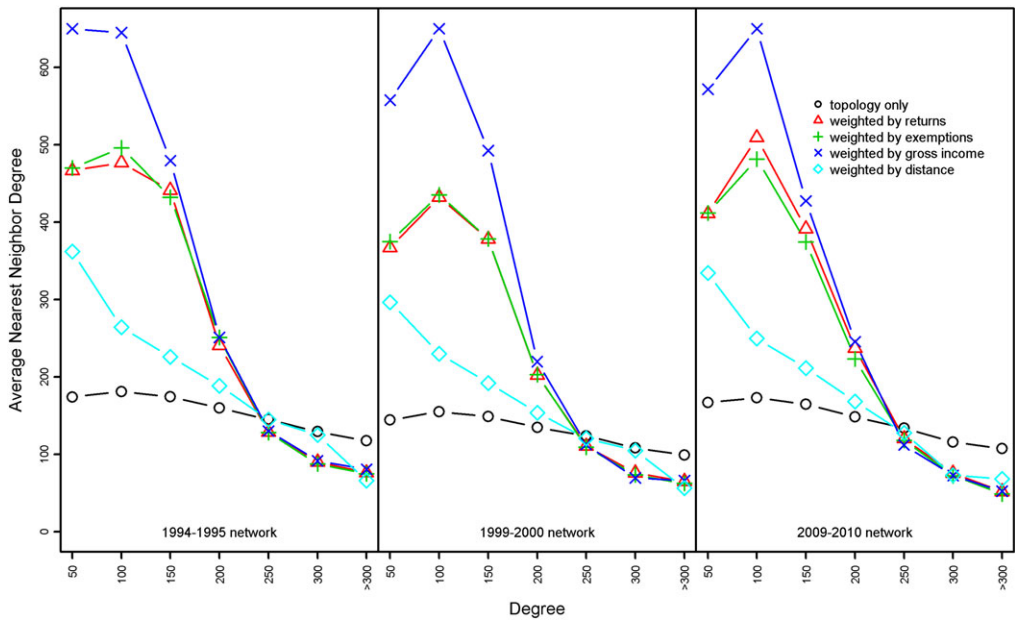


Fig. 6. Variation of topological and weighted average nearest neighbour degrees as the connection degree of vertices increase

3.5 Highly connected areas are more central

Vertices in social networks are considered being at central positions and therefore more influential when they are more connected or strategically located on the paths linking other vertices, or close to the rest of other vertices of the networks (Freeman 1979). Three centrality measures are commonly used, namely, degree, betweenness, and closeness centralities (Freeman 1977, 1979; Linton 1977; Freeman et al. 1991; Borgatti 2005). The degree centrality considers the connection degree of a vertex as a measure of the vertex's centrality. The closeness centrality measures how close a vertex to/from the rest of other vertices in the network, and it is defined as the inverse of the average shortest distance to/from all the rest of vertices in the network, $\theta_i^{close} = (N - 1) / \sum_{i \neq j} d_{i,j}$, where N is the number of vertices in the network, and $d_{i,j}$ represents the shortest distance between vertex i and j . The betweenness centrality measures to what degree a vertex (or edge) is located on the path between other vertices, and it is defined as the number of geodesic paths passing through a vertex (or edge), namely, $\theta_k^{betweenness} = \sum_i \sum_j p_{i \leftrightarrow k \leftrightarrow j} / p_{i \leftrightarrow j}$, $i \neq j \neq k$, where

$p_{i \leftrightarrow k \leftrightarrow j}$ represents the number of geodesic paths between vertices i and j that pass through vertex k , and $p_{i \leftrightarrow j}$ represents the total number of geodesic paths between vertices i and j .

In the population migration networks, the betweenness centralities of areas increase with area's connection degree (Figure 7). This implies that the highly connected areas have high betweenness, that is, these areas are located on the critical paths that connect the majority of the networks. However, the increasing pattern is different among the weighted and topological betweennesses. When the vertex's degree is less than around 200, the returns, exemptions, and income weighted betweennesses increase more than topological and distance-weighted betweennesses. For areas with degree greater than 200, their topological betweenness increases exponentially, while the returns, exemptions, and income weighted betweennesses increase slower, and the distance weighted betweenness does not increase as much except a surge at the degree of around 300. These differential patterns imply that the migration dynamics increases the betweenness of less connected counties while suppresses the betweenness of highly

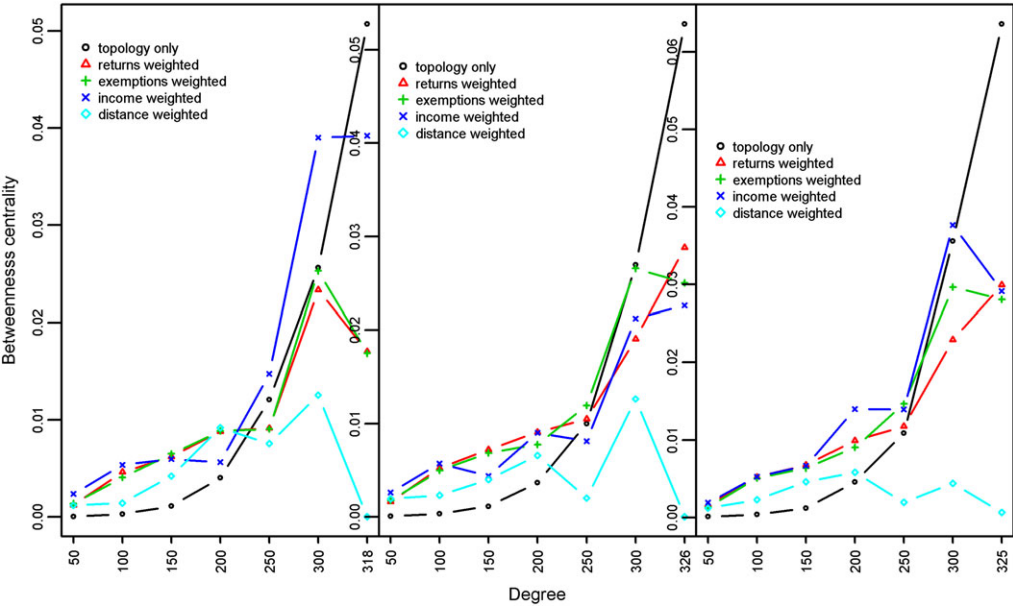


Fig. 7. Variation of betweenness centralities as the connection degree increases

connected counties. In other words, the migration dynamics make the betweenness of low-connection vertices higher than their topological betweenness, but the migration dynamics make the betweenness of high-connection vertices lower than their topological betweenness. The geographic distance reduces counties' betweenness the most.

The closeness centrality of an area characterizes how close the area is to the rest of other areas. The weighted closenesses show the same increasing pattern, a rapid increase when vertex's degree is less than 200 and a very slow increase afterwards (Figure 8). This is a pattern

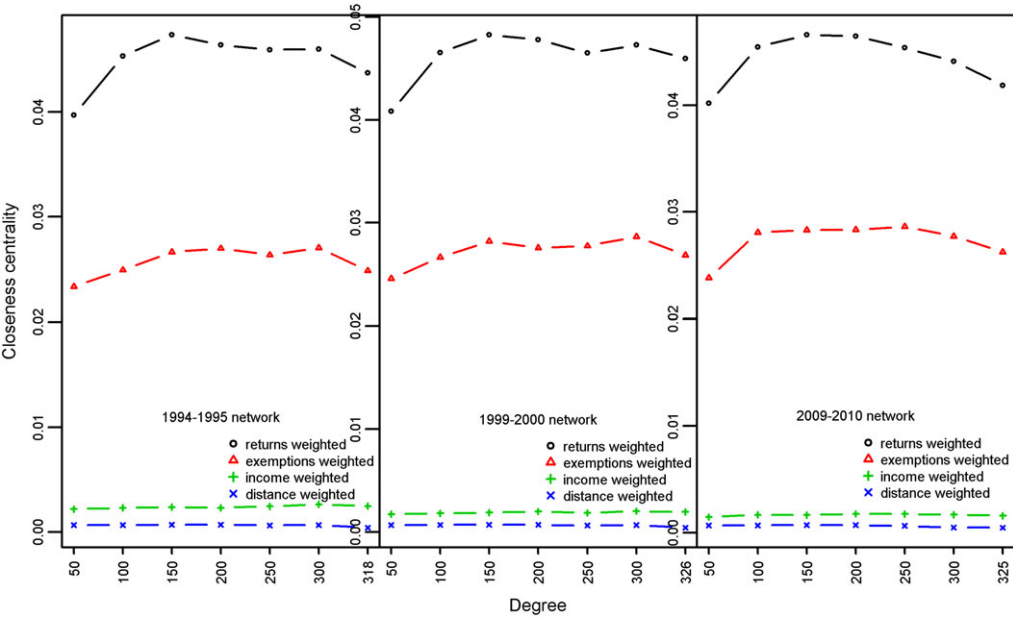


Fig. 8. Variation of closeness centralities as the connection degree increases

of highly connected networks, in which the vertices just need to take very few steps to reach most of the vertices when they have a reasonable number of connections.

4 Summary and conclusions

We constructed the population migration networks among BEA economic areas in the United States for every consecutive year from 1990 to 2011 from the IRS county-to-county migration data. We reported the violate one-year population migration dynamics among US counties. Since the US counties have several issues in regards use as the unit of migration analysis, we aggregated the county-to-county data to population migration networks among the BEA economic areas in the continental United States, and analysed the various aspects of the structural properties of the networks, and how the migration dynamics mediate the topological structure of the networks.

We find that the IRS county-to-county migration data are valuable to reveal the violate migration dynamics among US counties over one-year interval. Aggregated on the county-to-county migration data, the population migration networks among economic areas in the United States have tremendous heterogeneity or highly skewed distributions in the number of migration connections of areas, number of returns and exemptions, aggregated adjusted gross income, and distance of the inter-area migration. A small number of areas have exchanged migrants to a large number of other areas, and only a small number inter-area migration has exchanged large volume of migrants. Four aspects of the network structural properties were explored: connectivity, clustering, assortativity, and centrality. We particularly focused on how these properties have varied on areas of different connectivity, and how these properties have been mediated by the migration dynamics. Analysis on these structural properties has pointed to a fundamental structure of the inter-area population migration networks in the United States, that is, the hub-and-spoke dissortative structure in which a small number of highly connected areas exchange high volume and long distance migration while connecting to many less connected areas.

Recent advances in network science have provided approaches to study various aspects of the complex structures of the networks. This study has demonstrated how a few complex network approaches help to better understand the structural properties and how migration dynamics mediate the structural properties of the population migration networks among economic areas in the United States. When researchers modified the gravity model as constrained models or intervening opportunity models, and examined errors in migration models, they have realized that the complex interaction among places should not be overlooked to correctly and fully understand migration system. In addition to the network metric techniques, this study demonstrated the complex migration interaction in the US internal migration system in hope of informing more accurate migration model.

References

- Anselin L (1996) The Moran scatterplot as an ESDA tool to assess local instability in spatial association. *Spatial Analytical Perspectives on GIS* 111: 111–125
- Barabási A-L (2009) Scale-free networks: A decade and beyond. *Science* 325: 412–413
- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286: 509–512
- Barabási A-L, Frangos J (2014) *Linked: The new science of networks science of networks*. Basic Books, New York
- Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America* 101: 3747–3752

- Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U (2006) Complex networks: Structure and dynamics. *Physics reports* 424: 175–308
- Borgatti SP (2005) Centrality and network flow. *Social Networks* 27: 55–71
- Brown LA, Odland J, Golledge RG (1970) Migration, functional distance, the urban hierarchy. *Economic Geography* 46: 472–485
- Davis KF, D'Odorico P, Laio F, Ridolfi L (2013) Global spatio-temporal patterns in human migration: A complex network perspective. *PloS one* 8: e53723
- Engels RA, Healy MK (1981) Measuring interstate migration flows: An origin-destination network based on Internal Revenue Service records. *Environment and Planning A* 13: 1345–1360
- Fagiolo G, Mastroiello M (2013) International migration network: Topology and modelling. *Physical Review E* 88: 012812
- Fagiolo G, Reyes J, Schiavo S (2009) World-trade web: Topological properties, dynamics, and evolution. *Physical Review E* 79: 036115
- Fienberg SE (1970) An iterative procedure for estimation in contingency tables. *The Annals of Mathematical Statistics* 41: 907–917
- Fotheringham AS (1984) Spatial flows and spatial patterns. *Environment and Planning A* 16: 529–543
- Fotheringham AS, Webber M (1980) Spatial structure and the parameters of spatial interaction models. *Geographical Analysis* 12: 33–46
- Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry* 40: 35–41
- Freeman LC (1979) Centrality in social networks conceptual clarification. *Social Networks* 1: 215–239
- Freeman LC, Borgatti SP, White DR (1991) Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks* 13: 141–154
- Frey WH, Liaw KL, Wright R, White MJ (2005) Migration within the United States: Role of race-ethnicity [with comments]. In: *Brookings-Wharton papers on urban affairs*. Brookings Institution Press, Washington, DC, pp. 207–262
- Gastner MT, Newman ME (2006) Shape and efficiency in spatial distribution networks. *Journal of Statistical Mechanics: Theory and Experiment* 2006: P01015
- Goodchild M, Smith T (1980) Intransitivity, the spatial interaction model, US migration streams. *Environment and Planning A* 12: 1131–1144
- Greenwood MJ (1985) Human migration: Theory, models, empirical studies. *Journal of Regional Science* 25: 521–544
- Greenwood MJ (1988) Changing patterns of migration and regional economic growth in the US: A demographic perspective. *Growth and Change* 19: 68–86
- Gross E (2003) US population migration data: Strengths and limitations. Internal Revenue Service Statistics of Income Division, Washington, DC. URL: http://www.irs.gov/pub/irs-soi/99gross_update.doc
- Haggett P, Chorley RJ (1969) *Network analysis in geography*. Edward Arnold, London
- Haggett P, Cliff AD, Frey A (1977) *Locational analysis in human geography*. Wiley, New York
- Holland SC, Plane DA (2001) Methods of mapping migration flow patterns. *Southeastern Geographer* 41: 89–104
- Kandel DB (1978) Homophily, selection, socialization in adolescent friendships. *American Journal of Sociology* 84: 427–436
- Lin J, Ban Y (2014) The evolving network structure of US airline system during 1990–2010. *Physica A: Statistical Mechanics and its Applications* 410: 302–312
- Linton CF (1977) A set of measures of centrality based on betweenness. *Sociometry* 40: 35–41
- Long LE (1988) *Migration and residential mobility in the United States*. Russell Sage Foundation, New York
- Manson GA, Groop RE (2000) US inter-county migration in the 1990s: People and income move down the urban hierarchy. *The Professional Geographer* 52: 493–504
- Masser I, Brown PJ (1977) Spatial representation and spatial interaction. *Papers in Regional Science* 38: 71–92
- Massey DS, Arango J, Hugo G, Kouaouci A, Pellegrino A, Taylor JE (1993) Theories of international migration: A review and appraisal. *Population and Development Review* 19: 431–466
- May R, Lloyd MAL (2001) Infection dynamics on scale-free networks. *Physical Review E* 64: 066112
- McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27: 415–444
- Mueser P (1989) The spatial structure of migration: An analysis of flows between states in the USA over three decades. *Regional Studies* 23: 185–200
- Newman ME (2002) Assortative mixing in networks. *Physical Review Letters* 89: 208701
- Newman ME (2003a) Mixing patterns in networks. *Physical Review E* 67: 026126
- Newman ME (2003b) The structure and function of complex networks. *SIAM review* 45: 167–256
- Newman ME, Girvan M (2003) Mixing patterns and community structure in networks. In: Pastor-Satorras R, Rubi J, Diaz-Guilera A (eds) *Statistical mechanics of complex networks*. Springer, Berlin, pp. 66–87
- Noldus R, Van Mieghem P (2015) Assortativity in complex networks. *Journal of Complex Networks* 3: 507–542
- Olsson G (1965) Distance and human interaction. A migration study. *Geografiska Annaler. Series B, Human Geography* 47: 3–43

- Pandit K (1997a) Cohort and period effects in US migration: How demographic and economic cycles influence the migration schedule. *Annals of the Association of American Geographers* 87: 439–450
- Pandit K (1997b) Demographic cycle effects on migration timing and the delayed mobility phenomenon. *Geographical Analysis* 29: 187–199
- Pandit K (2000) Regional variation in mobility levels and timing in the United States. *The Professional Geographer* 52: 483–493
- Pastor-Satorras R, Vespignani A (2002) Epidemic dynamics in finite size scale-free networks. *Physical Review E* 65: 035108
- Pedersen PJ, Pytlikova M, Smith N (2008) Selection and network effects: Migration flows into OECD countries 1990–2000. *European Economic Review* 52: 1160–1186
- Pierce K (2015) SOI migration data: A new approach. *Statistics of Income Bulletin* 34: 91–94
- Plane DA (1982) An information theoretic approach to the estimation of migration flows. *Journal of Regional Science* 22: 441–456
- Plane DA (1984) Migration space: Doubly constrained gravity model mapping of relative interstate separation. *Annals of the Association of American Geographers* 74: 244–256
- Plane DA (1999) Time series perspective and physical geography analogies in migration research. In: Pandit K, Withers SD (eds) *Migration and restructuring in the United States: A geographic perspective*. Rowman and Littlefield, Boston, pp. 313–335
- Plane DA, Jurjevich JR (2009) Ties that no longer bind? The patterns and repercussions of age-articulated migration. *The Professional Geographer* 61: 4–20
- Rae A (2009) From spatial interaction data to spatial interaction information? Geovisualisation and spatial structures of migration from the 2001 UK census. *Computers, Environment and Urban Systems* 33: 161–178
- Ravasz E, Barabasi AL (2003) Hierarchical organization in complex networks. *Physical Review E* 67: 026112
- Rayer S, Brown DL (2001) Geographic diversity of inter-county migration in the United States, 1980–1995. *Population Research and Policy Review* 20: 229–252
- Schachter J, Franklin RS, Perry MJ (2003) Migration and geographic mobility in metropolitan and nonmetropolitan America, 1995 to 2000. US Department of Commerce, Economics and Statistics Administration, US Census Bureau, Washington, DC
- Slater P (1984a) A partial hierarchical regionalization of 3140 US counties on the basis of 1965–1970 inter-county migration. *Environment and Planning A* 16: 545–550
- Slater P (1984b) Measuring migration fields of US counties. *Geographical Analysis* 16: 65–73
- Slater P (1985) Point-to-point migration functions and gravity model renormalization: Approaches to aggregation in spatial interaction modelling. *Environment and Planning A* 17: 1025–1044
- Tobler WR (1987) Experiments in migration mapping by computer. *The American Cartographer* 14: 155–163
- Watts DJ (2004) The ‘new’ science of networks. *Annual Review of Sociology* 30: 243–270
- Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393: 440–442
- Wolpert J (1967) Distance and directional bias in inter-urban migratory streams. *Annals of the Association of American Geographers* 57: 605–616
- Xu Z, Harriss R (2008) Exploring the structure of the US intercity passenger air transportation network: A weighted complex network approach. *GeoJournal* 73: 87–102
- Xu Z, Sui D (2009) Effect of small-world networks on epidemic propagation and intervention. *Geographical Analysis* 41: 263–282



Resumen. Este estudio construye las redes de migración de población entre áreas económicas en los Estados Unidos para cada año consecutivo desde 1990 a 2011, y examina sus propiedades estructurales y la dinámica de migración de la población. Se exploran varios aspectos de las propiedades estructurales de las redes, como la conectividad, la conglomeración, la asortatividad y la centralidad. Se encontró que estas propiedades estructurales están influidas por las dinámicas de migración y la distancia entre áreas, y que los patrones de diferentes propiedades estructurales entre áreas de diferente conectividad revelan la estructura de eje central y radios de las redes. Se pone de manifiesto que existe una tremenda complejidad en la conectividad y dinámica migratoria en el sistema migratorio interno de los Estados Unidos.

抄録: 本稿では、1990年から2011年までの毎年の米国の経済領域間の人口移動ネットワークを構築し、それぞれの構造的特性と人口移動のダイナミクスを検討する。コネクティビティ、クラスタリング、同類度(assortativity)、中心性など、ネットワークの構造的特性の様々な側面を検討する。これらの構造的特性は移動ダイナミクスと地域間の距離が媒介しており、コネクティビティの異なる地域における様々な構造的特性のパターンから、ネットワークがハブアンドスポーク型の構造であることがわかる。米国国内の人口移動システムにおける移動のコネクティビティとダイナミクスは非常に複雑であることは明らかである。