Network Analysis of Migration Patterns in the United States

Travis Goldade
Mechanical Engineering
University of Nevada Reno

Batyr Charyyev and Mehmet Hadi Gunes
Computer Science and Engineering
University of Nevada Reno

Abstract: United States (US) internal migration is a seldom studied and poorly understood topic that lies mainly within the realm of economics. This paper attempts to utilize complex network analysis to gain a better understanding of the migration patterns in the US, and how it evolves over the time. In particular, we analyze migration patterns in the US during differing times of the economic prosperity of the housing boom (i.e., 2004-2007) and the recession (i.e., 2008-2011) through complex network analyses. Additionally, we explore whether there is a correlation between the political affiliation of a county and the migration patterns.

I. Introduction

In general, people tend to move to areas seeking opportunities and there is a correlation between net immigration and economic growth [1, 2, 7, 8]. It is very seldom the case that an individual will move to an area with less economic opportunities than where they currently are located.

Historically (i.e., post-World War II to 1980's) the United States (US) economy was driven by the manufacturing of goods, and people tended to immigrate towards areas of high manufacturing [1, 4]. People also tended to move from East to West, North to South, and toward areas of higher population. This was due to the West and South being considered to be a more desirable place to live; and the larger population areas typically had more manufacturing based jobs [1, 2, 4, 6]. In other words, people follow the money, and if given the option they will move to a more desirable place to live.

After the Internet revolution, the economy has been restructured. No longer is the economy based on the production of goods; now it is primarily driven on services and technology. This revolution has caused migration patterns to change [4]. People are less likely to move in the last 20 years than they have been historically. If they do move, the destination typically is areas of high population. However, it is less common than it was in the past, and, in many cases, depopulation of urban centers has been witnessed [4]. The reason being is that the jobs are not as centralized as they used to be. In general, manufacturing jobs are more limited than technology and service based industries. Thus, people are not required to head to the old economic hubs for job opportunities.

Another common pattern of migration is that during times of economic prosperity (e.g., booms and bubbling) migration is ordered. People tend to move towards either certain hubs of job opportunity, or areas that have desirable living locations if they are financially secure [1]. During times of economic hardship (e.g., recessions) people generally leave areas that were once economically prosperous and scatter towards different areas (e.g., bubble bursting) [1, 2, 3]. After bubbles burst and in times of recession the migration patterns are chaotic.

In most migration studies there are certain factors that attract people to certain locations, conversely there are also factors that repel people from their current location [6, 7, 8]. These are called pulling and pushing factors, respectively. Common pulling factors for internal migration are, among others, affordable housing, attractive climate, better employment opportunities, and family ties. Some of the most common internal pushing factors are higher taxes, unemployment, natural disasters, and low chances of marriage.

Even though researchers have analyzed international migration patterns as a complex network [7, 8], none has performed network analysis of internal migration patterns. Analysis of international migration studies found that (i) over time the world has become interconnected and is displaying small world behavior, (ii) there has been the formation of communities within the global community, and (iii) nodes have power law degree distributions.

In this paper, we analyze the network structure of US migration on a yearly basis and focus the housing boom (i.e., 2004-2007) and recession (i.e., 2008-2011) periods. We analyze the networks to identify communities that arise, the trends of these communities (e.g., are they geographically bound, do they have a similar political affiliation, and how do they evolve over the time).

II. Methodology

We collected migration data to and from each county from the Internal Revenue Service (IRS), US Migration data set between the years of 2004 and 2011. This data shows the residence changes of each citizen that migrates to a new county provided they change their county address on their IRS Tax Return.

The data for the political affiliation of each county is obtained from the political voter affiliation of the county during the presidential elections. Note that voter affiliation of the county is not necessary the same as the local county government political affiliation. Admittedly, the local government's actions are much more reflective of the local economic and social policies that are in effect in an area. However, such data is not easily available.

The collected data is processed and analyzed for during the different timeframes. The network is constructed between *counties*, where the directed weighted edges represent the population that has migrated from one county to another. The in-degree and out-degree represent immigration or emigration respectively. The migration flux is found by simply dividing the immigration and emigration of the county by its local population of the same year.

III. Migration Patten Analysis

Figure 1 shows a graph of the overall migration network of the US from 2004 to 2011. There are 3,144 nodes each representing a US county and each edge representing the total weighted directed migration from one county to another over the time period of 2004 to 2011 as reported by the IRS database. Nodes are colored based on the communities they belong to and edges are colored with the color of the node they originate from. The nodes are sized by their degree and if they are above a certain threshold, the text color indicates whether the county is Republican or Democratic in presidential elections.

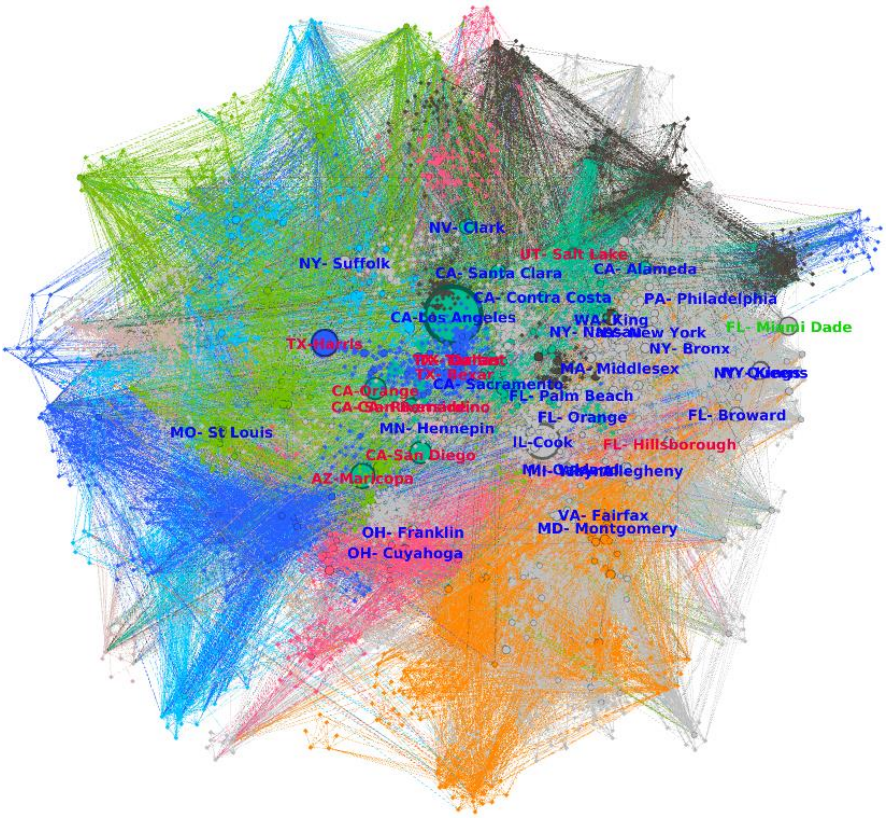

Figure 1: US Migration Network between 2004-2011

Note that more than 98% of the nodes are part of the giant component and other nodes are dropped from the subsequent analysis. The graph showed dissasortative correlation between node degrees indicating most of the migration was between dissimilar size counties.

Table 1 summarizes the network characteristics of different periods. Results show that the US migration network has a small world characteristics as there is a

high clustering and low average path length in all samples. Additionally, there are few hubs, and, as shown in the following section, there is an inverse correlation between clustering coefficient and weighted in-degree.

Table 1: US Migration Network Characteristics

Time frame	Giant component	Edges	Average degree	Average weighted degree	Diameter	Average path length	Clustering Coefficient	Modularity	Communities		
									Republican	Swing	Democratic
2005	3,179	85,016	26.74	1,805	8	3.13	0.540	0.628	10	3	4
2006	3,173	83,343	26.26	1,727	8	3.15	0.545	0.625	9	2	3
2007	3,198	85,551	26.75	1,769	8	3.14	0.542	0.628	8	3	4
2008	3,184	83,085	26.09	1,717	8	3.14	0.547	0.619	9	6	1
2009	3,168	78,294	24.71	1,658	8	3.19	0.538	0.635	7	10	0
2010	3,176	81,117	25.54	1,679	8	3.16	0.546	0.635	6	8	3
2011	3,203	89,527	27.95	1,940	9	3.11	0.55	0.647	5	7	5
2004-2007	3,415	122,292	35.81	6,533	7	2.96	0.577	0.627	10	2	4
2008-2011	3,389	120,623	35.59	6,569	7	2.95	0.566	0.646	8	3	5
2004-2011	3,597	147,881	41.11	12,392	7	2.87	0.573	0.642	9	3	4

Figure 2 displays the ranking of weighted in-degree and out-degree of nodes. Interestingly, the Out-degree ranking overlaps with the In-degree ranks except at the small ranks. We observe that while some counties have large influx of migration over the 2004-2011 time period, 363 have no incoming migration over the 7 years. On the other hand, largest out-degree is higher than in-degree and only 160 counties have no outgoing migration over the 7 years.

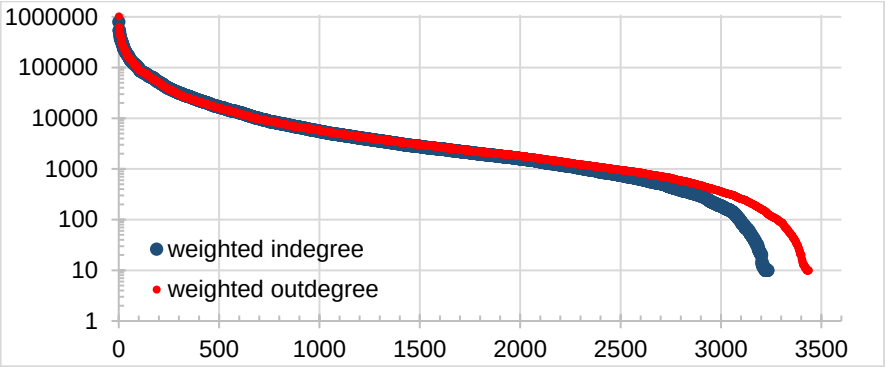

Figure 2: County Ranking by Weighted In- and Out-degree

Central Hubs

We observe that both total node population and hubs have a strong influence on the weighted in-degree of a node. Figure 3 show the top 10 ranking of counties based on populations, weighted in-degree, degree, Eigenvector, Betweenness, Pagerank, and hub centralities over the 2004 to 2011 time period. It shows that the strongest hubs are AZ-Maricopa County and NV-Clark County. These two counties had some of the lowest housing prices during the housing boom. As such they are central to the US migration network although their populations are relatively lower than other counties. Other hubs include CA-Los Angeles County, TX-Harrison County, and CA-San Diego County. All of these counties are found on the top ten rankings of all centralities. This indicates a strong connection between population and hubs in the US migration network.

Figure 3: Top 10 Ranking of Counties based on Centralities

Figure 4 shows correlation between the clustering coefficient and degree. Lack of high clustering for high degree nodes indicate there is not mutual migration between all counties. This also means counties migrating to high degree hubs are not likely to migrate to each other. On the other hand, low degree counties show a mixed pattern with some having very high clustering of 1 indicating migration between all of its neighbors or very low clustering of 0 indicating no interaction with its neighbors. Overall, however, we observe a high clustering for majority of counties indicating high migration among counties. 231 counties (i.e., 6.42%) have a clustering of 1 and 3275 (91.05%) have a clustering of 1/3 or higher.

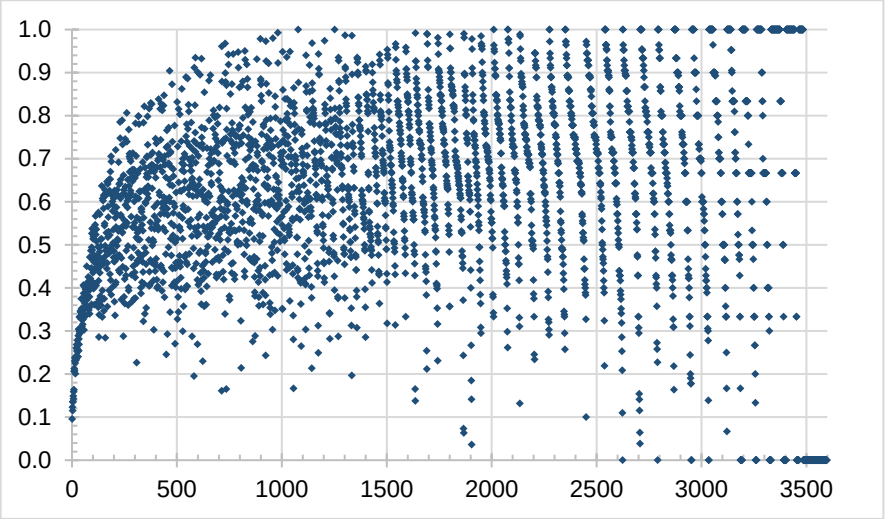

Figure 4: Clustering Coefficient of Counties Ranked by Degree

Communities

As is shown in Table 1, there are very few changes in the US migration network over the studied time periods. This indicates that the migration patterns are robust to the varying economic and political conditions.

Perhaps the most interesting part of this table is the formation of communities. Throughout the time frame studied, the total number of communities range from 15 to 17. All of the communities are geographically bound, with the largest community forming over most of the West, including California, Arizona, Texas, Nevada, Oregon, New Mexico, Washington and Utah. This community includes several hubs, such as San Diego, Los Angels, Maricopa, and Clark County. The other communities are much smaller, with the majority of them forming in the East. This suggests that there is a strong pull from the other communities into this large community in the West. This community encompasses most of what is called the Sun Belt and includes most of the South West. It is attractive for migrants because of low housing costs and desirable climate.

An interesting observation was that between 2009-2010 Michigan formed its own community, the only state in the network to do so. Michigan was one of the hardest hit states in the recession.

Another observation was the vast majority of communities formed were either Republican leaning or neutral prior to the recession. In 2008, when the housing market crashed the number of Democratic leaning communities doubled, from 3 to 6. Then in 2009 the number of Democratic leaning communities increased again to a total of 10. This is to be expected, however, as a Democratic president was elected in 2008, and several counties switched from Neutral or Republican leaning to Democratic.

Political Afflation

We also looked into whether there is a correlation with county political affiliation and migration. There were two methods in which this was analyzed. The first statistical method was simply looking for a regression between political afflation percentile and weighted in-degree. This method suggests a random distortion between these two variables over the studied time period.

Table 2 shows the statistical significance of political affiliation and inflow migration of each county over the entire time frame. It suggests that people are slightly attracted towards moderate Republican nodes. 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. These findings were further supported by the community analysis.

Table 2: 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%

To further explore the structure of US migration network, the community structure during the housing boom and the housing bust were analyzed. The overall network of each time period can be seen in Figures 5 and 6, respectively. The nodes are colored based on the communities they are affiliated and edges are based on the origin node. Nodes are sized based on degree and colored based on political affiliation. Both networks show similar structure and majority of the migration corridors remain intact for each community.

Figures 7, 8, 9 and 10 show the community structure of the Republican and Democrat communities for the housing boom (2004-2007) and housing bust (2008-2011) time periods. We observe that there are more Republican counties but they tend to have smaller population. On the other hand, the fewer Democratic counties often have large populations.

Removing nodes with either political affiliation greatly disrupts the network, with the Giant Component being broken into 35 and 47 communities when the Republican and Democratic nodes are removed respectively. This suggests that even though there is an afflation between political parties there are still several nodes that have cross afflation with the other counties.

Close to 50% of the total migrant population crosses the political afflation network over the time frame of this study. All of these suggests that political afflation does not have a major impact on the US migration network.

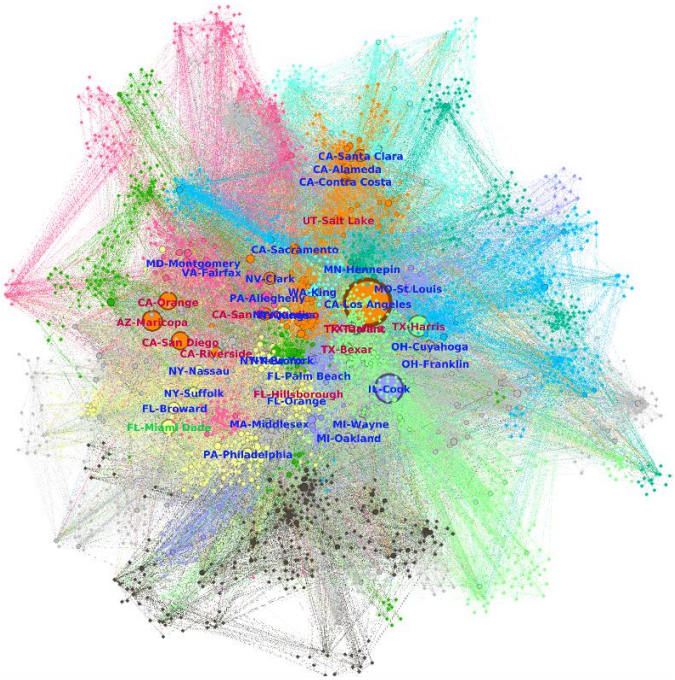

Figure 5: Housing Boom (2004-2007)

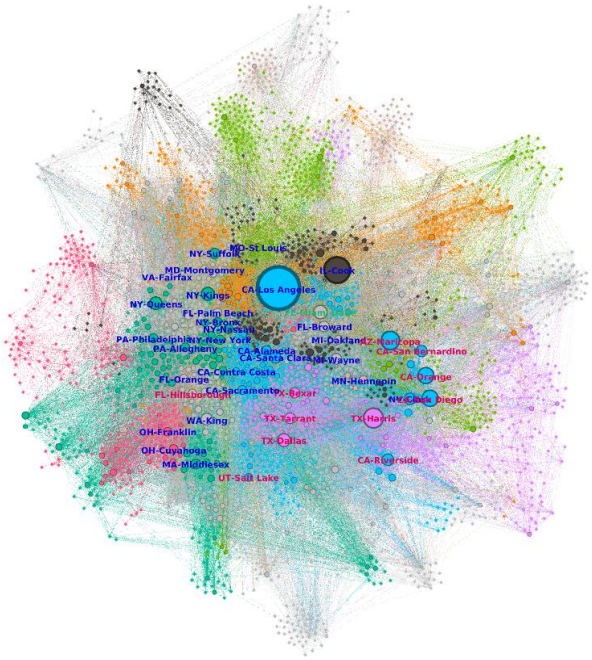

Figure 6: Housing Bust (2008-2011)

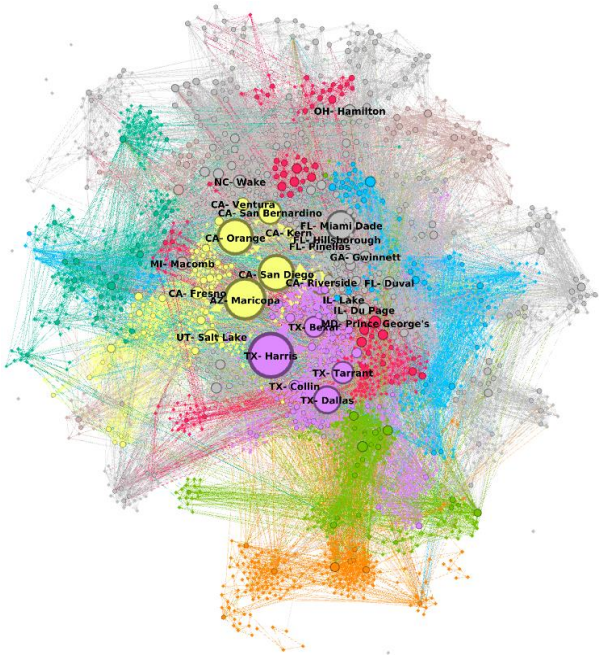

Figure 7: Republican Communities during Housing Boom (2004-2007)

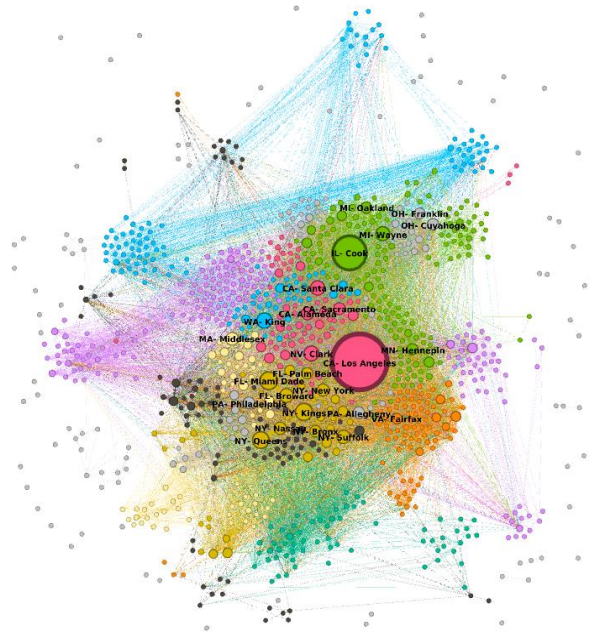

Figure 8: Democratic Communities during Housing Boom (2004-2007)

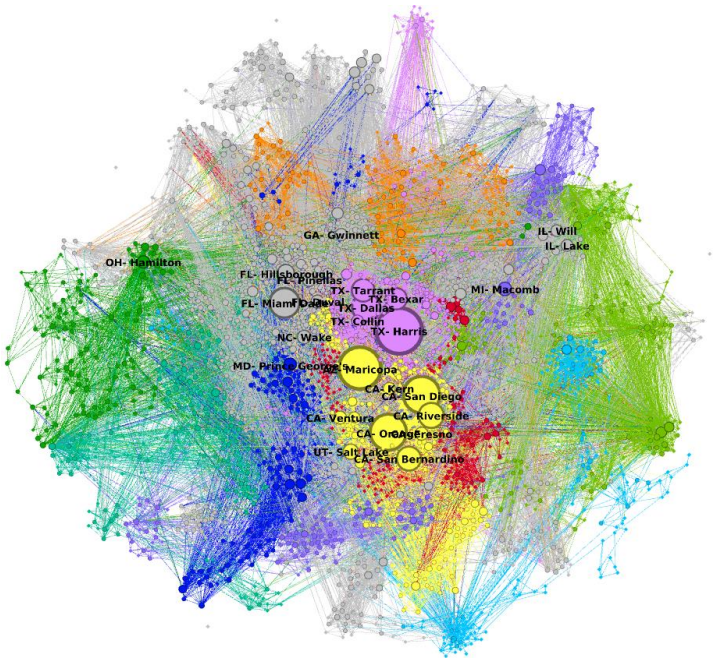

Figure 9: Republican Communities during Housing Bust (2008-2011)

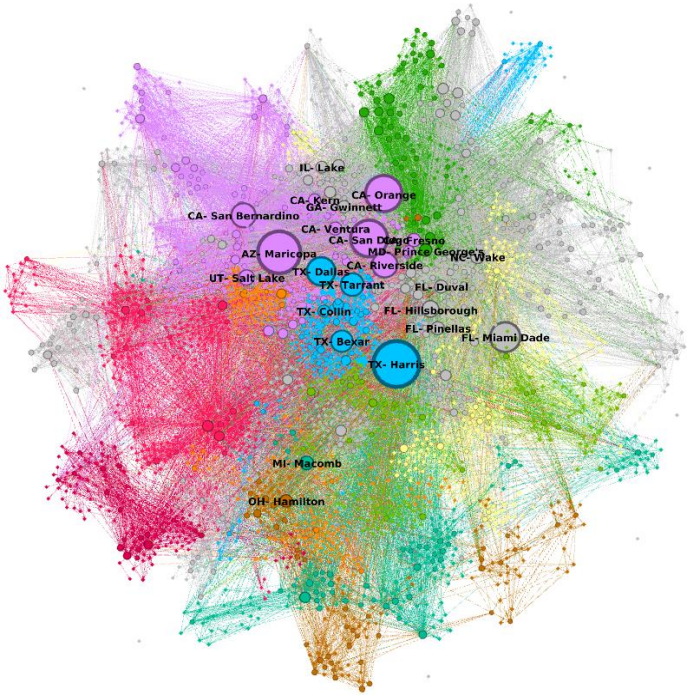

Figure 10: Democratic Communities during Housing Bust (2008-2011)

IV. Conclusion

The results of our study of migration patterns as complex network confirm what was found in the literature about migration dynamics. We found that population and geographic considerations were indeed strong indicators of US migration. Other indicators such as housing costs or family ties were beyond the scope of this study.

It is generally though that local government should have an influence on the local tax structure and therefore should have an impact on the US migration network. However, contrary to expectations, the network has more or less maintained its characteristics during the last decade of political and economic instability.

As a future work, we plan to extend this study to cover earlier time periods. Additionally, we would like to explore use of local election results to better capture political dynamics rather than the presidential election.

Acknowledgements This material is based upon work in part supported by the National Science Foundation under grant number EPS- IIA-1301726.

V. References

[1] P.E. Beesona, D.N. DeJonga, W. Troeskenb, "Population growth in U.S. counties, 1840–1990" in Regional Science and Urban Economics, Nov 2001, pp 669-699.

[2] S. Rayer and D.L. Brown, "Geographic Diversity of Inter-County Migration in the United States, 1980-1995" in Population Research and Study, Jun 2001, pp 229-252.

[3] O.R. Galle, J.A. Burr, and L.B. Potter, "Rethinking Measures of Migration: On the decomposition of net migration" in Social Indicators Research, Feb 1993, pp 157-171.

[4] W.H. Frey "The New Urban Revival in the United States" in Urban Studies, 1993, pp 741-774.

[5] R. Molloy, C.L. Smith, A. K. Wozniak, "Internal Migration in the United States" in Journal of Economic Perspectives, American Economic Association, 2011, pp 173-96.

[6] K.E Mchugh and P. Gober, "Short-Term Dynamics of the U.S. Interstate Migration System, 1980–1988" Oct 1992, pp 428-445.

[7] G. Fagiolo and M. Mastrorillo, "The International-Migration Network" in arXiv:1212.3852 [physics.soc-ph], Dec 2012.

[8] E. Tranos, M. Gheasi, and P. Nijkamp, "International Migration: A Global Complex Network" in Environment and Planning B, Feb 2015.