RESEARCH   **Open Access**

Complex network of United States migration

Batyr Charyyev and Mehmet Hadi Gunes*

*Correspondence:
mgunes@unr.edu
Department of Computer
Science and Engineering,
University of Nevada Reno,
Reno, NV, USA

Abstract

Economists and social scientists have studied the human migration extensively. However, the complex network of human mobility in the United States (US) is not studied in depth. In this paper, we analyze migration network between counties and states in the US between 2000 and 2015 to analyze the overall structure of US migration and yearly changes using temporal analysis. We aggregated network on different time windows and analyzed for both county and state level. Analyzing flow between US counties and states, we focus on the migration during different periods such as economic prosperity of the housing boom and economic hardship of the housing bust. We observed that nodes at county and state level usually remain active, but there are considerable fluctuations on links. This indicates that migration patterns change over the time. However, we could identify a backbone at both county and state levels using disparity filter. Finally, we analyze impact of the political and socioeconomic factors on the migration. Using gravity model, we observe that population, political affiliation, poverty, and unemployment rate have influence on US migration.

Introduction

Research has shown that there is a correlation between economic growth and net migration as people seek opportunities [1]. People typically move to places with greater economic opportunity than they were located. Post-World War II to 1980s, the United States (US) economy was driven by manufacturing and people were inclined to move areas of high manufacturing [2]. Since larger population areas usually had more manufacturing based jobs, migration was towards areas of high population [3]. In addition, people typically moved from North to South and East to West.

As factories have been relocated to countries with less manufacturing costs, there has been a steady decline of manufacturing jobs in the US since the 1980s. As US economy has become primarily driven on services and technology, the migration patterns have changed [4]. As new jobs are not centralized anymore, people are less likely to move in the last three decades than they have been historically. In some cases, depopulation of urban centers has been witnessed. As technology and service-based jobs are not concentrated in certain areas, people do not need to head to hubs.

Common migration patterns can be observed during times of economic prosperity such as booms and bubbling and economic hardship such as recessions. During economic prosperity, financially secure people move to places with desirable living

location and movement to certain hubs of job opportunities also increased [5]. On the other hand, during economic hardship, migration patterns are chaotic as people scatter towards different places and leave areas, which were once economically prosperous [6].

Certain factors pull people to certain locations and pulling factors include affordable housing, attractive climate, better employment opportunities, and family ties [7]. Other factors push people from their current location and pushing factors include higher taxes, unemployment, natural disasters, and low chances of marriage.

While researchers have analyzed international migration patterns as a complex network, to our knowledge, US migration patterns have not been analyzed as a complex network. Studies of international migration networks found out that (i) country migration network has a power law degree distribution; (ii) the world has become more interconnected and shows small world network characteristics; and (iii) there are communities among the countries, where people migrate.

In this paper, we analyze the network structure of US migration at county level and state level between 2000 and 2015. First, we analyze common network properties. We aggregate the data at different time windows to analyze emergence and robustness of network properties over the time. Then, we look at dynamic properties of the network. We analyze evolutionary dynamics at the node and link levels at different timescales and identify the structural backbone of migration in the US. Finally, we inspect political and socioeconomic factors affecting migration, and evaluate factors using gravity model of migration. In an earlier study [8], we analyzed the migration between 2004–2008 focusing on political aspects.

In the rest of the paper, we summarize related work in "Related work", present our methodology in "Methodology", perform detailed analysis of the US migration in "U.S. migration network", analyze temporal network characteristics in "Network dynamics", rank counties and states in "Central counties and states", and analyze political and socioeconomic factors in "Migration factors". Finally, we concluded in "Conclusions".

Related work

In this section, we provide an overview of related work on analysis of complex migration network, dynamic temporal networks, and socioeconomic factors affecting migration.

Researchers have analyzed interregional and international migration patterns. Kemper [9] analyzes migration between Eastern and Western Germany before and after unification. Niedomysl [10] analyzes interregional migration of Sweden focusing on how demographic, socioeconomic, and geographical aspects determine residential preferences. Davis et al. [11] analyzes global migration focusing on the correlation of migrant destination choice along with historical, cultural, and economic factors. Fagiolo and Mastrorillo [12] study correlation of migration and international trade from complex network perspective. Tranos et al. [13] analyzes international migration using gravity model and network-based regression techniques. Authors reveal the importance of physical and cultural proximity in migration along with the existence of both pull effects of prosperity and push effects of young population. Aleskerov et al. [14] analyzes central countries in the international migration. The common objectives in these analyses are: how the structure and topology of the network evolve with time? Is there clustering between specific

countries or regions? Which countries or regions are central in the migration network? These studies show the importance of socioeconomic, geographical, and political factors in shaping the migration network structure.

To analyze different social and economic factors, most studies use gravity model for migration. These studies focus on factors like infant mortality rate, education quality, political view, and income growth. Rayp and Ruyssen [15] extends gravity model of immigration focusing on economic and demographic factors in sub-Saharan African migration. Authors found that growth prospects and opportunities for employment and education are the main factors for migration rather than sociopolitical circumstances. Bergh et al. [16] analyzes effect of institutional quality and poverty on migration using a gravity model. They observed that migration decision depends on the expectations about future income levels and poor institutions act as a push factor, but absolute poverty in origin country limits migration. Similarly, [17] examined internal migration flows in New Zealand using the gravity model. Authors reveal that the deterrence effect of distance on migration increased in all periods until the 1996–2001 period and then decreased slightly in the 2001–2006 period. Furthermore, improvements in connectivity through reduced travel time have not increased migration flows.

As migration networks are not static, it is important to analyze the dynamic properties of the system to reveal important patterns. Holme and Saramäki [18] presents measures of dynamic networks including representation of dynamic networks as static graphs and models of temporal networks. Some studies have analyzed similar temporal networks. For instance, [19] analyzes dynamics of link utilization using records of communication in a large social network. They observed that roles of nodes change dramatically from day to day, and hence, authors assert that interventions targeting hubs will have significantly less effect than previously thought. Gautreau et al. [20] analyzed US airport network and observed that the links that disappear have similar properties as the ones that appear, and the disappearing and appearing links have a weight that is low on average but broadly distributed. They also reveal that links between airports with different traffic are very volatile. Bajardi et al. [21] examines dynamic patterns of cattle trade movements. They aggregate cattle trade movement data into different time windows and study temporal properties of the network. Authors observed that the nodes and links forming the backbone strongly vary depending on the time window and that the memory of the backbone rapidly fades away from one snapshot to the successive one. They also found that centrality of nodes fluctuate strongly over the time which hinders assessment of the spreading potential of premises.

Methodology

Data collection

We collected migration data among US counties between 2000 and 2015 from the Internal Revenue Service (IRS) data [22]. As each citizen needs to report their income and pay taxes, IRS collects addresses of citizens every year. IRS records address changes and provide county-level summaries, which we can use to calculate number of people migrating from one county to another each year.

We determined political affiliation of each county from the votes each party obtained during the presidential elections. We retrieved presidential election results for each county

from the United States Census, every 4 year when the presidential elections happen [23]. Note that local election or congressional election results would provide a better picture of political inclinations of people in a county at every 2 year, but those results were not readily available.

We also gathered the unemployment and poverty rate data from the US Department of Agriculture [24] and federal expenditure and crime rate from the US Census [23].

Data cleaning

To ensure consistency, collected data needed some cleaning. For the data between 2000 and 2012, IRS applied a threshold of 20 to report migration from one county to the other. Hence, if less than 20 people migrated between two counties, it was not reported in the data. However, between 2013 and 2015, the threshold was increased to 40. Therefore, we increased the threshold to 40 for all periods. In addition, there was a change in the code of six counties. For example, Shannon County, SD with a county code of 46–113 changed its name to Oglala Lakota County, AD with a county code of 46–102. We updated entries for these six counties to account for changes in county codes. Moreover, we removed outlier data points in four counties of Kansas, as one had a very high weighted in-degree, whereas the other three had high weighted out-degrees. These values were higher than county populations and cancelled each other. Hence, we removed these four data points as a reporting anomaly. Finally, poverty rate data were not available for 2000–2003. Thus, we excluded these 4 years from analyses of socioeconomic factors affecting migration.

Network construction

The collected data are cleaned and analyzed by constructing migration network between counties. We used 2010 FIPS State and County codes to match entries in different data sets.

In the migration network, nodes represent counties, and directed and weighted edges represent the population that has migrated from one county to another. Immigration and emigration are represented by in-degree and out-degree, respectively. We find the migration flux of a county by dividing the migration of the county by its population for each year. In our analyses for each snapshot, we only considered nodes with an in-degree or out-degree greater than zero, and named them active nodes, where only active edges are present in the network.

Aggregating the data over various times, window Δt enables us to create static snapshots that we can analyze with the usual techniques of network theory. In particular, we can create $16/\Delta t$ consecutive snapshots, as we have 16 years of data. By analyzing different values of Δt, we can identify network properties that are prone to changes with the value of Δt. In addition to $\Delta t=1$, we considered $\Delta t=2$ years and $\Delta t=4$ years which produce 8 and 4 aggregated networks, respectively. These periods enable us to analyze the effects of local elections with $\Delta t=2$ years and presidential elections with $\Delta t=4$ years. The 4 year periods also correspond to the housing boom between 2004 and 2007 and recession between 2008 and 2011.

U.S. migration network

In this section, we analyze networks aggregated at different timescales and levels (county and state) to identify properties that remain stable or change across the time and levels. Network size and characteristics depend on the time window Δt. Table 1 summarizes

Table 1 US migration network characteristics

	Nodes	Edges	Average degree	Avrg weighted degree	Diameter	Average path length	Clustering coefficient	Modularity	Assortativity
County level									
Δt=1 year (16 networks)									
Average	2.9×10^3	5.7×10^4	19.3	5×10^3	9.4	3.5	0.5	0.7	0.26
Variance	3.8×10^3	6.2×10^7	5.5	1.6×10^5	0.9	1.5×10^{-2}	3.7×10^{-4}	1.2×10^{-4}	9.9×10^{-5}
Δt=2 years (8 networks)									
Average	3×10^3	6.6×10^4	22.2	1×10^4	8.8	3.4	0.5	0.7	0.26
Variance	2.5×10^3	5.7×10^7	5	2.2×10^5	0.7	3.7×10^{-2}	2.2×10^{-4}	8.7×10^{-5}	4.9×10^{-5}
Δt=4 years (4 networks)									
Average	3×10^3	7.7×10^4	25.5	2×10^4	8	3.2	0.6	0.7	0.26
Variance	5.3×10^2	2.5×10^7	2.2	1.3×10^5	0	1.8×10^{-3}	9.4×10^{-5}	1.8×10^{-3}	7.4×10^{-5}
Δt=16 years (1 network)									
Average	3×10^3	10×10^4	33.386	8×10^4	8	3	0.6	0.7	0.26
Variance	–	–	–	–	–	–	–	–	–
State level									
Δt=1 year (16 networks)									
Average	51	1.7×10^3	34.4	2.9×10^5	2.7	1.3	0.8	0.3	0.8
Variance	0	1.4×10^4	5.4	6.5×10^8	0.2	2.2×10^{-3}	2×10^{-4}	2.7×10^{-4}	1.4×10^{-3}
Δt=2 years (8 networks)									
Average	51	1.8×10^3	36.6	5.9×10^5	2.2	1.3	0.8	0.3	0.9
Variance	0	7.8×10^3	3	1×10^9	0.2	1.2×10^{-3}	1.4×10^{-4}	3.7×10^{-4}	6.5×10^{-4}
Δt=4 years (4 networks)									
Average	51	1.9×10^3	38.6	1.2×10^6	2	1.2	0.8	0.3	0.9
Variance	0	1.5×10^3	0.6	3.3×10^8	0	2.4×10^{-4}	5.8×10^{-5}	3×10^{-4}	1.8×10^{-4}
Δt=16 years (1 network)									
Average	51	2.2×10^3	42.3	4.7×10^6	2	1.2	0.9	0.3	0.9
Variance	–	–	–	–	–	–	–	–	–

Average of the number of nodes, edges, avrg degree, avrg weighted degree, diameter, avrg path length, clustering coefficient, modularity, and assortativity for different *Δt* at county and state levels

the basic properties of networks for county and state levels. Results show that the US migration network has a small world network characteristic, as each network at each Δt has a high clustering coefficient and a low average path length. We observe that some of the 3112 counties are not part of the migration network for every year. Over 15 years, 71 counties have incoming migration of less than 40 people and 57 counties have outgoing migration of less than 40 people, while 47 of the counties are common in both sets. Network characteristics are similar for each Δt within the network level, but differ between the county and state. As Δt timescale increases, the number of edges increases by an order magnitude, but the number of nodes does not change much. This indicates that there is not much activation/deactivation of nodes, as they are active most of the time.

The number of edges and average weighted degree is lowest in 2014, while the average path length is the highest in both county and state levels (see Additional file 1: Table S1 and Additional file 2: Table S2). This indicates among the analyzed years, the least migration among counties occurred in 2014. Comparing the housing boom, i.e., 2004–2007, and housing bust, i.e., 2008–2011, periods, we observe that people migrated between more counties during the housing boom than the housing bust with similar overall network characteristics. Migration seems to have reduced between 2013 and 2015.
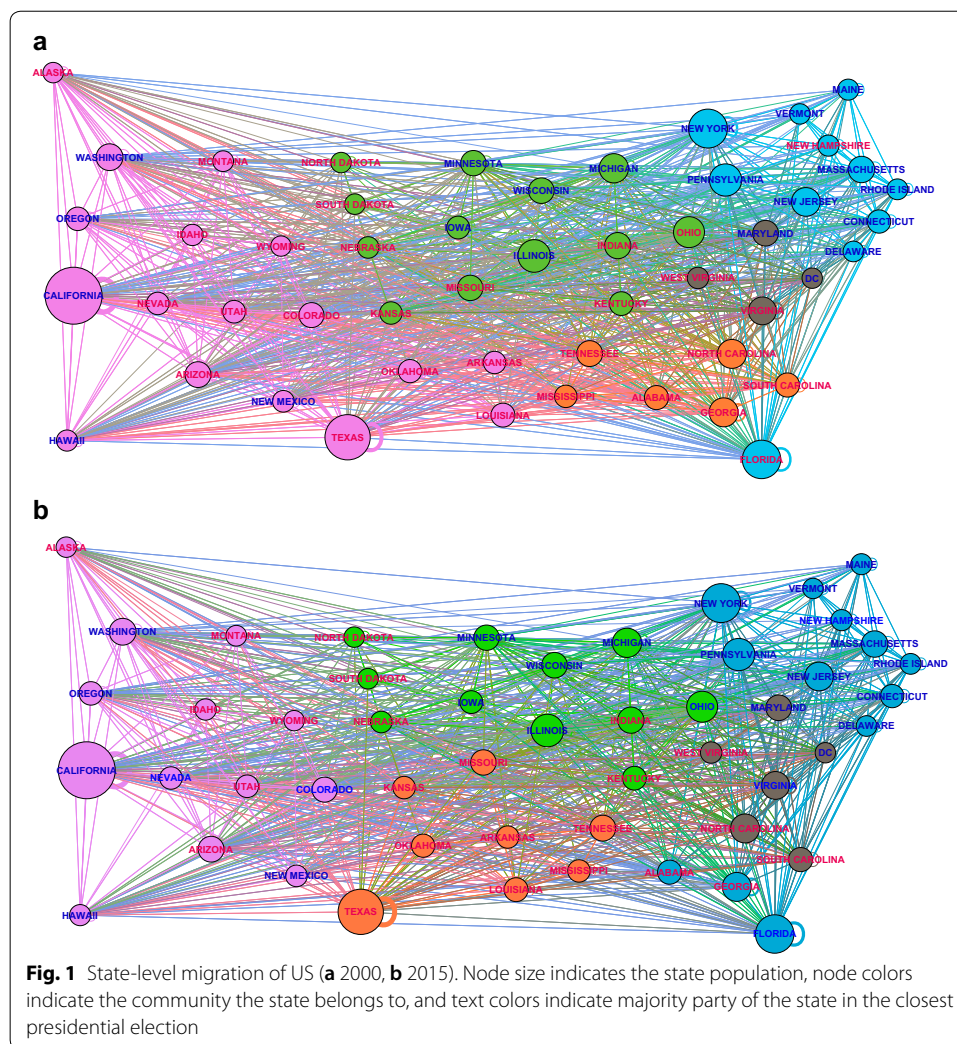
High modularity of the networks indicates that there are communities in each of the network. In addition, the assortativity of networks are slightly positive indicating a slight assortativity for all years with 2014 being the lowest. The assortative correlation between nodes indicates that people slightly prefer to move between counties of similar size rather than between counties of different sizes, i.e., large to small or small to large. State-level network has larger assortativity and smaller modularity compared to the county level, as it has a much denser network.

Analyzing migration over the 2000–2015 period, we detected 18 communities among counties using [25] algorithm implemented in gephi [26]. Additional file 3: Figure S1 presents a graph of the US migration network, where nodes represent US counties and are colored based on communities they belong. We observe that counties with the highest incoming population belong to the same community if they are geographically close by. Los Angeles, CA is in the same community with Clark, NV and Maricopa, AZ along with other counties in CA, whereas the top NY and TX counties are in their own community.

At the state level, we observe similar communities even when different periods are analyzed, as shown in Fig. 1, for years of 2000 and 2015. Additional file 4: Table S4 presents the communities of each state for each year. We observe that for each of the 16 networks, there are five communities among the states, but the community sets differ with each period, as some of the states belong to different communities during different years. For instance, some of the Midwest states are part of different communities. We observe that states are not segregated based on the political affiliation, but are clustered based on the geographic closeness.

We observe that 54.5% of edges are between counties of different states, whereas 29.3% of migration from the state is interstate (see Additional file 5: Table S3). That is, significant majority of migration is within the state, while there is a greater variation in the destination county in interstate migration. The only year, where majority of edges is within the state, is 2014 which indicates that there was less incentive to migrate to another state.

Fig. 1 State-level migration of US (**a** 2000, **b** 2015). Node size indicates the state population, node colors indicate the community the state belongs to, and text colors indicate majority party of the state in the closest presidential election

These results confirm that people migrate to geopolitically similar regions as reported in earlier studies of [10, 11]. In addition, US migration network has similar characteristics across the analyzed time periods, with some variation between the years that will be further analyzed.

An important characteristic to summarize a network is the degree distribution. Figure 2 presents the degree distribution for US County Migration network. We observe that overall shape of the degree distribution is an exponential distribution, and their distributions are similar across different time scales and periods (others not shown). We observe that, in general, people migrate from a larger number of counties than the number of counties they migrate to, and the difference increased during housing bust. Overall, the spread between in and out migration is higher during the housing boom, indicating an imbalanced flow among counties, whereas it is narrower during the housing bust, indicating a balanced flow to/from a county. Over the 15-year duration, the migration total of a county is sometimes greater than its population in the year 2000 (see Additional file 6: Figure S2-A). The counties with smaller inflow generally lose greater number of its population. In addition, migration to and from most of the counties is

Fig. 2 Degree distribution of US county migration network (**a** PDF, **b** CCDF). **a** The probability distribution function of the different periods and **b** the complementary cumulative distribution function
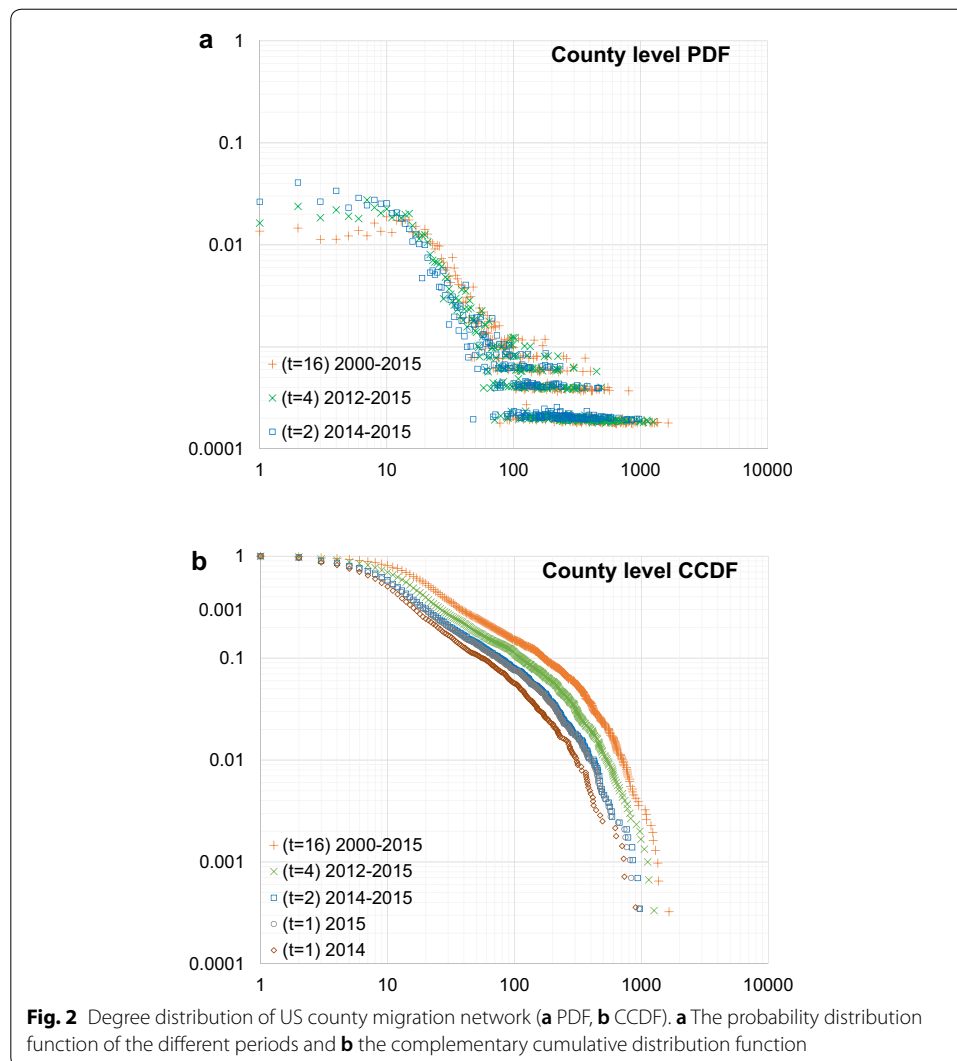
close to 0 for the 15 year duration, but largest counties have a larger net loss or gain through migration. For instance, CA-Los Angeles has the highest net loss, whereas CA-Riverside highest net gain.

Considering the state level, we observe that, over the 15 years, the total in and out migration of DC and in migration of Nevada is greater than its population of the year 2000 (see Additional file 6: Figure S2-B). Considering net migration over 15 years, the largest population gain by percentage is in Nevada and the largest loss is in New York states.

Network dynamics

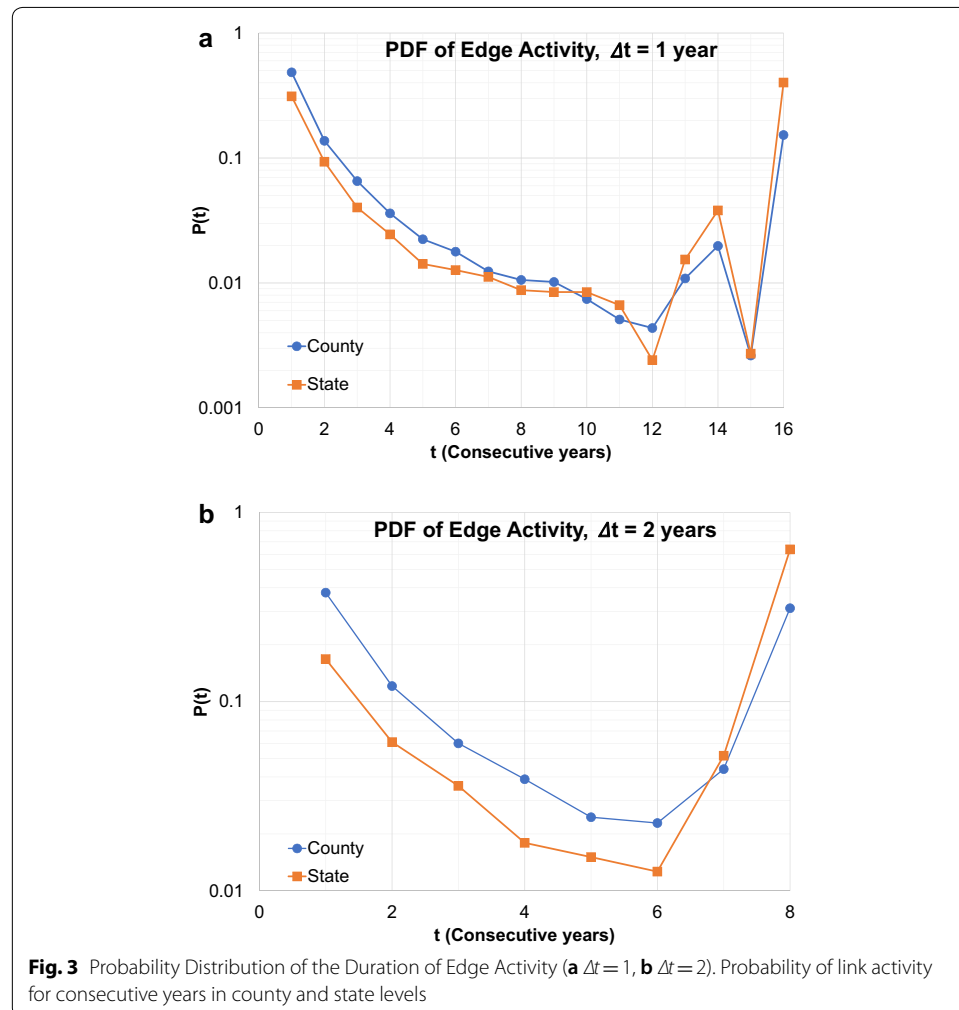Network dynamics can be characterized by activity and inactivity of nodes and edges along with how their properties change over the time once they are active. In the county-level network, nodes are usually active with a probability of 0.8 for being continuously active over the 15 year period. On the other hand, edges have a probability of 0.15 for being always active. Hence, in this section, we focus on the edge dynamics. We analyze
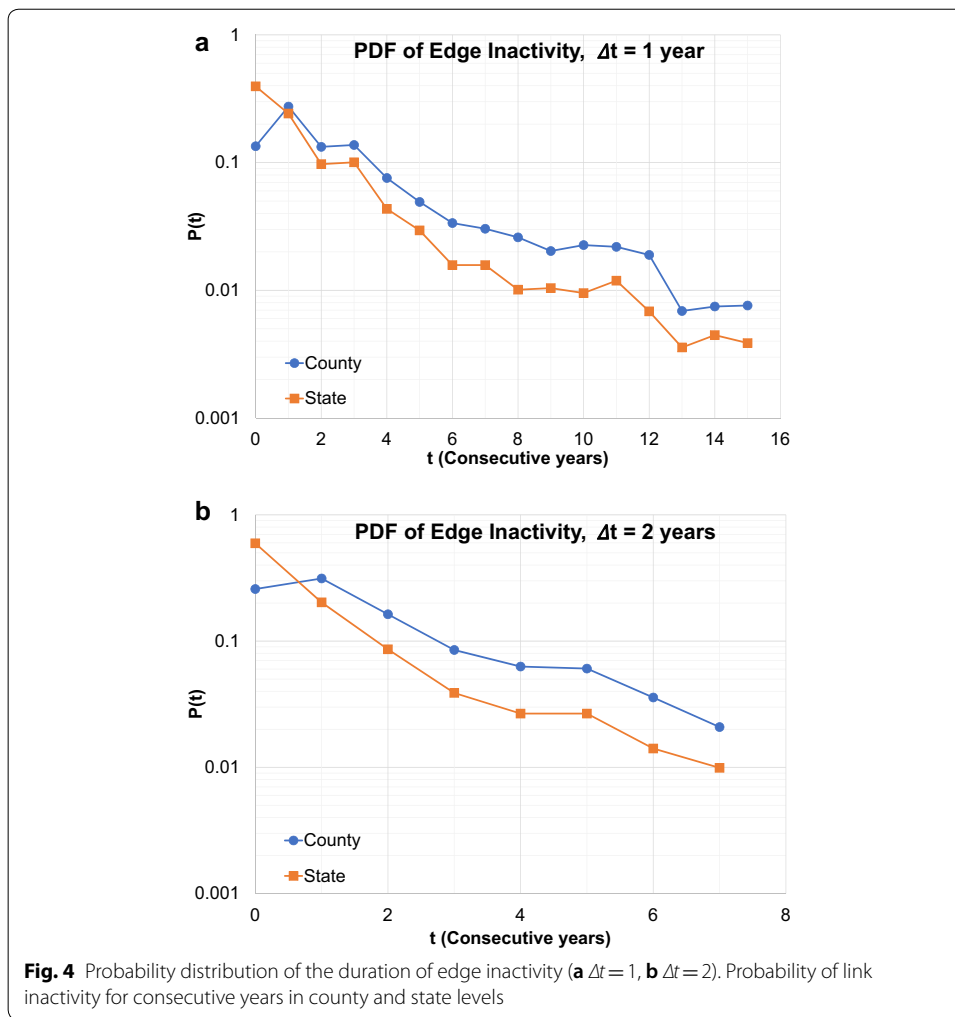
how links appear and disappear from the network for different time windows of Δt, and once they are active how their weight change.

Period of activity for a given value of Δt is defined as the number of consecutive time steps in which the link is active. Figures 3 and 4 show activity and inactivity distribution of links for different values of Δt, respectively. We observe that both distributions (activity and inactivity) in both levels (county and state) show similar behavior. When Δt is 1 year, there is a slight fluctuation of $P(t)$ when $t = 2014$ and 2015 because of the decline of migration in 2014 compared to other years. Some of the links, which were continuously active until 2014, change their state to inactive in 2014 and then back to active in 2015. With $\Delta t = 2$ years or higher, this slight fluctuation disappears among different networks with the same Δt.

In similar studies of cattle trade movements [21] and air transportation network [20], most of the nodes and links are continuously active or inactive, as they have daily data. Our network shows similar behavior in terms of activity and inactivity of links, as our data are yearly. Yet, there are considerable number of links that are active for long durations. In particular, the migration between counties/states that are central or

Fig. 3 Probability Distribution of the Duration of Edge Activity (**a** $\Delta t = 1$, **b** $\Delta t = 2$). Probability of link activity for consecutive years in county and state levels

Fig. 4 Probability distribution of the duration of edge inactivity (**a** $\Delta t = 1$, **b** $\Delta t = 2$). Probability of link inactivity for consecutive years in county and state levels

geographical close is always active. For example, Los Angeles, CA and Harris, TX are central nodes in the county-level network or Texas and California are central nodes in the state-level network, and they have migration between these pairs every year. Similarly, Washoe, NV–Pershing, NV counties Utah–Colorado pair are geographically close and have yearly migration between each pair.
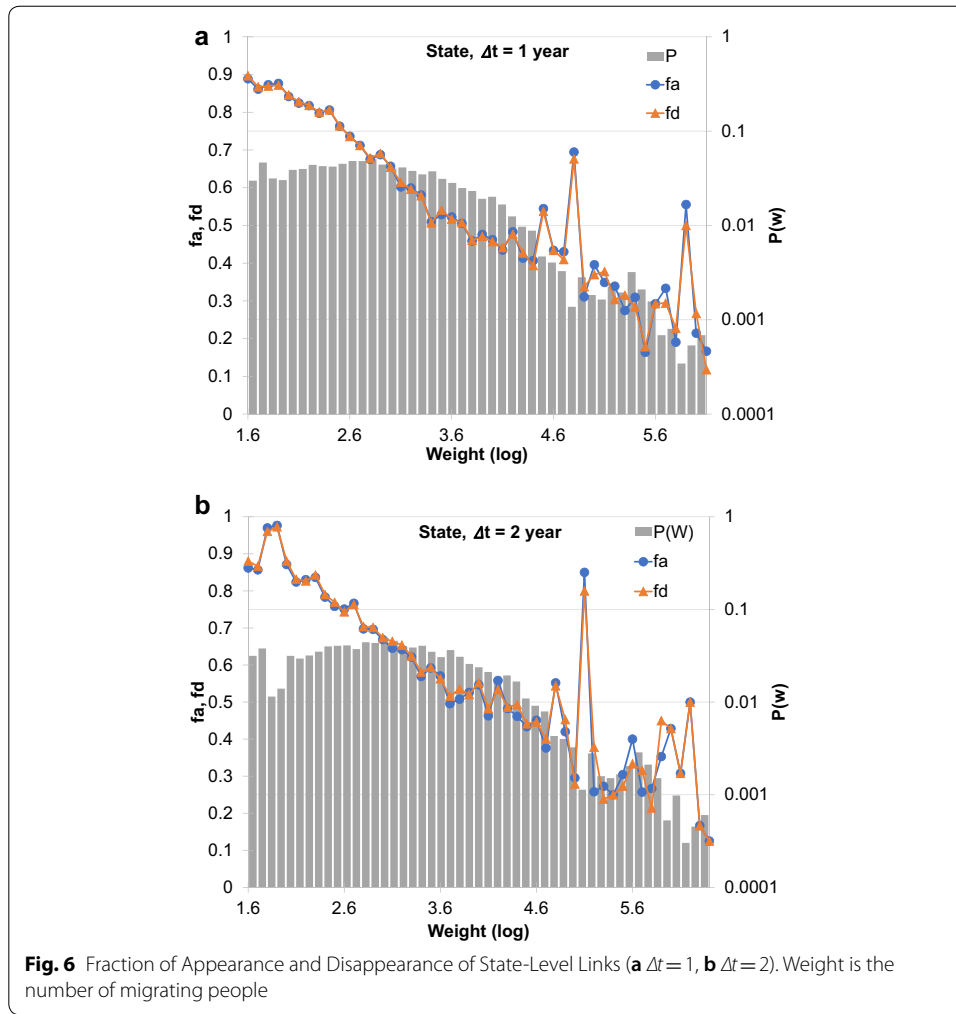
Since most of the links are active or inactive for short periods of time, we want to analyze the mechanism behind appearance and disappearance of links. We used mechanism proposed by [20] to evaluate the fraction of appearing f^a and disappearing f^d links as a function of their weight. This will enable us to see if there is a correlation between stability of links and the number of people migrating. Fraction of appearing links can be formulized as $f^a(w) = E^a(w|t)/E(w|t)$, where $E(w|t)$ is the number of links with weight w at time t and $E^a(w|t)$ is the number of such links that were not active at time t and thus appearing at time t. Similar logic is applied for fraction of disappearance f^d, where $E^d(w|t)$ is the number of links of weight w active at time $t-1$ but not active at t and $E(w|t)$ is similar as for f^a. Figures 5 and 6 provide results of f^a and f^d for county and state levels, respectively, for different Δt timescales.

Fig. 5 Fraction of appearance and disappearance of county-level links (**a** $\Delta t = 1$, **b** $\Delta t = 2$). Weight is the number of migrating people
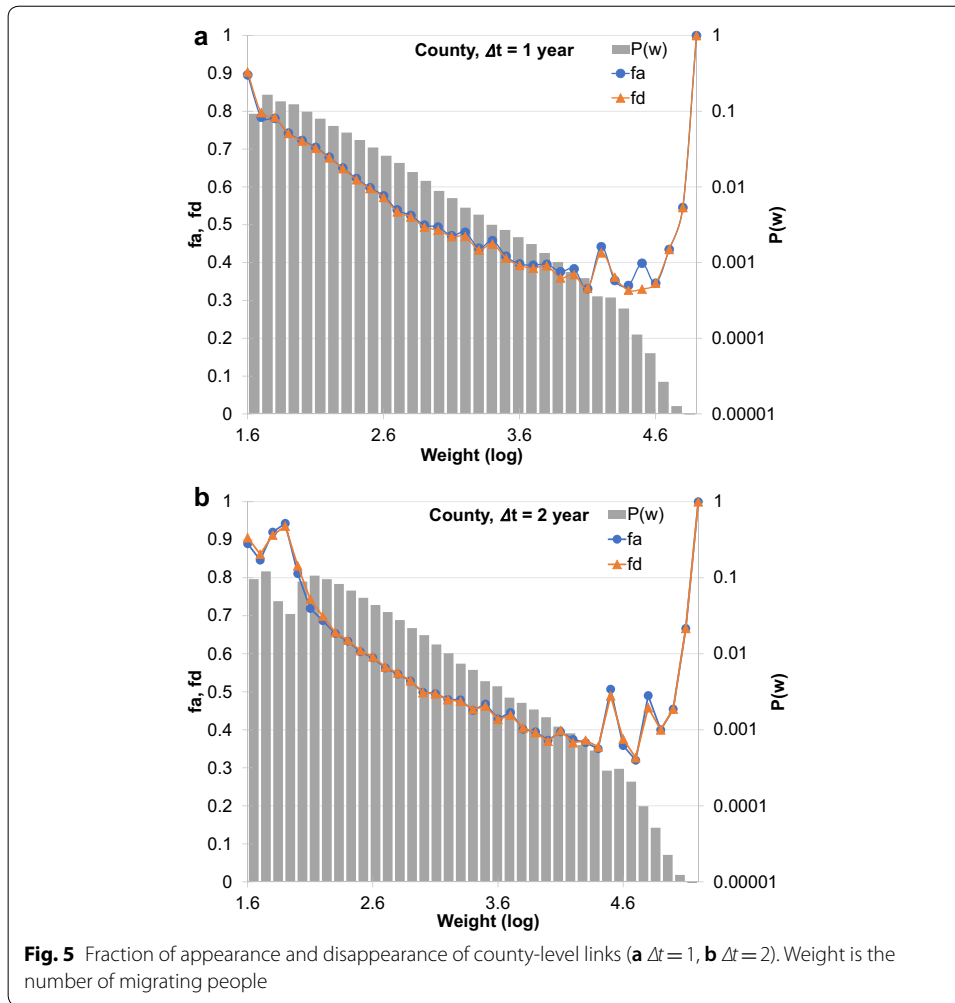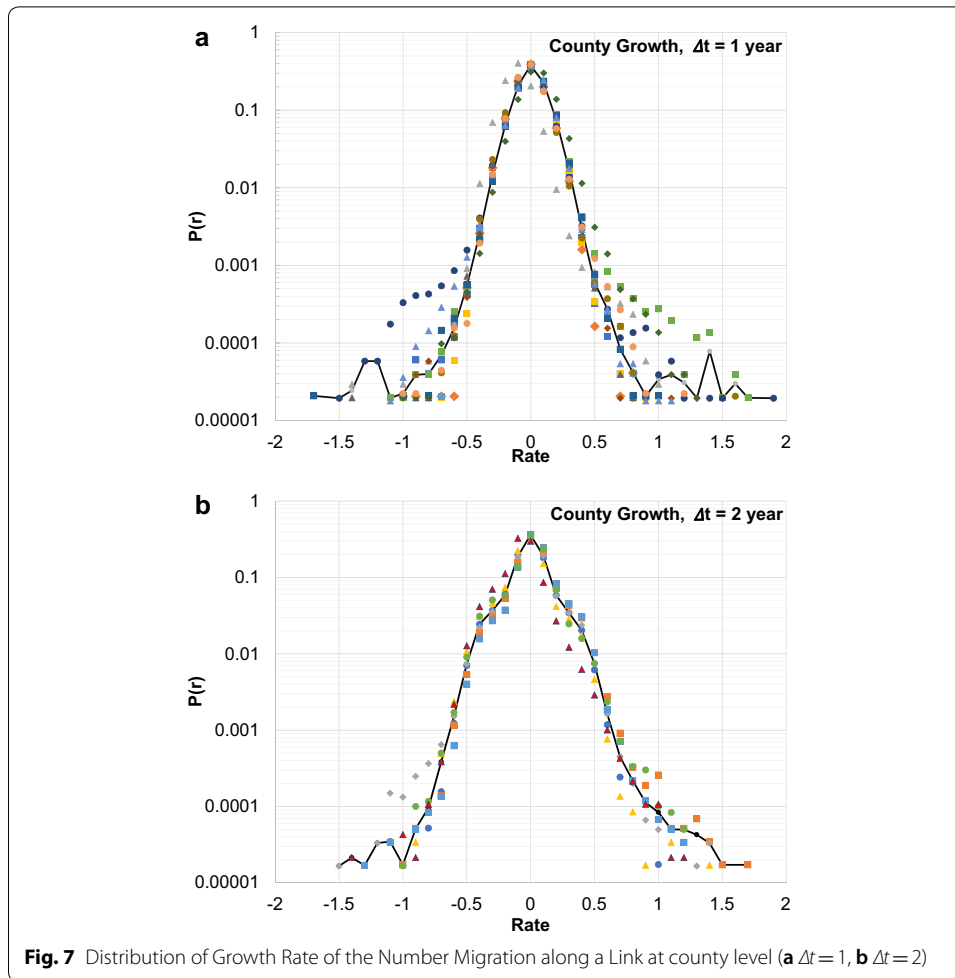
We observe that f^a and f^d have almost identical behavior similar to the cattle trade movements [21] and airline transportation network [20]. However, overall behavior of f^a and f^d differs from those studies. In the cattle trade network, links with intermediate weight are stable, whereas links with small or large weight are unstable because of different commercial forces and limitations on a premise's capacity. In air travel network, there is a positive correlation between links' stability and weight, as links with large weights are economically convenient and involve hubs that are not necessarily the source or destination of a traveler. Our migration analysis shows some similarity to the air travel. However, in migration network, links with very large weights show unstable behavior, whereas they were more stable in the air transportation network. This indicates that mass migration does not always happen between two counties and fluctuates over the years.

When we compare county and state-level networks, it seems that there is a correlation between stability of a link and its weight with slight fluctuations in high weight links at the state level. However, when we remove self-edges from analyses (see Additional file 7: Figure S3), we can see that county and state levels have identical behavior in terms of the fraction of appearance and disappearance of links.

Fig. 6 Fraction of Appearance and Disappearance of State-Level Links (**a** $\Delta t = 1$, **b** $\Delta t = 2$). Weight is the number of migrating people

In our previous analyses, we concluded that migration from one county to another may fluctuate in different years. Thus, we analyze evolution of link weight over the time utilizing growth rate $r_{ij}(t) = \log(w_{ij}(t+1)/w_{ij}(t))$ function [21]. From Figs. 7 and 8, we can observe that most of the weight increments are small, but sudden large increase or decrease of weights can be observed with a non-negligible probability. This behavior can be seen at different network levels (i.e., county and state) or different Δt timeframe. Similar results were obtained from analyses of firm growth [27], air transportation [20], and cattle trade movement [21].

We observed that links in a network fluctuate in terms of weight, as there is a strong instability of links with small weights. Hence, we want to identify the backbone of network if there is any. Global thresholding of links as a function of weight may give misleading results, because it may dismiss links with small weights which are locally very important [21]. Therefore, we utilized a disparity filter, which was introduced in [28], to identify backbone of the network. We created backbone of the network for each year (16 networks) for county and state levels. Then, we computed the overlap between $(16 \times 15)/2$ pairs of the backbones. Overlap of two networks is calculated as the ratio of intersection to union of edges, i.e., $| E_1 \cap E_2 |/| E_1 \cup E_2 |$.
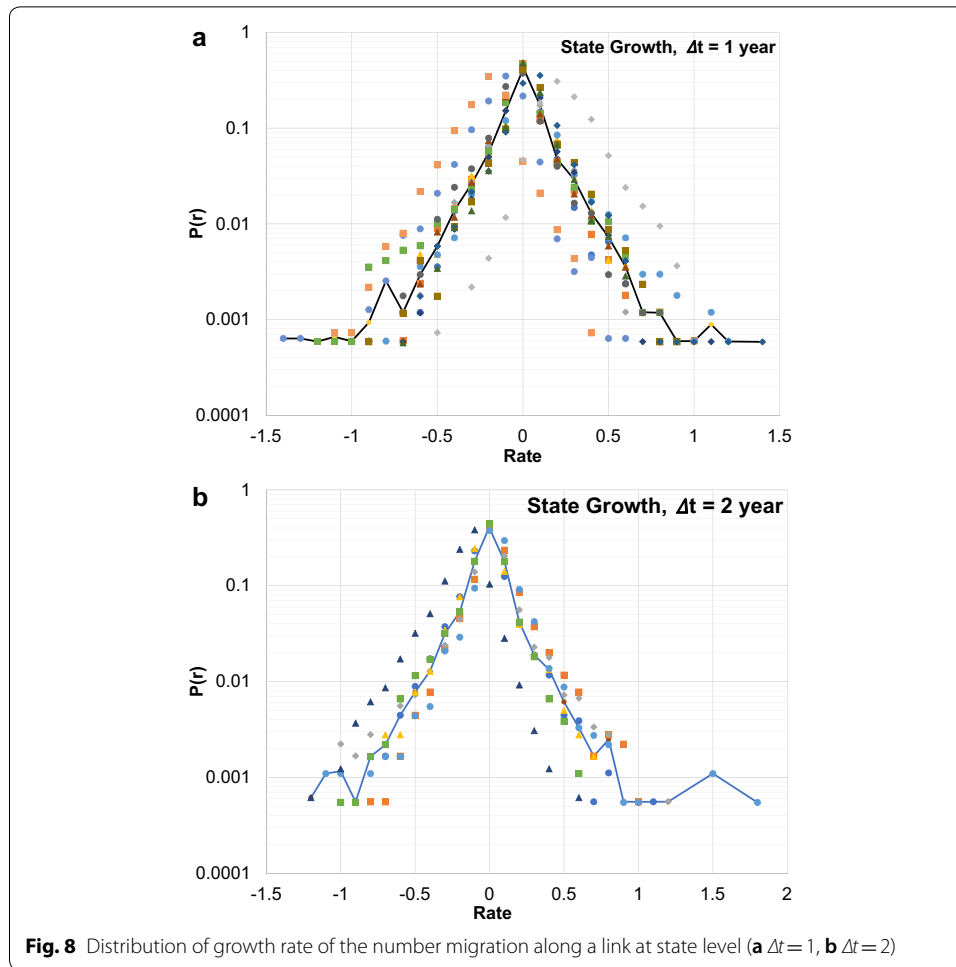
Fig. 7 Distribution of Growth Rate of the Number Migration along a Link at county level (**a** $\Delta t = 1$, **b** $\Delta t = 2$)

In a disparity filter, significance level of filtering can be configured with the α parameter and we repeated the analysis for different values of α. Tables 2 and 3 present overlaps of the backbones in color-coded matrix. Overlap of two successive backbones is around 80% for the county level and 90% for the state level. Moving away from diagonal may reduce this value to 65% for the county level and to 80% for the state level. These results indicate that there exists a backbone for migration network at both levels, even though the county-level links are not as stably active.

We determine the backbone networks, as the ones produced with $\alpha = 0.5$. Figure 9 presents the county- and state-level backbone networks, where node communities are colored with the same color.

Central counties and states

In this section, we compare different centrality metrics to determine and analyze top-ranked counties by each centrality metric. In particular, we compare population, weighted out-degree, weighted in-degree, degree, hub, eigenvector, PageRank, and betweenness centralities. Figure 10 compares the top 10 ranking of counties and states based on their populations in 2000, weighted out-degree, weighted in-degree, degree,

Fig. 8 Distribution of growth rate of the number migration along a link at state level (**a** $\Delta t = 1$, **b** $\Delta t = 2$)

hub, eigenvector, PageRank, and betweenness centralities considering the backbone identified in the previous section.

Betweenness centrality ranks nodes that are in between most of the shortest paths among node pairs. In our case, counties/states that are on the pathways of most migrations will be ranked highest. *Hub centrality* ranks nodes that are connected to authority nodes that receive links from hubs. Hence, a highly central hub links to highly central authorities, and vice versa. In our case, a high hub centrality indicates counties/states that have outgoing links to central counties receiving migration from many. *Eigenvector centrality* ranks nodes based on the centrality of nodes that they are connected to. In our case, a high-ranked county/state is connected to other high-ranked counties with incoming or outgoing migration. *PageRank centrality* adjusts the generous recognition of the eigenvector centrality, and ranks a node higher if it is linked from other important and stingy nodes or if it is highly linked.

Comparing the backbone network with the overall network ($\Delta t = 16$) (see Additional file 8: Figure S4-A), we observe overlap in the top five central counties. Except pagerank, most of the centralities shows similar ranking with higher ranks. Overall, we observe that sun-belt counties attract large inflow of migration are among the top 10 ranked counties. Even though Maricopa, AZ has relatively lower population

Table 2 Evolution of yearly county-level migration network backbones

α=0.5	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
2015	0.66	0.67	0.68	0.68	0.68	0.67	0.68	0.68	0.69	0.71	0.71	0.69	0.69	0.79	0.73	1.00
2014	0.58	0.59	0.60	0.60	0.59	0.58	0.59	0.59	0.61	0.64	0.62	0.59	0.58	0.75	1.00	0.73
2013	0.66	0.67	0.68	0.68	0.68	0.66	0.68	0.68	0.69	0.72	0.71	0.69	0.69	1.00	0.75	0.79
2012	0.70	0.70	0.70	0.70	0.71	0.71	0.72	0.73	0.73	0.72	0.74	0.77	1.00	0.69	0.58	0.69
2011	0.69	0.70	0.70	0.70	0.70	0.70	0.72	0.73	0.73	0.73	0.75	1.00	0.77	0.69	0.59	0.69
2010	0.70	0.71	0.71	0.71	0.72	0.71	0.73	0.75	0.76	0.77	1.00	0.75	0.74	0.71	0.62	0.71
2009	0.70	0.71	0.72	0.72	0.72	0.71	0.73	0.75	0.77	1.00	0.77	0.73	0.72	0.72	0.64	0.71
2008	0.71	0.71	0.72	0.72	0.73	0.73	0.75	0.77	1.00	0.77	0.76	0.73	0.73	0.69	0.61	0.69
2007	0.71	0.72	0.72	0.72	0.74	0.75	0.77	1.00	0.77	0.75	0.75	0.73	0.73	0.68	0.59	0.68
2006	0.71	0.72	0.73	0.73	0.75	0.76	1.00	0.77	0.75	0.73	0.73	0.72	0.72	0.68	0.59	0.68
2005	0.72	0.73	0.73	0.74	0.76	1.00	0.76	0.75	0.73	0.71	0.71	0.70	0.71	0.66	0.58	0.67
2004	0.73	0.75	0.75	0.77	1.00	0.76	0.75	0.74	0.73	0.72	0.72	0.70	0.71	0.68	0.59	0.68
2003	0.74	0.76	0.77	1.00	0.77	0.74	0.73	0.72	0.72	0.72	0.71	0.70	0.70	0.68	0.60	0.68
2002	0.75	0.77	1.00	0.77	0.75	0.73	0.73	0.72	0.72	0.72	0.71	0.70	0.70	0.68	0.60	0.68
2001	0.77	1.00	0.77	0.76	0.75	0.73	0.72	0.72	0.71	0.71	0.71	0.70	0.70	0.67	0.59	0.67
2000	1.00	0.77	0.75	0.74	0.73	0.72	0.71	0.71	0.71	0.70	0.70	0.69	0.70	0.66	0.58	0.66

Legend: 1.00 0.90 0.80 0.70 0.60 0.50 0.40

α=0.1	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
2015	0.65	0.66	0.67	0.67	0.68	0.66	0.68	0.67	0.69	0.71	0.70	0.65	0.66	0.79	0.71	1.00
2014	0.54	0.54	0.55	0.55	0.55	0.53	0.55	0.54	0.56	0.59	0.57	0.52	0.52	0.72	1.00	0.71
2013	0.64	0.64	0.65	0.65	0.66	0.64	0.66	0.66	0.68	0.70	0.69	0.64	0.64	1.00	0.72	0.79
2012	0.70	0.70	0.70	0.71	0.72	0.72	0.73	0.75	0.75	0.73	0.75	0.79	1.00	0.64	0.52	0.66
2011	0.68	0.69	0.69	0.69	0.70	0.71	0.72	0.74	0.74	0.73	0.75	1.00	0.79	0.64	0.52	0.65
2010	0.71	0.71	0.72	0.72	0.73	0.73	0.75	0.77	0.79	0.80	1.00	0.75	0.75	0.69	0.57	0.70
2009	0.71	0.71	0.72	0.73	0.73	0.72	0.75	0.76	0.79	1.00	0.80	0.73	0.73	0.70	0.59	0.71
2008	0.72	0.72	0.73	0.73	0.74	0.75	0.77	0.79	1.00	0.79	0.79	0.74	0.75	0.68	0.56	0.69
2007	0.73	0.74	0.74	0.75	0.77	0.78	0.80	1.00	0.79	0.76	0.77	0.74	0.75	0.66	0.54	0.67
2006	0.74	0.74	0.74	0.76	0.78	0.79	1.00	0.80	0.77	0.75	0.75	0.72	0.73	0.66	0.55	0.68
2005	0.73	0.74	0.75	0.77	0.79	1.00	0.79	0.78	0.75	0.72	0.73	0.71	0.72	0.64	0.53	0.66
2004	0.74	0.75	0.77	0.79	1.00	0.79	0.78	0.77	0.74	0.73	0.73	0.70	0.72	0.66	0.55	0.68
2003	0.76	0.79	0.81	1.00	0.79	0.77	0.76	0.75	0.73	0.73	0.72	0.69	0.71	0.65	0.55	0.67
2002	0.78	0.81	1.00	0.81	0.77	0.75	0.74	0.74	0.73	0.72	0.72	0.69	0.70	0.65	0.55	0.67
2001	0.80	1.00	0.81	0.79	0.75	0.74	0.74	0.74	0.72	0.71	0.71	0.69	0.70	0.64	0.54	0.66
2000	1.00	0.80	0.78	0.76	0.74	0.73	0.74	0.73	0.72	0.71	0.71	0.68	0.70	0.64	0.54	0.65

α=0.001	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
2015	0.62	0.64	0.64	0.64	0.64	0.63	0.64	0.63	0.65	0.66	0.66	0.60	0.60	0.76	0.64	1.00
2014	0.48	0.48	0.50	0.49	0.48	0.47	0.48	0.47	0.49	0.50	0.50	0.45	0.44	0.66	1.00	0.64
2013	0.61	0.62	0.62	0.62	0.63	0.62	0.63	0.62	0.64	0.66	0.65	0.59	0.58	1.00	0.66	0.76
2012	0.71	0.70	0.68	0.69	0.71	0.71	0.72	0.75	0.75	0.73	0.73	0.78	1.00	0.58	0.44	0.60
2011	0.68	0.67	0.67	0.67	0.68	0.70	0.70	0.72	0.74	0.72	0.72	1.00	0.78	0.59	0.45	0.60
2010	0.70	0.71	0.71	0.71	0.72	0.70	0.73	0.77	0.80	0.79	1.00	0.72	0.73	0.65	0.50	0.66
2009	0.69	0.70	0.70	0.71	0.72	0.70	0.73	0.76	0.79	1.00	0.79	0.72	0.73	0.66	0.50	0.66
2008	0.72	0.71	0.72	0.72	0.74	0.74	0.76	0.80	1.00	0.79	0.80	0.74	0.75	0.64	0.49	0.65
2007	0.73	0.73	0.72	0.73	0.76	0.77	0.79	1.00	0.80	0.76	0.77	0.72	0.75	0.62	0.47	0.63
2006	0.74	0.74	0.74	0.75	0.78	0.79	1.00	0.79	0.76	0.73	0.73	0.70	0.72	0.63	0.48	0.64
2005	0.74	0.74	0.75	0.76	0.80	1.00	0.79	0.77	0.74	0.70	0.70	0.70	0.71	0.62	0.47	0.63
2004	0.74	0.76	0.77	0.78	1.00	0.80	0.78	0.76	0.74	0.72	0.72	0.68	0.71	0.63	0.48	0.64
2003	0.77	0.79	0.79	1.00	0.78	0.76	0.75	0.73	0.72	0.71	0.71	0.67	0.69	0.62	0.49	0.64
2002	0.77	0.80	1.00	0.79	0.77	0.75	0.74	0.72	0.72	0.70	0.71	0.67	0.68	0.62	0.50	0.64
2001	0.80	1.00	0.80	0.79	0.76	0.74	0.74	0.73	0.71	0.70	0.71	0.67	0.70	0.62	0.48	0.64
2000	1.00	0.80	0.77	0.77	0.74	0.74	0.74	0.73	0.72	0.69	0.70	0.68	0.71	0.61	0.48	0.62

Table 3 Evolution of yearly state-level migration network backbones

α=0.5	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
2015	0.80	0.79	0.80	0.80	0.82	0.79	0.83	0.83	0.83	0.81	0.84	0.82	0.81	0.89	0.80	1.00
2014	0.70	0.71	0.71	0.72	0.71	0.68	0.72	0.70	0.72	0.72	0.72	0.70	0.67	0.82	1.00	0.80
2013	0.77	0.77	0.77	0.77	0.78	0.76	0.79	0.79	0.80	0.80	0.81	0.80	0.79	1.00	0.82	0.89
2012	0.83	0.82	0.83	0.79	0.83	0.82	0.85	0.87	0.87	0.83	0.88	0.91	1.00	0.79	0.67	0.81
2011	0.84	0.82	0.84	0.79	0.82	0.81	0.84	0.87	0.87	0.87	0.89	1.00	0.91	0.80	0.70	0.82
2010	0.84	0.84	0.86	0.84	0.84	0.81	0.86	0.87	0.89	0.89	1.00	0.89	0.88	0.81	0.72	0.84
2009	0.84	0.84	0.85	0.84	0.85	0.84	0.88	0.89	0.90	1.00	0.89	0.87	0.83	0.80	0.72	0.81
2008	0.86	0.85	0.86	0.82	0.85	0.86	0.91	0.93	1.00	0.90	0.89	0.87	0.87	0.80	0.72	0.83
2007	0.85	0.85	0.87	0.83	0.87	0.89	0.94	1.00	0.93	0.89	0.87	0.87	0.87	0.79	0.70	0.83
2006	0.86	0.84	0.87	0.85	0.90	0.90	1.00	0.94	0.91	0.88	0.86	0.84	0.85	0.79	0.72	0.83
2005	0.85	0.84	0.85	0.84	0.89	1.00	0.90	0.89	0.86	0.84	0.81	0.81	0.82	0.76	0.68	0.79
2004	0.87	0.88	0.89	0.91	1.00	0.89	0.90	0.87	0.85	0.85	0.84	0.82	0.83	0.78	0.71	0.82
2003	0.83	0.88	0.89	1.00	0.91	0.84	0.85	0.83	0.82	0.84	0.84	0.79	0.79	0.77	0.72	0.80
2002	0.87	0.90	1.00	0.89	0.89	0.85	0.87	0.87	0.86	0.85	0.86	0.84	0.83	0.77	0.71	0.80
2001	0.89	1.00	0.90	0.88	0.88	0.84	0.84	0.85	0.85	0.84	0.84	0.82	0.82	0.77	0.71	0.79
2000	1.00	0.89	0.87	0.83	0.87	0.85	0.86	0.85	0.86	0.84	0.84	0.84	0.83	0.77	0.70	0.80

Legend: 1.00 / 0.90 / 0.80 / 0.70 / 0.60 / 0.50 / 0.40

α=0.1	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
2015	0.81	0.84	0.85	0.84	0.81	0.81	0.82	0.84	0.83	0.84	0.86	0.82	0.81	0.90	0.85	1.00
2014	0.76	0.76	0.77	0.75	0.72	0.74	0.73	0.73	0.74	0.76	0.75	0.71	0.70	0.83	1.00	0.85
2013	0.80	0.80	0.80	0.80	0.80	0.80	0.78	0.82	0.83	0.84	0.86	0.81	0.82	1.00	0.83	0.90
2012	0.80	0.81	0.81	0.84	0.79	0.80	0.79	0.84	0.87	0.86	0.88	0.91	1.00	0.82	0.70	0.81
2011	0.81	0.83	0.83	0.88	0.81	0.84	0.82	0.85	0.87	0.87	0.88	1.00	0.91	0.81	0.71	0.82
2010	0.86	0.89	0.88	0.92	0.86	0.86	0.86	0.90	0.92	0.94	1.00	0.88	0.88	0.86	0.75	0.86
2009	0.83	0.86	0.86	0.89	0.83	0.85	0.84	0.88	0.93	1.00	0.94	0.87	0.86	0.84	0.76	0.84
2008	0.84	0.87	0.85	0.87	0.86	0.87	0.86	0.91	1.00	0.93	0.92	0.87	0.87	0.83	0.74	0.83
2007	0.86	0.86	0.85	0.88	0.88	0.90	0.89	1.00	0.91	0.88	0.90	0.85	0.84	0.82	0.73	0.84
2006	0.82	0.84	0.84	0.89	0.89	0.93	1.00	0.89	0.86	0.84	0.86	0.82	0.79	0.78	0.73	0.82
2005	0.87	0.88	0.89	0.90	0.94	1.00	0.93	0.90	0.87	0.85	0.86	0.84	0.80	0.80	0.74	0.81
2004	0.85	0.90	0.89	0.91	1.00	0.94	0.89	0.88	0.86	0.83	0.86	0.81	0.79	0.80	0.72	0.81
2003	0.88	0.91	0.92	1.00	0.91	0.90	0.89	0.88	0.87	0.89	0.92	0.88	0.84	0.80	0.75	0.84
2002	0.93	0.96	1.00	0.92	0.89	0.89	0.84	0.85	0.85	0.86	0.88	0.83	0.81	0.80	0.77	0.85
2001	0.94	1.00	0.96	0.91	0.90	0.88	0.84	0.86	0.87	0.86	0.89	0.83	0.81	0.80	0.76	0.84
2000	1.00	0.94	0.93	0.88	0.85	0.87	0.82	0.86	0.84	0.83	0.86	0.81	0.80	0.80	0.76	0.81

α=0.001	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
2015	0.91	0.91	0.91	0.92	0.89	0.88	0.92	0.92	0.94	0.95	0.95	0.98	0.95	1.00	0.97	1.00
2014	0.91	0.91	0.91	0.89	0.89	0.88	0.92	0.92	0.94	0.92	0.92	0.95	0.92	0.97	1.00	0.97
2013	0.91	0.91	0.91	0.92	0.89	0.88	0.92	0.92	0.94	0.95	0.95	0.98	0.95	1.00	0.97	1.00
2012	0.92	0.92	0.92	0.94	0.91	0.89	0.94	0.94	0.92	0.94	0.94	0.97	1.00	0.95	0.92	0.95
2011	0.89	0.89	0.89	0.91	0.88	0.86	0.94	0.94	0.95	0.94	0.94	1.00	0.97	0.98	0.95	0.98
2010	0.95	0.95	0.95	0.97	0.94	0.92	0.94	0.97	0.95	1.00	1.00	0.94	0.94	0.95	0.92	0.95
2009	0.95	0.95	0.95	0.97	0.94	0.92	0.94	0.97	0.95	1.00	1.00	0.94	0.94	0.95	0.92	0.95
2008	0.94	0.94	0.94	0.92	0.92	0.88	0.95	0.98	1.00	0.95	0.95	0.95	0.92	0.94	0.94	0.94
2007	0.95	0.95	0.95	0.94	0.94	0.89	0.97	1.00	0.98	0.97	0.97	0.94	0.94	0.92	0.92	0.92
2006	0.92	0.92	0.92	0.91	0.94	0.89	1.00	0.97	0.95	0.94	0.94	0.94	0.94	0.92	0.92	0.92
2005	0.91	0.91	0.91	0.92	0.92	1.00	0.89	0.89	0.88	0.92	0.92	0.86	0.89	0.88	0.88	0.88
2004	0.95	0.95	0.95	0.94	1.00	0.92	0.94	0.94	0.92	0.94	0.94	0.88	0.91	0.89	0.89	0.92
2003	0.98	0.98	0.98	1.00	0.94	0.92	0.91	0.94	0.92	0.97	0.97	0.91	0.94	0.92	0.89	0.92
2002	1.00	1.00	1.00	0.98	0.95	0.91	0.92	0.95	0.94	0.95	0.95	0.89	0.92	0.91	0.91	0.91
2001	1.00	1.00	1.00	0.98	0.95	0.91	0.92	0.95	0.94	0.95	0.95	0.89	0.92	0.91	0.91	0.91
2000	1.00	1.00	1.00	0.98	0.95	0.91	0.92	0.95	0.94	0.95	0.95	0.89	0.92	0.91	0.91	0.91

than other counties in 2010, it is the top county with respect to most of the centrality measures.

Comparing the housing boom (2004–2007) and housing bust (2008–2012) with the backbone network, we observe that weighted in- and out-degree ranking of counties changed between different periods except the top-ranked Los Angeles, CA. In addition, while in-degree ranking of some counties were higher during housing boom than overall

Fig. 9 Backbone of networks (**a** county, **b** state)

Fig. 10 Top 10 ranking of counties (**a**) and States (**b**) in the backbone based on different centralities
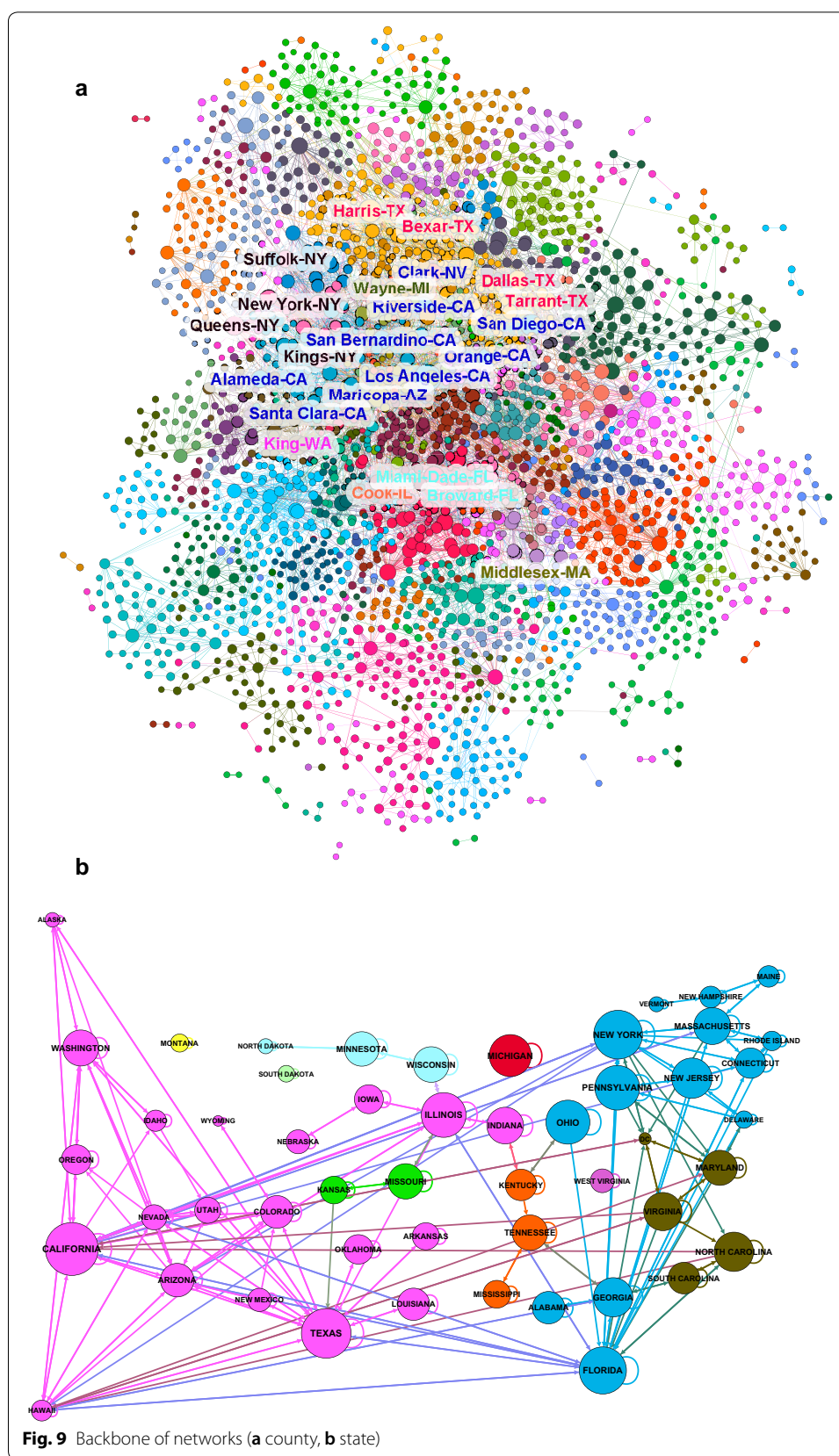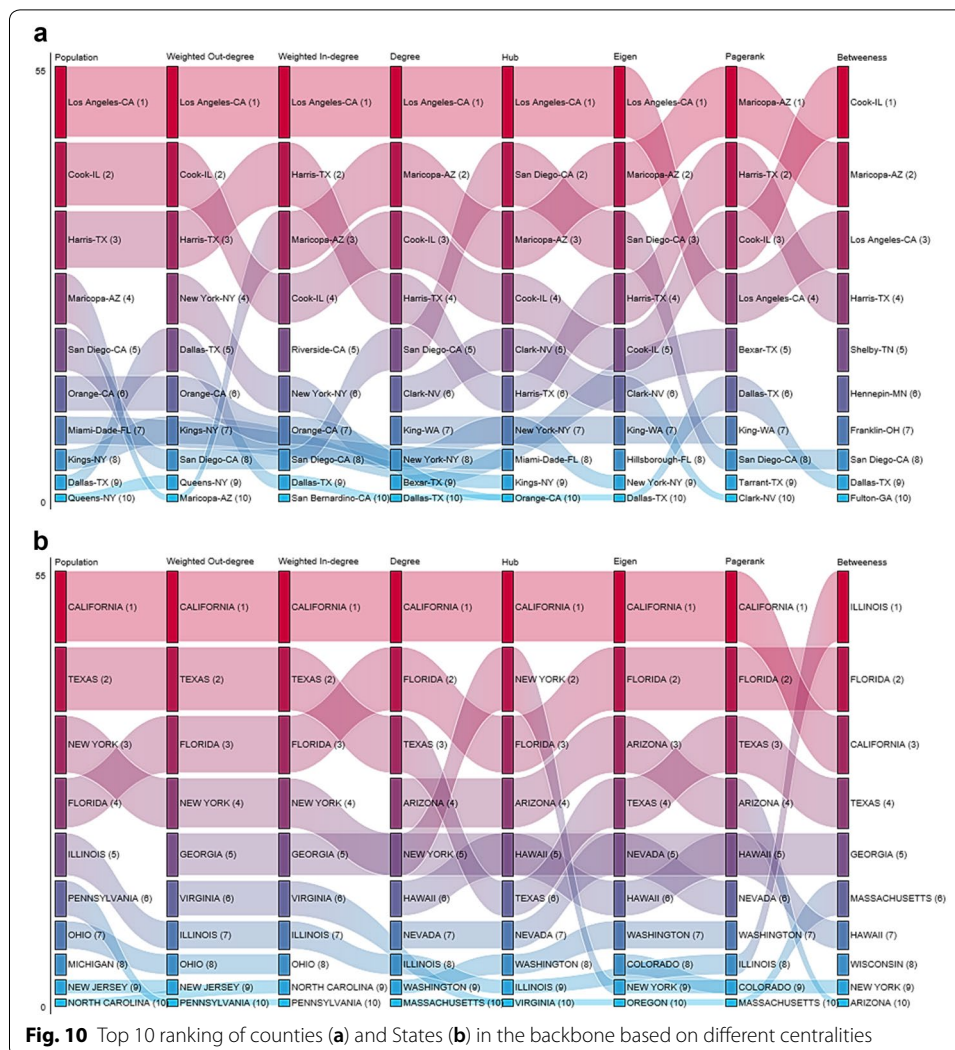
(e.g., San Diego, CA and Orange, CA), their rankings fell after housing boom. Housing bust seem to be beneficial to some counties (e.g., Maricopa, AZ), with higher in-degree and lower out-degree. We observe that except Maricopa, AZ, ranking of top between-ness counties change with the housing boom or bust. Except Maricopa, AZ, top six counties are in the same rank of hub centrality during different periods. Harris, TX has lower eigenvector centrality during housing boom, indicating fewer connections to cen-tral counties, while it has higher centrality during housing bust, indicating more connec-tions to central counties. A reverse effect is observed with Clark, NV. We observe that majority of the counties in top PageRank are counties that are in the sun-belt. Among top-ranked counties, Clark, NV loses most ranking with the housing bust.

Figure 10b presents the top 10 states based on the centralities considering backbone network of states identified in the previous section. Similar to backbone of county net-work, in state backbone, top ranks in centralities are shared by California, Texas, New York, Florida, and Arizona. These states have high ranking due to their counties with high centrality. For example, California is driven by Los Angeles county, while Texas is

driven by Harris county. Considering the overall network ($\Delta t = 16$) (see Additional file 8: Figure S4-B), New York State had the highest overall net population loss, even though it had the second highest incoming migration and fourth outgoing, and is not in top nine in any of the other centralities.

Migration factors

In this section, we analyze impacts of socioeconomic factors (such as unemployment, poverty, crime rate, and federal expenditure) and political factors on migration, and then perform a gravity model based analysis of identified factors.

Socioeconomic factors

We obtained the unemployment and poverty rate data from US Department of Agriculture [24] and federal expenditure and crime rate from US Census [23]. In our analysis, we calculated the difference of the unemployment and poverty rates at the destination county with respect to the source county. As the migration data are for a whole year we used the previous year's data for source county and adjusted for annual percentage changes.

Figure 11 presents the migration change between source and destination county, where each bar shows how many individuals migrated between counties with a difference of a given percentage of poverty. For instance, in 2005, 1,172,986 individuals migrated to another county with a poverty 1% less than their origin county and 1,130,399 migrated to a county with a 1% higher poverty. Overall, people moved to a county with 1.16% less poverty than their source county in 2005 and 2.47% less poverty in 2009. These results indicate that people had stronger motivation to move to counties that are more prosperous especially in 2009.

Figure 12 presents the migration change between source and destination county, where each bar shows how many individuals migrated between counties with a difference of a given percentage of unemployment. For instance, in 2005, 3,152,814 individuals migrated to another county with an unemployment 1% less than their origin county and 3,825,457 migrated to a county with a 1% higher unemployment. Overall, people moved to a county with 0.28% fewer unemployment in 2005 and 3.72% higher unemployment in 2009. This indicates, in 2009, that a strong consideration for migration between counties was unemployment rate of destination county compared to 2005, where people moved to counties with slightly fewer jobs than their origin county.

When crime rate of counties were analyzed, we observed that in 2005 and 2009, people migrated to a county with a 0.006% higher and 0.02% fewer crime rate than their origin county, on average. Similarly, when we analyzed federal expenditure per county, we observed that in 2005 and 2009, people migrated to a county with a 0.53% and 0.1% fewer expenditure than their origin county, on average. These results indicate that crime rate or federal expenditure of a county was not a major factor in people's decision for migration.

Fig. 11 Migration with respect to the poverty (**a** 2005, **b** 2009)
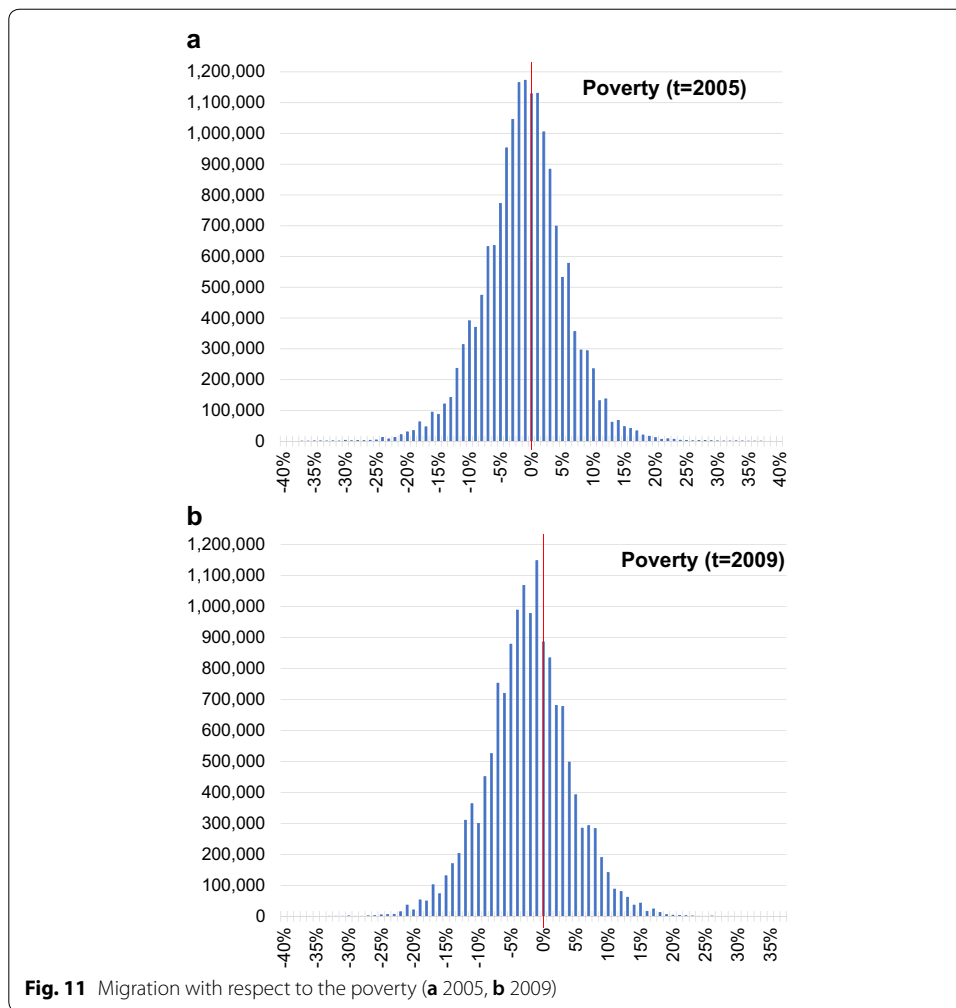
Political factors

We then analyze whether political affiliation of a county has a correlation with its migration. We utilized the presidential election results of 2004 and 2008, approximately at the beginning of housing boom and housing burst periods, obtained from US census [23]. Migration patterns of 2000 and 2012 were similar to the 2004 and 2008, respectively, as the same party had won the respective elections.

Table 4 shows the statistical significance of political affiliation and weighted in-degree of each county over the entire time frame. Results suggests that people are slightly attracted towards moderate Republican counties. This was expected; as in general, the housing boom was located in Republican counties. Another interesting observation is that being strongly politically affiliated (either Republican or Democratic) had a negative influence on the migration influx. This suggests that people tend to move towards more neutral areas as opposed to strong Republican or Democratic counties.

Figure 13 presents the flow of people in 2005 based on the Republican Party vote in the 2004 Presidential election, which was won by the Republican candidate. Each block shows the number of people that migrated from a county with a range of vote for the

Fig. 12 Migration with respect to the unemployment (**a** 2005, **b** 2009)
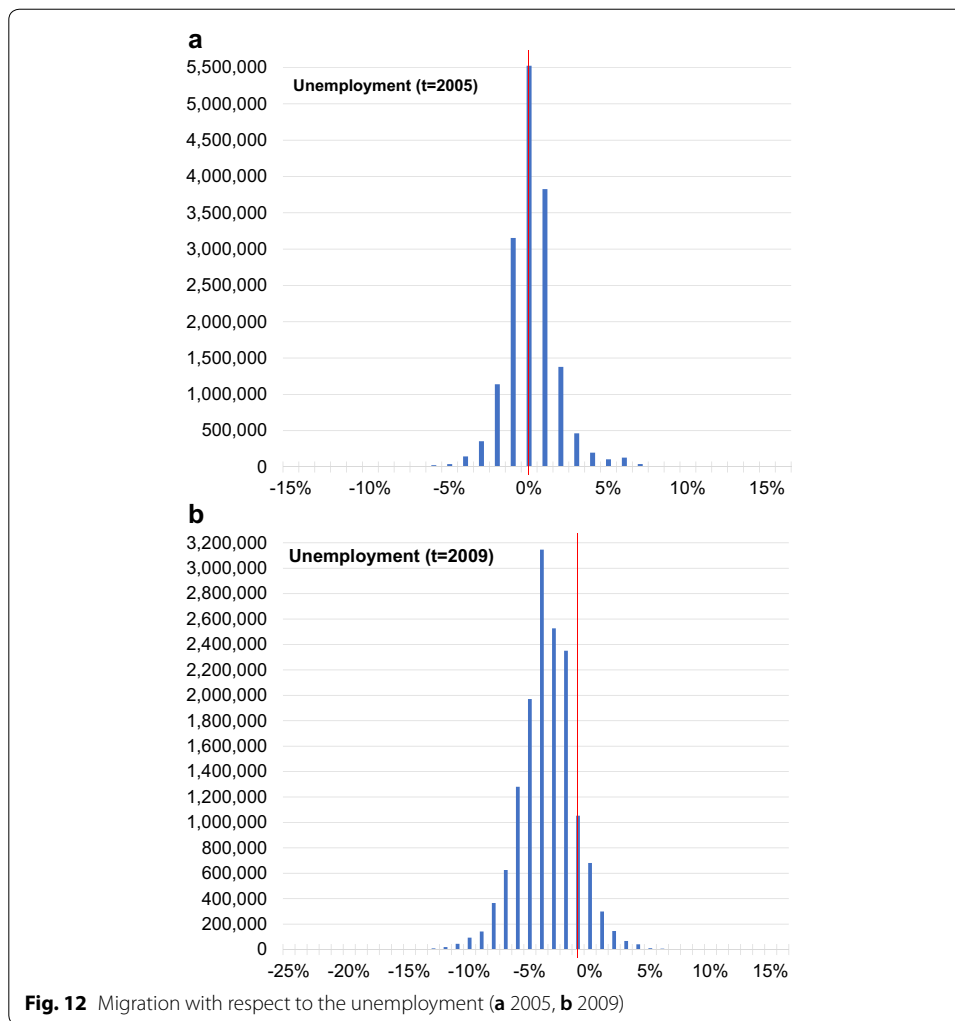
Republican Party. For instance, first block indicates that 1,144,070 people migrated in 2005 from counties with a Republican vote less than 25% in the 2004 Presidential election. Some of these people migrated to another county with a similar Republican vote and a slim portion of them moved to a significantly Republican county with over 75% Republican vote. Overall, 756,276 people moved to a county with less than 25% vote in 2005. Note that people that did not migrate are not counted in these figures. Similarly, Fig. 14 shows the flow of people in 2009 based on the Democratic Party vote in the 2008 Presidential election, which was won by the Democratic candidate. Compared to the prior election, we observe greater mobility for highly Democratic counties.
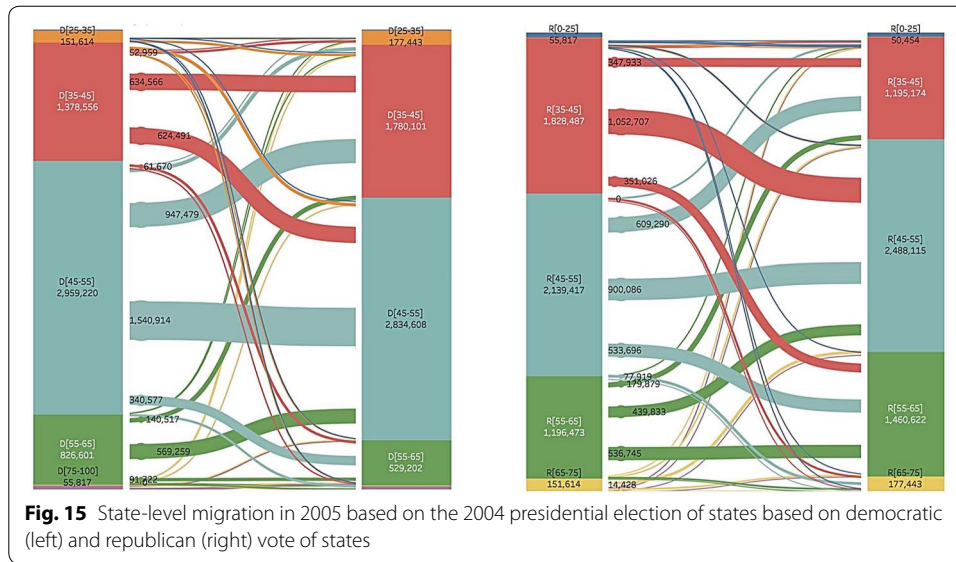
Table 4 Political affiliation and growth percentage

Percentile	> 25 (%)	25–35 (%)	35–45 (%)	45–55 (%)	55–65 (%)	65–75 (%)	< 75 (%)
Republican	− 0.03	− 0.04	− 0.05	0.01	0.24	0.06	− 0.22
Democratic	0.03	0.04	0.27	0.02	− 0.52	− 1.69	− 6.81

Fig. 13 County-level migration in 2005 based on the republican party vote in the 2004 presidential election

Fig. 14 County-level migration in 2009 based on the democratic party vote in the 2008 presidential election
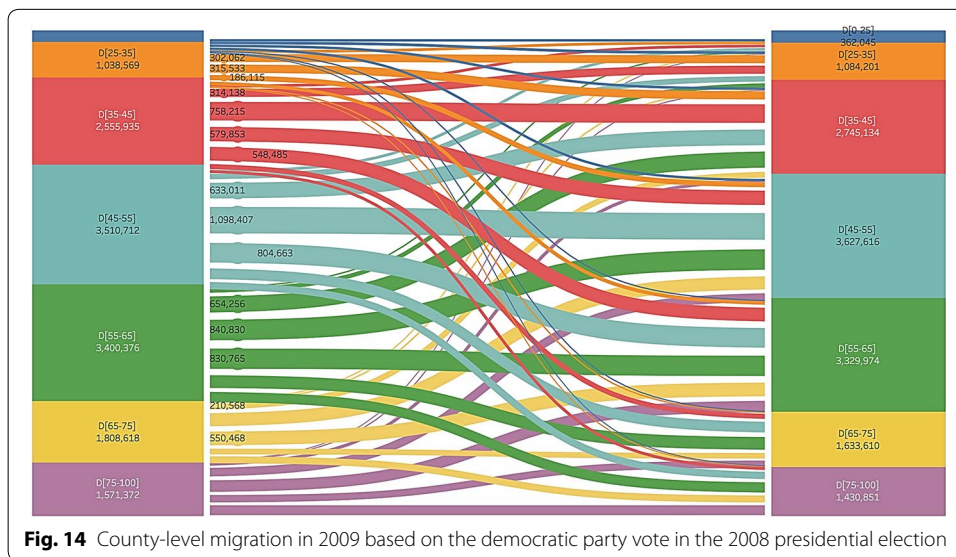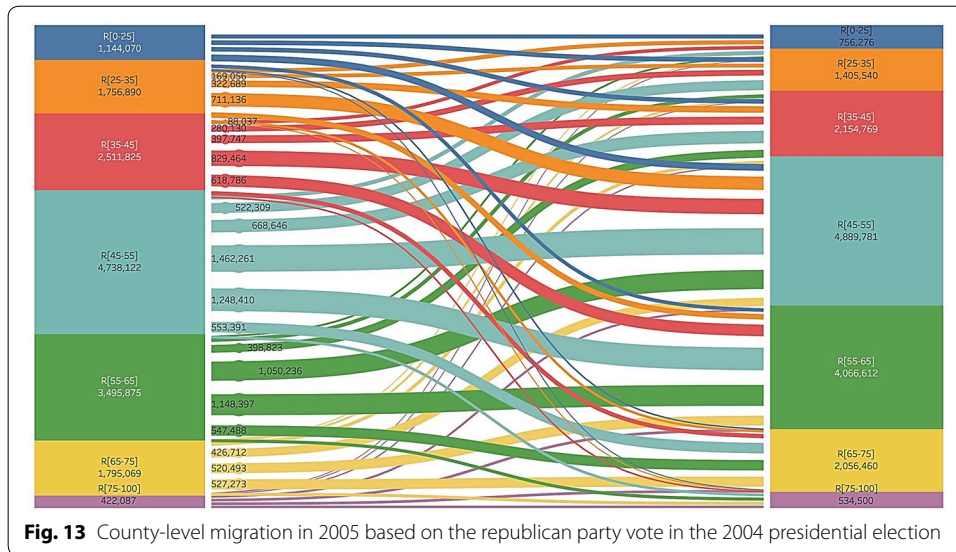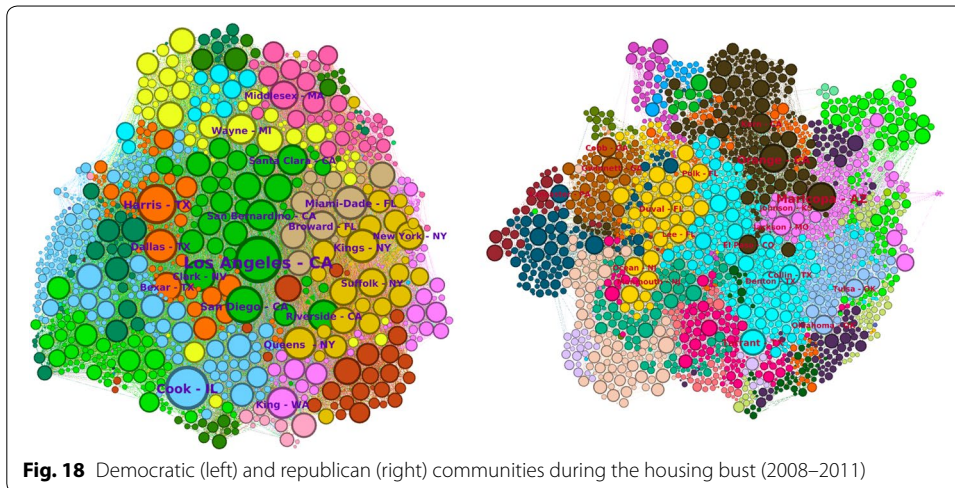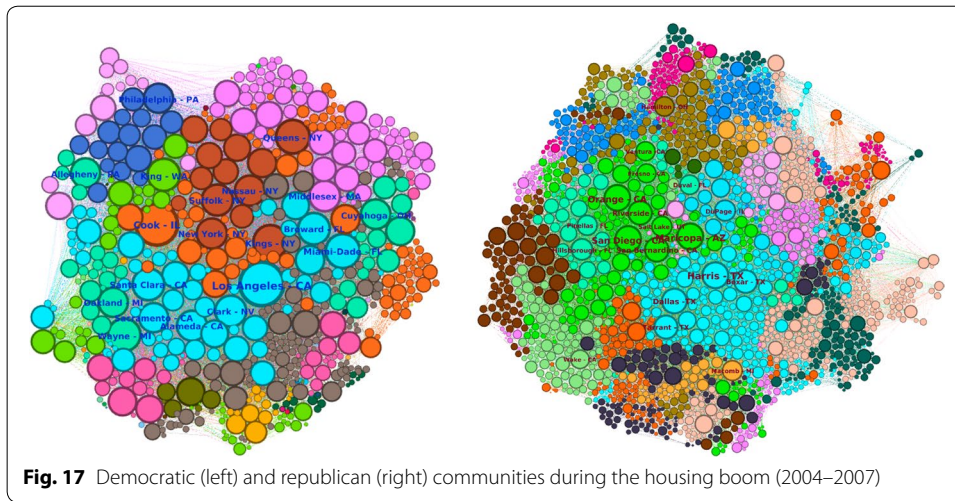
We then analyzed the state-level migration based on political affiliation of States. In Fig. 15, we observe that after the 2004 Presidential election that was won by the Republican candidate, people migrated towards more Republican States. However, after the 2008 Presidential election that was won by the Democratic candidate, people still moved to the Republican States, as shown in Fig. 16.

To further explore the structure of the migration network between Democrat and Republican counties, we analyzed the community structure during the housing boom and the housing bust, where only counties that are majority Democrat or Republican. Figure 17 shows the community structure of the Democrat and Republican communities during the housing boom (2004–2007) without the other. We observe that 612 Democratic counties form 56 communities, but 47 are not connected to the

Fig. 15 State-level migration in 2005 based on the 2004 presidential election of states based on democratic (left) and republican (right) vote of states

Fig. 16 State-level migration in 2009 based on the 2008 presidential election of states based on democratic (left) and republican (right) vote of states
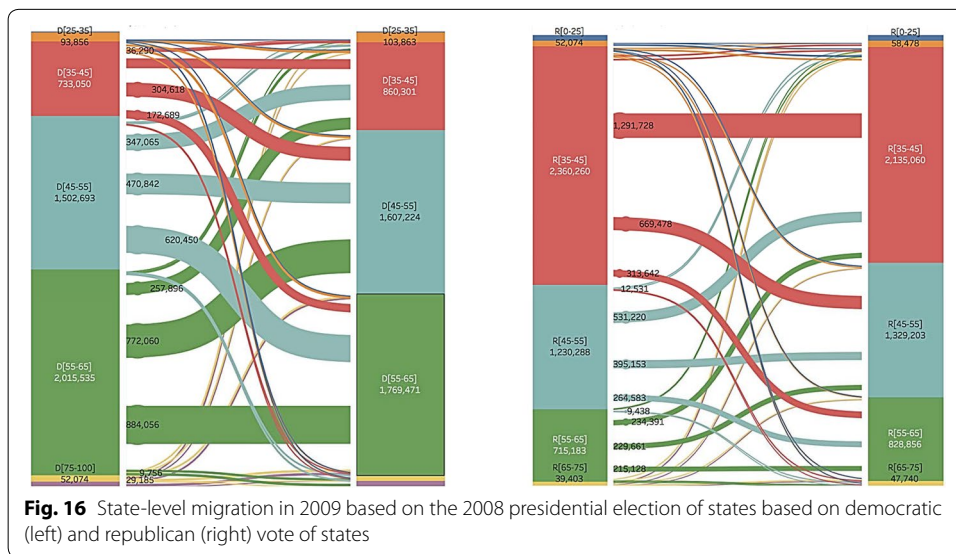
giant component. Similarly, 2521 Republican counties form 106 communities with 86 being isolated. Figure 18 presents the community structure of the Democrat and Republican communities during the housing bust (2008–2011). The giant component gets broken into 35 and 47 communities, respectively, when the Republican and Democratic nodes are removed from the network. We observe that 913 Democratic counties form 41 communities, but 38 are not connected to the giant component. Similarly, 2226 Republican counties form 115 communities with 94 being isolated.

Overall, we observe that there are fewer Democratic counties, but they have larger populations. On the other hand, we see that there are more Republican counties, but with smaller populations. Removing nodes with either political affiliation disrupts the migration network. This suggests that even though there is an afflation between political parties, there are still several nodes that have cross afflation with the other

Fig. 17 Democratic (left) and republican (right) communities during the housing boom (2004–2007)

Fig. 18 Democratic (left) and republican (right) communities during the housing bust (2008–2011)

counties. Close to half of the total migrant population crosses the political afflation network over the period of this study. These results suggests that political afflation, especially of the destination, has some impact on the US migration.

Gravity model

We utilized the gravity model of migration to analyze individual impact of each factor. We estimate the migration by ordinary least squares, as formulated in [29]

$$ln\,M_{ij} = \delta + \alpha\,ln\,P_i + \beta\,ln\,P_j - \gamma\,ln\,D_{ij} + \varepsilon_{ij}, \tag{1}$$

where M_{ij} is number of people migrated from i to j and P_i, P_j, are populations of origin and destination, respectively, and D_{ij} is straight line distance between origin and destination in km. Distance between counties were retried from National Bureau of Economic Research [30]. The parameters δ, α, β, γ are estimated by Ordinary Least Squares (OLS) regression in which ε is random error term.

We normalized the data to get normalized values between 0 and 1. Then, we applied OLS regression on Eq. (1) by projecting migration flow on population,

Table 5 Parameter estimations of the gravity model of migration

	2004–2007 ($\Delta t = 4$)		2008–2011 ($\Delta t = 4$)		2012–2015 ($\Delta t = 4$)		2004–2015 ($\Delta t = 16$)	
	Average	Variance	Average	Variance	Average	Variance	Average	Variance
Population								
δ	−5.82	1×10^{-2}	−5.90	1.2×10^{-2}	−5.51	2.8×10^{-2}	−5.74	4.5×10^{-2}
α (source)	0.55	1.7×10^{-5}	0.53	4.5×10^{-5}	0.46	4.5×10^{-3}	0.51	2.8×10^{-3}
β (destination)	0.56	8×10^{-6}	0.55	4.9×10^{-5}	0.46	6.4×10^{-3}	0.52	4.2×10^{-3}
γ (distance)	0.82	5×10^{-5}	0.81	2×10^{-4}	0.70	1.2×10^{-2}	0.78	6.6×10^{-3}
MSE	2.10	5.2×10^{-4}	2.11	2.2×10^{-4}	1.42	2×10^{-1}	1.88	1.7×10^{-1}
Political affiliation								
δ	−6.38	1.7×10^{-2}	−6.52	1.2×10^{-2}	−5.38	1.2×10^{-2}	−6.09	3.2×10^{-1}
α (source)	0.89	1.1×10^{-3}	0.83	8.3×10^{-4}	0.36	2.5×10^{-4}	0.70	6.3×10^{-2}
β (destination)	0.97	2.6×10^{-4}	0.90	1.2×10^{-3}	0.94	2×10^{-2}	0.93	6.6×10^{-3}
γ (distance)	0.05	2.6×10^{-6}	0.09	1.8×10^{-5}	0.02	7.5×10^{-5}	0.05	8.7×10^{-4}
MSE	2.84	4.3×10^{-4}	2.78	1.3×10^{-4}	2.01	3.5×10^{-1}	2.5	2.5×10^{-1}
Poverty								
δ	−4.88	4.6×10^{-2}	−4.64	6.6×10^{-2}	−4.38	4.8×10^{-2}	−4.63	9×10^{-2}
α (source)	0.80	2.7×10^{-3}	0.86	9.4×10^{-3}	0.78	7.6×10^{-3}	0.81	6.5×10^{-3}
β (destination)	0.85	1.4×10^{-3}	0.90	1.2×10^{-3}	0.82	8×10^{-3}	0.86	4.2×10^{-3}
γ (distance)	0.01	2.4×10^{-4}	0.03	2.1×10^{-4}	0.03	2.8×10^{-4}	0.02	2.8×10^{-4}
MSE	3.14	1.6×10^{-3}	3.15	2.4×10^{-3}	2.20	3.8×10^{-1}	2.83	3.2×10^{-1}
Unemployment								
δ	−3.69	1.2×10^{-1}	−3.86	1.5×10^{-1}	−2.95	2.5×10^{-1}	−3.50	3.1×10^{-1}
α (source)	1.32	1.7×10^{-3}	1.43	1.7×10^{-2}	1.38	2.8×10^{-2}	1.38	1.5×10^{-2}
β (destination)	1.34	2×10^{-3}	1.43	1.5×10^{-2}	1.40	1.8×10^{-2}	1.39	9.9×10^{-3}
γ (distance)	0.05	8.2×10^{-4}	0.14	5.8×10^{-4}	0.10	1.6×10^{-3}	0.10	2.1×10^{-3}
MSE	3.14	3×10^{-3}	3.03	7.8×10^{-4}	2.15	3.5×10^{-1}	2.77	3.1×10^{-1}

Each parameter is independently calculated for α, source, destination, and distance coefficients. In addition, mean squared error is reported for each network sample

poverty, unemployment, and political affiliation independently. Projection of population, poverty, and unemployment is straightforward. However, when projecting politic affiliation, we first identified if the source is Democratic or Republican. If the source is Democratic, we used the Democratic percentage of votes for the source and the destination based on the last presidential election. We applied OLS regression for each year 2004 to 2015. Summary of parameters with mean square error for different Δt time periods is provided in Table 5. We observe similar impact of each parameter for source and destination across different Δt periods. In addition, Table 6 presents the parameters when all factors are considered. Combined results indicate that unemployment has the highest impact, while source political affiliation has the least.

Overall for population analyses, α and β converged to [0.5, 0.6] and γ converged to [0.7, 0.8]. Similar results were obtained for analyzes of migration in UK [31] with γ converging to [1.5, 1.6] and for analyzes of New Zealand [17] with α, β, and γ converging to [0.8, 0,9]. However, results for poverty and unemployment rate in US migration differ from related studies. For example, in analyses of internal migration in Africa [15], α converges to 0.38 and β converged to 6.4 for unemployment, while α converges to 0.32 and β converged to −0.23 for poverty.

Table 6 Combined parameter estimations of the gravity model of migration

	2004–2007 ($\Delta t = 4$)		2008–2011 ($\Delta t = 4$)		2012–2015 ($\Delta t = 4$)		2004–2015 ($\Delta t = 16$)	
	Average	Variance	Average	Variance	Average	Variance	Average	Variance
δ	− 1.94	2.9×10^{-2}	− 2.37	4×10^{-2}	− 2.28	4.7×10^{-2}	− 2.19	6.8×10^{-2}
γ (distance)	0.69	1×10^{-2}	0.80	1.1×10^{-4}	0.78	3×10^{-5}	0.76	5.4×10^{-3}
MSE	1.62	2×10^{-1}	2.41	2×10^{-3}	2.38	2×10^{-3}	2.14	2×10^{-1}
Population								
α (source)	0.47	4.1×10^{-3}	0.54	3.3×10^{-4}	0.57	2.7×10^{-4}	0.53	3.2×10^{-3}
β (destination)	0.45	5.6×10^{-3}	0.56	1.2×10^{-4}	0.57	1.2×10^{-4}	0.53	4.5×10^{-3}
Political affiliation								
α (source)	0.22	2×10^{-4}	0.01	6.3×10^{-4}	0.14	1.2×10^{-3}	0.12	8.4×10^{-3}
β (destination)	0.64	9.5×10^{-3}	0.42	3.9×10^{-3}	0.61	8.7×10^{-4}	0.56	1.5×10^{-2}
Poverty								
α (source)	0.27	1.5×10^{-2}	0.57	8.4×10^{-3}	0.42	1.9×10^{-3}	0.42	2.3×10^{-2}
β (destination)	0.34	1.8×10^{-2}	0.57	1.2×10^{-2}	0.43	4.2×10^{-3}	0.45	1.9×10^{-2}
Unemployment								
α (source)	0.73	1.8×10^{-3}	0.60	1.7×10^{-2}	0.64	4×10^{-3}	0.66	9.4×10^{-3}
β (destination)	0.71	7.8×10^{-3}	0.58	2×10^{-2}	0.62	7×10^{-3}	0.64	1.2×10^{-2}

Conclusions

In this paper, we analyze the US migration as a complex network at county and state levels. Our results confirm what was found in the literature about migration dynamics. We observe that counties are usually active, but show a strong fluctuation in their links. For instance, links with small weights are typically unstable, but there are some links with very large weight that are fluctuating as well. Furthermore, we observed that once edges are active their weight increments are typically small, but might have sudden large increase or decrease. Despite fluctuations in link level, there are backbones in both the county and state-level networks. We observe that the migration network has maintained its characteristics during the political and economic instability, since the millennium and only recently has exhibited slightly differing network characteristics.

In analyses of political and socioeconomic factors, we observe that political inclination of destination and economic factors such as poverty and unemployment rate plays a role in the US migration, whereas crime rate and federal expenditure do not. In addition, population and geographic considerations were indeed strong indicators of US migration flow. Compared to earlier studies on migration, US migration conforms to the gravity model, but has some differences from other countries.

Finally, compared to other temporal networks of cattle movement and air transportation, US migration network has higher activity of nodes over the years and has edge dynamics somewhat similar to the airflow.

Additional files

Additional file 1: Table S1. County-level US Migration Network Characteristics. Number of nodes, edges, avrg degree, avrg weighted degree, diameter, avrg path length, clustering coefficient, modularity, and assortativity for each network of Δt at county level.

Additional file 2: Table S2. State-level US Migration Network Characteristics. Number of nodes, edges, avrg degree, avrg weighted degree, diameter, avrg path length, clustering coefficient, modularity, and assortativity for each network of Δt at state level.

Additional file 3: Figure S1. County-Level Migration Network between 2000 and 2015. Node size is determined by its population. Counties with the highest 20 incoming migration, i.e., weighted in-degree, are labeled and label color shows whether the county was Republican or Democratic in the 2000 presidential elections. Nodes are colored based on the community they belong to and edges are colored based on the node they originate from.

Additional file 4: Table S4. Communities among States. Communities of states for each year.

Additional file 5: Table S3. Interstate and Intrastate Migration. Yearly links and migration within and between states.

Additional file 6: Figure S2. County (left) and State (right) Ranking by Population. Counties and States are ranked by the population in 2000.

Additional file 7: Figure S3. Fraction of Appearance and Disappearance of State-Level Links with no Self Edge (left: Δt = 1, right: Δt = 2). Weight is the number of migrating people.

Additional file 8: Figure S4. Top 10 Ranking of Counties in the Overall Network based on Different Centralities. The network is obtained for Δt=16.

Authors' contributions
BC performed the research, data collection and analyzes. MG designed the research, provided guidance and contributed to the writing of the manuscript. Both authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
Data and materials are available at https://github.com/ComplexNetwork-USMigration/data.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Rayer S, Brown DL. Geographic diversity of inter-county migration in the United States, 1980–1995. Popul Res Study. 2001;20(3):229–52.
2. Beeson PE, DeJong DN, Troesken W. Population growth in U.S. counties 1840–1990. Reginal Science and Urban Economics. 2001.
3. McHugh KE, Gober P. Short term dynamics of U.S. Interstate Migration System. 1992.
4. Frey W. The new urban revival in the United States (vol. 30). Urban Studies. 1993.
5. Molloy R, Smith CL, Wozniak A. Internal migration in the United States. J Econ Perspect. 2011;25(3):173–96.
6. Galle OR, Burr JA, Potter LB. Rethinking measures of migration: on the decomposition of net migration. Soc Indicators Res. 1993;28(2):157–71.
7. Fagiolo G, Mastrorilllo M. The International Migration Network. 2012.
8. Goldade T, Charyyev B, Gunes MH. Network analysis of migration patterns in the United States. Complex networks & their applications. Berlin: Springer; 2017. p. 770–83.
9. Kemper F-J. Internal migration in eastern and western. Reg Stud. 2004;38(6):659–78.
10. Niedomysl T. Residential preferences for interregional migration in Sweden: demographic, so-cioeconomic, and geographical determinants. Environ Plann A. 2008;40(5):1109–31.
11. Davis KF, D'Odorico P. Global spatio-temporal patterns in human migration: a complex network per-spective. PLoS ONE. 2013;8:e53723.
12. Fagiolo G, Mastrorillo M. Does human migration affect international trade? A complex-network perspective. PLoS ONE. 2014;9:e97331.
13. Tranos E, Gheasi M, Nijkamp P. International migration: a global complex network. Environ Plann B. 2015;42(1):4–22.
14. Aleskerov F, Khutorskaya O, Buldyaev A, Yamilov A. Network analysis of international migration. In: International conference on network analysis. Berlin: Springer; 2016. p. 177–85.
15. Rayp G, Ruyssen I. Africa on the move: an extended gravity model of intra-regional migration Migration, a world in motion: a multinational conference on migration and migration policy. Maastricht: Association for Public Policy Analysis and Management Netherlands; 2010.
16. Bergh A, Mirkina I, Nilsson T. Pushed by poverty or by institutions? Determinants of global migration flows. IFN Working Paper. 2015.

17. Alimi O, Mare D, Poot J. Does distance still matter for internal migration and, if so, how? Evidence from 1986 to 2006. Labour, Employment and Work in New Zealand. 2015.
18. Holme P, Saramäki J. Temporal networks. Phys Rep. 2012;519(3):97–125.
19. Braha D, Bar-Yam Y. From centrality to temporary fame: dynamic. Complexity. 2006;12(2):59–63.
20. Gautreau A, Barrat A, Barthélemy M. Microdynamics in stationary complex networks. In: Proceedings of the national academy of sciences. 2009. p. 8847–52.
21. Bajardi P, Barrat A, Natale F, Savini L, Colizza V. Dynamical patterns of cattle trade movements. PLoS ONE. 2011;6(5):e19869.
22. SOI Tax Stats—migration data. (n.d.). From Internal Revenue Service: https://www.irs.gov/statistics/soi-tax-stats-migration-data. Accessed 30 Nov 2017.
23. United States Census. (n.d.). U.S. Census Bureau: https://www.census.gov/support/USACdataDownloads.html.
24. United States Department of Agriculture (n.d.). https://www.ers.usda.gov/data-products/county-level-data-sets/. Accessed 25 Jan 2018.
25. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech. 2008;2008(10):P10008.
26. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: International AAAI conference on weblogs and social media. 2009.
27. Stanley MH, Amaral LA. Scaling behavior in the growth of companies. Nature. 1996;379(6568):804–6.
28. Serrano MA, Boguná M. Extracting the multiscale backbone of complex weighted networks. Proc Natl Acad Sci. 2009;106(16):6483–8.
29. Poot J, Alimi O, Cameron MP, Maré DC. Gravity model of migration: the successful comeback of an ageing superstar in regional science. J Reg Res. 2016; 63–86.
30. Research NB. County distance. http://www.nber.org/data/county-distance-database.html. 2016.
31. Lomax N, Norman P. Subnational migration in the United Kingdom: producing a consistent time series using a combination of available data and estimates. J Popul Res. 2013;30(3):265–88.