Analysis of Metropolitan Human Migration Flow Patterns in the United States

Georgia Bass
Political and Social Networks PS753

INTRODUCTION:

Factors that drive human migration include environmental conditions, economic stability, and political unrest. Whether external or internal forces cause humans to move, visually mapping human migration across the United States can inform policy makers, economists, and contribute to a more in-depth understanding of why people move. In this paper, we analyzed in and out flow migration counts between the largest Metropolitan Statistical Areas of the United States between 2012 and 2016. Visualizing these internal migration flows through interactive tools makes it possible to see large data patterns through virtual telescopes. As Hidalgo and Almossawi cite, "these new visualizations [..] allow the emergence of a richer relationship between the visualization and the reader, who is now less of a spectator and more an explorer."

NETWORK DATA AND METHODS:

The dataset is from the United States Census Bureau and is measuring Metro Area-to-Metro Area Migration Flow. The data was gathered from 2012 to 2016 by the American Community Survey (ACS), documenting human migration flow patterns between Metropolitan Statistical Areas. Containing five-digit Federal Information Processing Standards (FIPS) codes of the MSAs and the United States associated territories, the spreadsheet divides migration counts into flow from geography B to geography A, from geography A to geography B, net migration from geography B to geography A[2], and gross migration between geography A and geography B[2]. The ACS's margin of error resides at a 90 percent confidence interval while the February 2013 MSA definitions are used. A unique feature of the Metro Area-to-Metro Area Migration Flow data is that the ACS tracked individuals moving from South America, Oceania and At Sea, Northern America, United States Island Areas, Europe, Caribbean, Central America, Asia, and Africa to their respective MSAs across the country. The original data is in a network edgelist matrix.

This network is a random sample of Americans that answered the American Community Survey for the 2012-2016 census. It asked respondents "age 1 year and over whether they lived in the same residence 1 year ago. For people who lived in a different residence, the location of their previous residence was collected."

The Metropolitan Statistical Areas (MSA) represents a node or vertex. The level of analysis is on an individual-scale. There are 399 nodes or vertices in this dataset.

The number of individuals moving from MSA(a) to MSA(b) represents a tie or edge. The ties are weighted and directed links – since there are unique values of people flowing from MSA(a) to MSA(b) and vice versa. The range of tie values are screenshot below:

```
> summary(E(network_igraph)$weight)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
    0.0     1.0    21.0   200.1    81.0 82951.0       2 
```

The tie measures the count of people that have migrated from one MSA to another MSA and is an adequate measure.

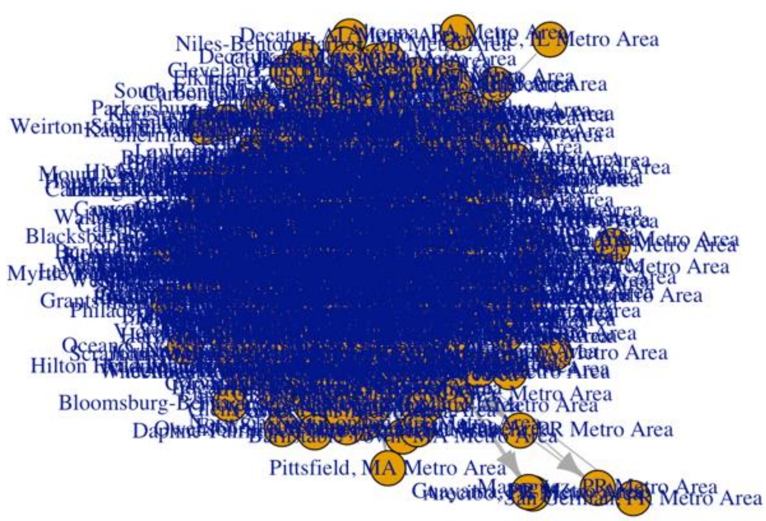
There is only 1 component within the complete network size (399 vertices) that represents weighted ties between MSA. Within the dataset 0 is entered to represent 0 people traveling between MSA(a) to MSA(b) or vice versa. If the rows with 0 were excluded from the data we would expect there to be isolates in the data representing MSA with no migration flow.

All nodes are fully connected and there are no missing edges within 71613 total edges.

In the context of migration flow patterns between MSA the network diameter and distribution of network geodesics would not be applicable. The path length refers to calculating the distances between two or more nodes when network ties are binary or unweighted. In this network the nodes are weighted and directed. The data refers to people traveling only from MSA(a) to MSA(b) or vice versa and there is no MSA(c).

The density of a network is defined as the proportion of ties present out of all possible ties in a network. There is no looping within the network and only one edge exists between nodes. The graph density of the network is: 0.451.

Although we set the threshold at the 95th percentile for *pct95.ig* and *number689.ig* the visualization of the network is still too large to be anything more than a hairball.



A threshold transformation was used to create the final network data used in the analysis. There are 71613 total edges between the 391 nodes. The migration network is dense and many MSA ties are zero because there is no migration occurring. For the purpose of this analysis we limit the definition of a tie between MSA. We create a distinct network in which the tie is defined by the frequency of the edges in relation to MSA by percentile:

a) *number689.ig and city689.stat*

The 95th percentile for the weighted network ties yields 689. With an edge count of 3579 and node count of 399 – this network is too large to visualize. We calculate a 99.95th percentile, which yields the following:

```
> quantile(E(network_igraph)$weight, .9995, na.rm = TRUE)
  99.95%
21910.31
```

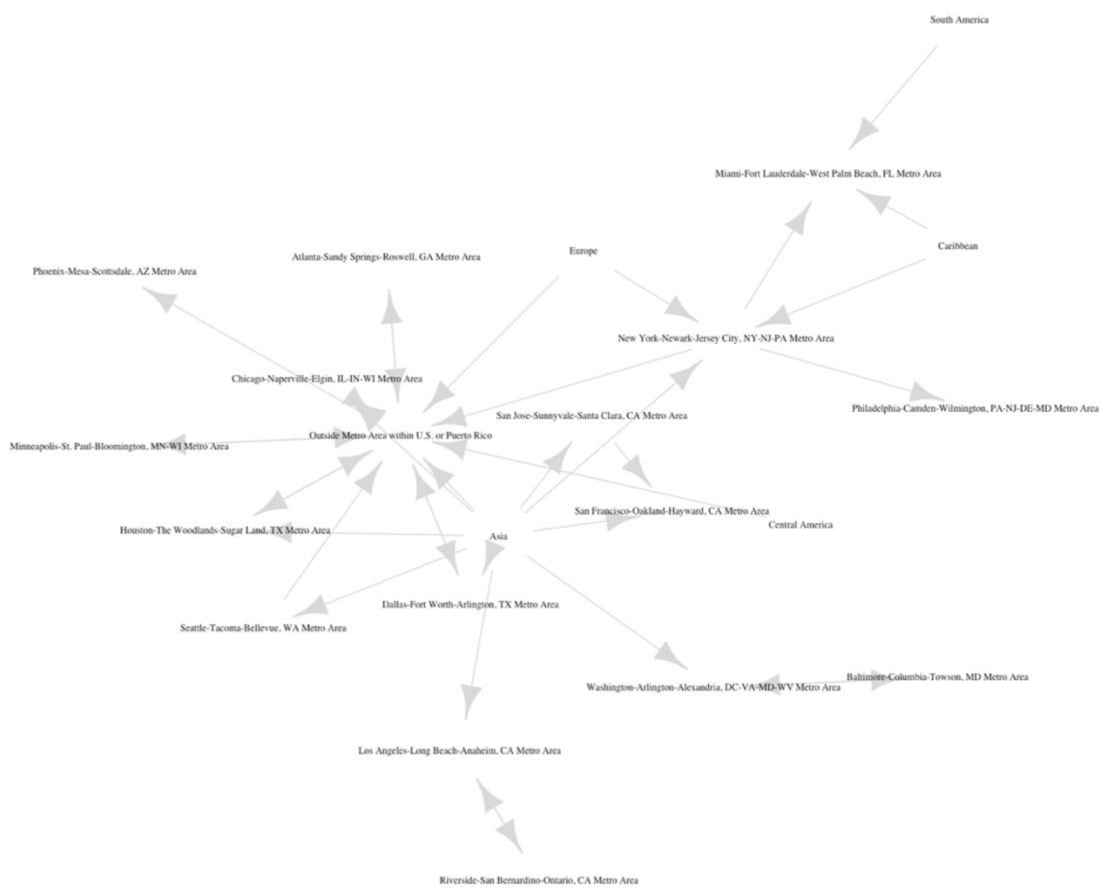
```
> ecount(number689.ig)
[1] 36
> vcount(number689.ig)
[1] 22
```

At the 99.95 percentile a tie exists if 21910.31 individuals or more are migrating from MSA to another MSA. The new graph density is 0.0779.

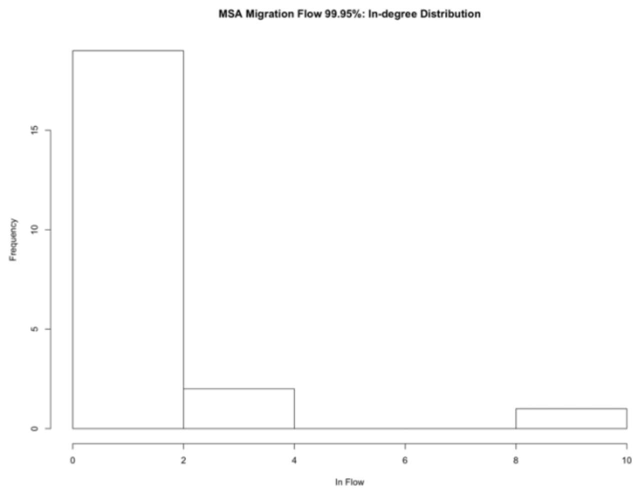
The MSAs listed are as followed for *number689.ig*:

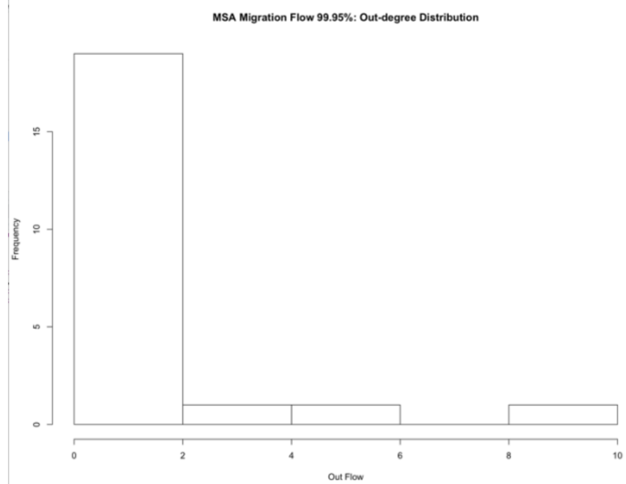
```
 [1] "Atlanta-Sandy Springs-Roswell, GA Metro Area"            "Baltimore-Columbia-Towson, MD Metro Area"
 [3] "Chicago-Naperville-Elgin, IL-IN-WI Metro Area"           "Dallas-Fort Worth-Arlington, TX Metro Area"
 [5] "Houston-The Woodlands-Sugar Land, TX Metro Area"         "Los Angeles-Long Beach-Anaheim, CA Metro Area"
 [7] "Miami-Fort Lauderdale-West Palm Beach, FL Metro Area"    "Minneapolis-St. Paul-Bloomington, MN-WI Metro Area"
 [9] "New York-Newark-Jersey City, NY-NJ-PA Metro Area"        "Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area"
[11] "Phoenix-Mesa-Scottsdale, AZ Metro Area"                  "Riverside-San Bernardino-Ontario, CA Metro Area"
[13] "San Francisco-Oakland-Hayward, CA Metro Area"            "Seattle-Tacoma-Bellevue, WA Metro Area"
[15] "Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area" "Outside Metro Area within U.S. or Puerto Rico"
[17] "Asia"                                                    "Central America"
[19] "Europe"                                                  "South America"
[21] "Caribbean"                                               "San Jose-Sunnyvale-Santa Clara, CA Metro Area"
```

The following is the visualization of the *number689.ig* network:



Average Node Degree Distribution:





HUMAN MIGRATION NETWORK:

The following is analysis using the *number689.ig and city689.stat* network.

There are 22 MSAs in which more than 21,910 people have migrated from MSA(a) to MSA(b). The MSAs have seen a large influx of migration from Asia and a large influx to Outside Metro Area within U.S or Puerto Rico.

Eigenvector centrality calculation shows "a node is important if it is linked to by other important nodes, [thus] a node with a high eigenvector centrality is not necessarily highly linked (the node might have few but important links)." Asia has the highest eigenvector centrality score of 0.5, while Europe is second with a score of 0.31. The network shows that eight of the 22 MSA's have the same eigenvector centrality score of 0.27, yet indegree/outdegree scores differ. Upon further investigation, these MSAs have higher indegree scores than outdegree scores – possibly meaning that these MSAs have seen an increase in populations between 2012 and 2016.

Bonacich power centrality of a node relates to the centrality scores of the neighboring nodes (Brynmar). A positive Bonacich measure means that neighboring nodes may have lower centrality scores and a negative measure means that neighboring nodes may have higher centrality score. The Bonacich scores of zero show that MSAs such as Miami-Fort Lauderdale FL or San Francisco CA are solely receiving people into the MSA (see graph). Whereas New York-Newark-Jersey City NY-NJ-PA may have the same indegree/outdegree score, the Bonacich score is dependent on the centrality score of those traveling to and from Europe, Miami FL, Caribbean, or Philadelphia PA.

Total Degree, Indegree, Outdegree, Eigenvector Centrality:

```
> city689.nodes
                                                 name totdegree indegree outdegree      eigen
1                                                Asia        10        0        10 0.50000000
2             Atlanta-Sandy Springs-Roswell, GA Metro Area         2        1         1 0.27272727
3               Baltimore-Columbia-Towson, MD Metro Area         2        1         1 0.00000000
4                                           Caribbean         2        0         2 0.04545455
5                                     Central America         1        0         1 0.27272727
6            Chicago-Naperville-Elgin, IL-IN-WI Metro Area         3        2         1 0.27272727
7              Dallas-Fort Worth-Arlington, TX Metro Area         3        2         1 0.27272727
8                                              Europe         2        0         2 0.31818182
9          Houston-The Woodlands-Sugar Land, TX Metro Area         3        2         1 0.27272727
10            Los Angeles-Long Beach-Anaheim, CA Metro Area         3        2         1 0.00000000
11     Miami-Fort Lauderdale-West Palm Beach, FL Metro Area         3        3         0 0.00000000
12       Minneapolis-St. Paul-Bloomington, MN-WI Metro Area         2        1         1 0.27272727
13         New York-Newark-Jersey City, NY-NJ-PA Metro Area         6        3         3 0.27272727
14            Outside Metro Area within U.S. or Puerto Rico        16       10         6 0.22727273
15  Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area         1        1         0 0.00000000
16                 Phoenix-Mesa-Scottsdale, AZ Metro Area         1        1         0 0.00000000
17         Riverside-San Bernardino-Ontario, CA Metro Area         2        1         1 0.00000000
18            San Francisco-Oakland-Hayward, CA Metro Area         2        2         0 0.00000000
19            San Jose-Sunnyvale-Santa Clara, CA Metro Area         2        1         1 0.00000000
20                 Seattle-Tacoma-Bellevue, WA Metro Area         2        1         1 0.27272727
21                                      South America         1        0         1 0.00000000
22 Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area         3        2         1 0.00000000
```

Bonacich Power:

```
> sna::bonpow(get.inducedSubgraph(city689.stat,v=c(1:16,18:21)))
                                                  Asia           Atlanta-Sandy Springs-Roswell, GA Metro Area
                                             0.3694845                                             -1.2931956
                  Baltimore-Columbia-Towson, MD Metro Area                                          Caribbean
                                             0.0000000                                              1.6626801
                                       Central America         Chicago-Naperville-Elgin, IL-IN-WI Metro Area
                                            -1.2931956                                             -1.2931956
                Dallas-Fort Worth-Arlington, TX Metro Area                                             Europe
                                            -1.2931956                                             -0.3694845
           Houston-The Woodlands-Sugar Land, TX Metro Area          Los Angeles-Long Beach-Anaheim, CA Metro Area
                                            -1.2931956                                              0.0000000
      Miami-Fort Lauderdale-West Palm Beach, FL Metro Area     Minneapolis-St. Paul-Bloomington, MN-WI Metro Area
                                             0.0000000                                             -1.2931956
          New York-Newark-Jersey City, NY-NJ-PA Metro Area          Outside Metro Area within U.S. or Puerto Rico
                                             0.1847422                                             -2.0321646
    Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area                 Phoenix-Mesa-Scottsdale, AZ Metro Area
                                             0.0000000                                              0.0000000
              San Francisco-Oakland-Hayward, CA Metro Area          San Jose-Sunnyvale-Santa Clara, CA Metro Area
                                             0.0000000                                              0.7389689
                   Seattle-Tacoma-Bellevue, WA Metro Area                                      South America
                                            -1.2931956                                              0.7389689
```

Derived and Reflected Centrality:

```
                                                               rc        dc
Atlanta-Sandy Springs-Roswell, GA Metro Area           0.17280285 0.8271971
Baltimore-Columbia-Towson, MD Metro Area               1.00000000 0.0000000
Chicago-Naperville-Elgin, IL-IN-WI Metro Area          0.09277253 0.9072275
Dallas-Fort Worth-Arlington, TX Metro Area             0.12925016 0.8707498
Houston-The Woodlands-Sugar Land, TX Metro Area        0.10930322 0.8906968
Los Angeles-Long Beach-Anaheim, CA Metro Area          0.24608252 0.7539175
Miami-Fort Lauderdale-West Palm Beach, FL Metro Area   0.46652451 0.5334755
Minneapolis-St. Paul-Bloomington, MN-WI Metro Area     0.15629822 0.8437018
New York-Newark-Jersey City, NY-NJ-PA Metro Area       0.23012154 0.7698785
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area 0.32878111 0.6712189
Phoenix-Mesa-Scottsdale, AZ Metro Area                 0.14948632 0.8505137
Riverside-San Bernardino-Ontario, CA Metro Area        1.00000000 0.0000000
San Francisco-Oakland-Hayward, CA Metro Area           0.14031796 0.8596820
Seattle-Tacoma-Bellevue, WA Metro Area                 0.06358590 0.9364141
Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area 0.13171731 0.8682827
Outside Metro Area within U.S. or Puerto Rico          0.41790202 0.5820980
Asia                                                   0.00000000 1.0000000
Central America                                        0.00000000 1.0000000
Europe                                                 0.00000000 1.0000000
South America                                          0.00000000 1.0000000
Caribbean                                              0.00000000 1.0000000
San Jose-Sunnyvale-Santa Clara, CA Metro Area          0.06832958 0.9316704
```

Constrain is a measure of the redundancy of a node's connections between 0 and 1. Where 0 is a complete lack of redundancy and 1 is complete redundancy. Many MSAs have complete redundancy such as Atlanta

GA, Baltimore MD, or Minneapolis MN. Because the network is directed, the nodes with complete redundancy score of zero have only one directed tie either to or from that MSA. Betweenness centrality measures “the extent to which a vertex lies on paths between other vertices. Vertices with high betweenness may have considerable influence within a network by virtue of their control over information passing between others.” The Outside Metro Area within U.S or Puerto Rico has the highest betweenness centrality score of 58, while New York-Newark-Jersey City, NY-NJ-PA has the next highest betweenness centrality score of 12 in the network. In the graphical representation of the network these two MSAs are central nodes.

```
> constraint(number689.ig)
          Atlanta-Sandy Springs-Roswell, GA Metro Area          Baltimore-Columbia-Towson, MD Metro Area
                                             1.0000000                                         1.0000000
         Chicago-Naperville-Elgin, IL-IN-WI Metro Area        Dallas-Fort Worth-Arlington, TX Metro Area
                                             0.6880244                                         0.7381488
       Houston-The Woodlands-Sugar Land, TX Metro Area      Los Angeles-Long Beach-Anaheim, CA Metro Area
                                             0.6990146                                         0.5792945
    Miami-Fort Lauderdale-West Palm Beach, FL Metro Area   Minneapolis-St. Paul-Bloomington, MN-WI Metro Area
                                             0.5115349                                         1.0000000
      New York-Newark-Jersey City, NY-NJ-PA Metro Area Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area
                                             0.2884826                                         1.0000000
             Phoenix-Mesa-Scottsdale, AZ Metro Area     Riverside-San Bernardino-Ontario, CA Metro Area
                                             1.0000000                                         1.0000000
          San Francisco-Oakland-Hayward, CA Metro Area           Seattle-Tacoma-Bellevue, WA Metro Area
                                             0.8276955                                         0.6241747
Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area      Outside Metro Area within U.S. or Puerto Rico
                                             0.5256975                                         0.1709651
                                                  Asia                                   Central America
                                             0.2557751                                         1.0000000
                                                Europe                                     South America
                                             0.6377676                                         1.0000000
                                             Caribbean      San Jose-Sunnyvale-Santa Clara, CA Metro Area
                                             0.7432629                                         0.8698590
```

Constraint

```
> igraph::betweenness(number689.ig, directed=TRUE)
          Atlanta-Sandy Springs-Roswell, GA Metro Area          Baltimore-Columbia-Towson, MD Metro Area
                                                     0                                                 0
         Chicago-Naperville-Elgin, IL-IN-WI Metro Area        Dallas-Fort Worth-Arlington, TX Metro Area
                                                     0                                                 0
       Houston-The Woodlands-Sugar Land, TX Metro Area      Los Angeles-Long Beach-Anaheim, CA Metro Area
                                                     0                                                 1
    Miami-Fort Lauderdale-West Palm Beach, FL Metro Area   Minneapolis-St. Paul-Bloomington, MN-WI Metro Area
                                                     0                                                 0
      New York-Newark-Jersey City, NY-NJ-PA Metro Area Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area
                                                    12                                                 0
             Phoenix-Mesa-Scottsdale, AZ Metro Area     Riverside-San Bernardino-Ontario, CA Metro Area
                                                     0                                                 0
          San Francisco-Oakland-Hayward, CA Metro Area           Seattle-Tacoma-Bellevue, WA Metro Area
                                                     0                                                 4
Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area      Outside Metro Area within U.S. or Puerto Rico
                                                     1                                                58
                                                  Asia                                   Central America
                                                     0                                                 0
                                                Europe                                     South America
                                                     0                                                 0
                                             Caribbean      San Jose-Sunnyvale-Santa Clara, CA Metro Area
                                                     0                                                 0
```

Betweenness Centrality

The two measures of structural equivalence used to visualize the network are network roles and cluster dendrogram. The network role measure is based on the cluster dendrogram. The color of the node correlates to the structural equivalence values or how closely related the nodes are to each other. The size of the node is based on the weight of the ties. The dendrogram shows hierarchical clustering where the height dimension on the y-axis indicates how different the nodes are from each other, with higher numbers indicating a greater degree of difference. The node indices are the numbers appearing on the x-axis. The difference between nodes is indicated by a link at the specified height.

Network Roles:

Cluster Dendrogram:

The range of observed values of the measurement in the dataset are:

```
> summary(E(number689.ig)$weight)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  22118   24262   28622   32692   34495   82951 
```

In the original dataset where thousands of people are migrating between 399 MSAs, it is surprising that the ties from Asia, Europe, Caribbean, South America are all greater than 21,910 (as defined by the 99.95 percentile). The three central nodes: Asia, Outside Metro Area within U.S or Puerto Rico, and New York-Newark-Jersey City, NY-NJ-PA – correlates with the expected number of people migrating to locations outside of MSAs, the New York City as largest metro area in the United States, and Asia as the largest continent.

Lower than expected scores (lower degree of difference) that appeared on the cluster dendrogram include:

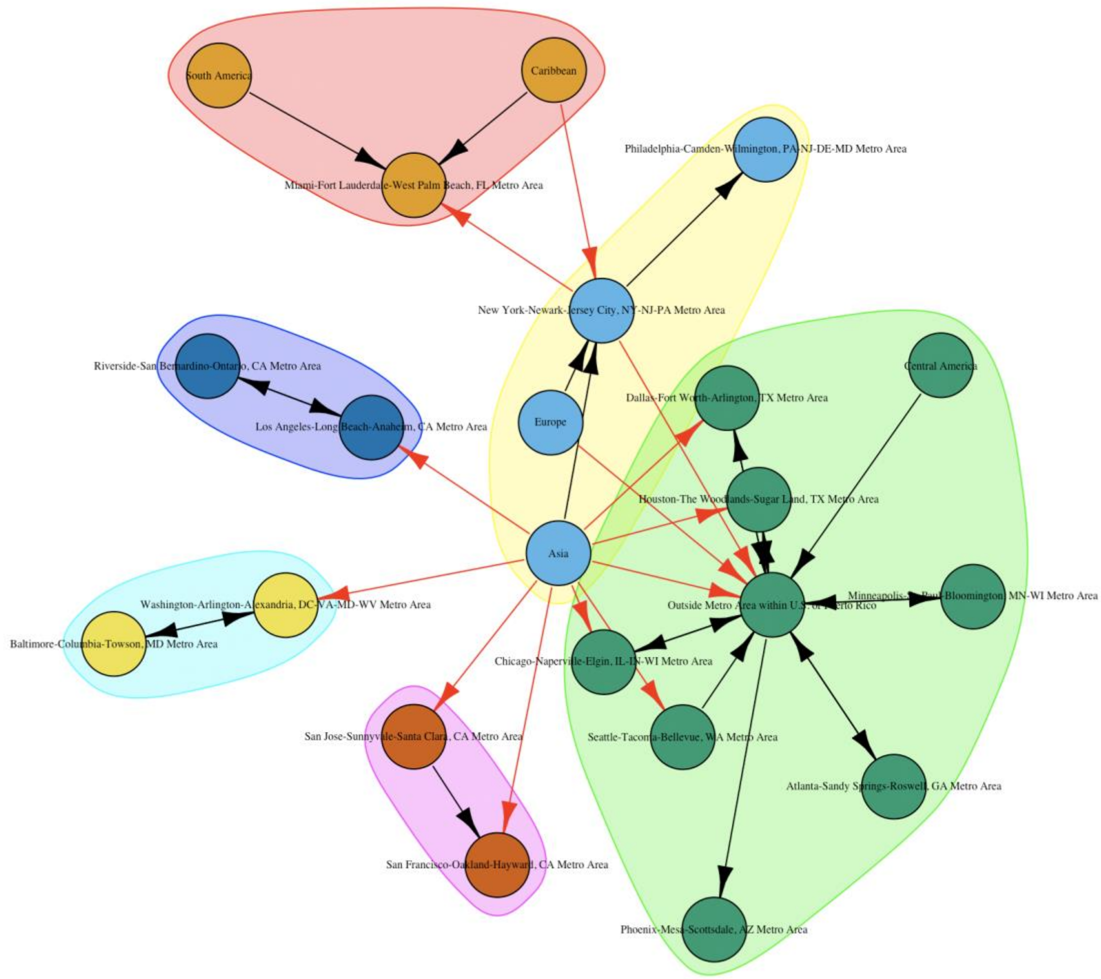
Baltimore-Columbia-Towson, MD
Atlanta-Sandy Springs-Roswell, GA

Higher scores (higher degree of difference) as expected that appeared on the cluster dendrogram include:

Washington-Arlington-Alexandria, DC-VA-MD-WV
South America

Because the network role measure is based on the cluster dendrogram, the higher the degree of difference – the more centralized the node. We expect all nodes that represent regions outside the United States to have migration flow in one direction: out. The MSA's represented within the network are economic hubs where large finance companies, technology firms, and booming tourism industries exist. It is expected that Outside Metro Area within U.S or Puerto Rico is the largest node, because more area exists in the United State outside the defined Metro Areas. Because the network is directed, a unique analysis would be to observe more in-depth of the direction and magnitude of the migration flows. For instance, Miami-Fort Lauderdale-West Palm Beach, FL receives over 21,910 individuals from three MSAs – but does not send out more than 21,910 individuals to any MSA. On the periphery of the network we can analyze nodes in which there is greater than 21,910 individuals flowing between MSAs in California or the District of Columbia and Maryland. Analyzing the community structure of the network will yield even more insight into the behavior of nodes and how they behave.

COMMUNITY STRUCTURE:



Throughout the paper we have used a top-down approach to analyze the migration flows between MSAs. We began with an analysis of an entire dataset, found an appropriate threshold to analyze the community structure and have continuously improved the network visualization. To identify communities, we use the walk-trap community detection. This algorithm detects communities based on random walks across the network. Unlike the fast and greedy community detection, the walk-trap detection can handle weighted networks.

The network above is plot with community coloring of the MSAs. Six communities were identified by the walk-trap algorithm.

```
> igraph::groups(walktrap.community(number689.ig, steps=50))
$`1`
[1] "Miami-Fort Lauderdale-West Palm Beach, FL Metro Area" "South America"
[3] "Caribbean"

$`2`
[1] "Los Angeles-Long Beach-Anaheim, CA Metro Area"   "Riverside-San Bernardino-Ontario, CA Metro Area"

$`3`
[1] "Atlanta-Sandy Springs-Roswell, GA Metro Area"       "Chicago-Naperville-Elgin, IL-IN-WI Metro Area"
[3] "Dallas-Fort Worth-Arlington, TX Metro Area"         "Houston-The Woodlands-Sugar Land, TX Metro Area"
[5] "Minneapolis-St. Paul-Bloomington, MN-WI Metro Area" "Phoenix-Mesa-Scottsdale, AZ Metro Area"
[7] "Seattle-Tacoma-Bellevue, WA Metro Area"             "Outside Metro Area within U.S. or Puerto Rico"
[9] "Central America"

$`4`
[1] "New York-Newark-Jersey City, NY-NJ-PA Metro Area"       "Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area"
[3] "Europe"

$`5`
[1] "Baltimore-Columbia-Towson, MD Metro Area"                "Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area"

$`6`
[1] "San Francisco-Oakland-Hayward, CA Metro Area"  "Asia"
[3] "San Jose-Sunnyvale-Santa Clara, CA Metro Area"
```

In analyzing the six communities, it is revealed that there are regional ties that make up communities between the MSAs. There are two main core peripheries while the periphery communities make up a set of ties between 2-3 MSAs. Most notably, two of the periphery communities exist between MSAs in California and the three-tie community is migration going into Florida from South America and the Caribbean. Another major observation is that individuals are traveling between MSAs on a regional-scale. Of the MSA ties with more than 21,910 migrating, these individuals are traveling up and down their respective east and west coast. And individuals from Outside the Metro Area tend to travel between periphery communities.

Using the walk-trap community detection algorithm on an even larger MSA network would make it more likely to find communities. Computing communities in large networks using random walk-trap is one of few statistical methods that exist. The fast and greedy method tries to detect particularly dense subgraphs by optimizing modularity scores but cannot detect weighted networks. While the goal of community detection is to identify groups of nodes with a higher density of ties within communities than between communities, a walk-trap algorithm is unique for higher weights increase the probability that random walker goes in that direction vs. the direction of a tie with a lower weight.

In looking at the 99.95 percentile of the network, I hypothesize that this network has a higher transitivity than we would expect from a random network. Transitivity refers "to the extent to which the relation that relates two nodes in a network that are connected by an edge is transitive. Perfect transitivity is rare in real networks because "it implies that each component is a clique or each pair of reachable nodes in the graph would be connected by an edge." But because we are looking at a migration network where ties equal 21,910 people or more traveling between MSAs – it is more likely that the network is connected because individuals are flowing between the highest receiving or dispersing MSAs.

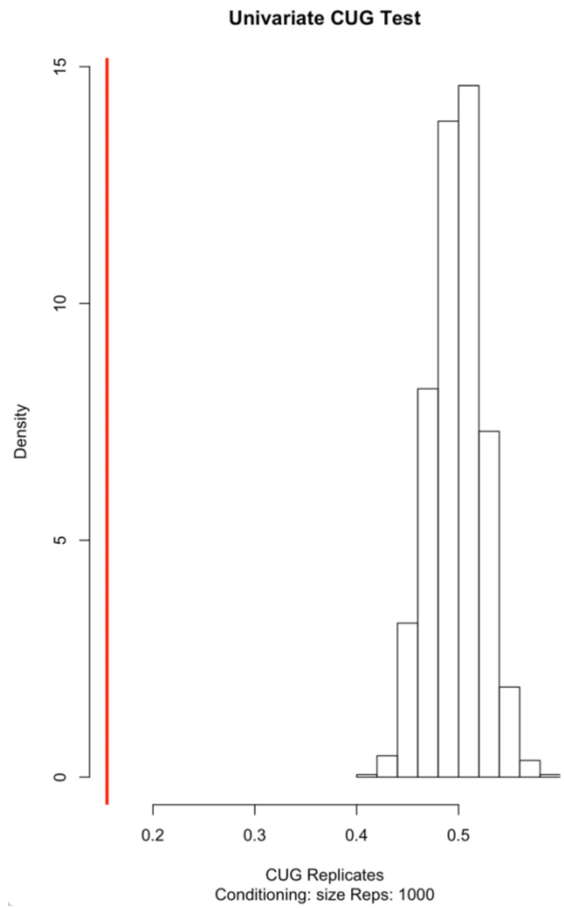
To test this hypothesis, we use the CUG-test to compare the structure of the network to baseline expectation with 1000 replications.

```
> trans.cug

Univariate Conditional Uniform Graph Test

Conditioning Method: size
Graph Type: digraph
Diagonal Used: FALSE
Replications: 1000

Observed Value: 0.1549296
Pr(X>=Obs): 1
Pr(X<=Obs): 0
```



There is a high probability that the observed network transitivity of 0.1549 could be randomly generated, conditional only on the size of the network. We feel confident accepting the hypothesis that the observed is higher than would be expected from a random network. In fact, in looking at how many standard errors away the observed transitivity value is from what we would expect, on average, we find that it is: -13.434. A negative t-value implies that the sample mean is less than the hypothesized mean.

```
> (trans.cug$obs.stat-mean(trans.cug$rep.stat))/sd(trans.cug$rep.stat)
[1] -13.43473
```

```
> igraph::transitivity(number689.ig)
[1] 0.1875
```

The transitivity score of the *number689.ig* network is: 0.1875. And the observed transitivity value is 0.1549. So, we fail to reject the hypothesis using the CUG-test.

CONCLUSION:

Looking at migration networks between MSAs across the United States is complex. Because the network was so dense, we had to set a high threshold of a 99.95 percentile. Although this yielded a less dense network and the ability to cleanly visualize the nodes and ties – we lost thousands of data points that would have significantly changed our data analysis results. Analyzing 22 nodes and 32 edges, we found there to be six distinct communities with two core peripheries and four peripheral communities. We also observed that the communities contained spatially regional MSAs and that individuals from Outside the Metro Area tend to travel between periphery communities. A limitation of the network analysis was the inability to calculate the Gould-Fernandez Brokerage. The Gould-Fernandez Brokerage would have calculated values within the network with respect not only to the pattern of ties between vertices, but also the identities of those vertices as captured in a node attribute.

Throughout our analysis of the MSA network, we continually defined the visualization of our network. The first network contained all 71613 total edges between the 391 nodes and produced a hairball of a network. Our second network visualizes the network role measure based on the cluster dendrogram. This network shows us direction of the weighted ties as well as labeled vertices. The final network visualized is produced from the community structure calculations. Labels are clearly defined, communities are highlighted, and the direction of the tie is present. This visualization gives us a clear understanding of the network.

For future research, we would investigate a network with a lower threshold. Lowering the threshold could highlight more periphery communities. We would also define the ties by proportion of total individuals flowing into an MSA. As seen below this alters the network:

> For network *pct95.ig* we calculated the proportion of all people traveling to MSA(b) divided by each MSA(a) that people are coming from. This yields the significance of MSA(a) tie to MSA(b). Next, we calculated the 95th percentile for the new proportion values within the network. Our result was 0.018. In the *pct95.ig* a tie exists if the number of people traveling into the MSA is 1.8 percent greater than the total proportion of all people traveling to that MSA. It gives us an edge count of 3597 and a node count of 391.

```
> arrange(temp1,to)
# A tibble: 71,613 x 5
   from                                         to                         weight totalin    pctin
   <chr>                                        <chr>                       <int>   <int>    <dbl>
 1 Albuquerque, NM Metro Area                   Abilene, TX Metro Area         10   15396 0.000650
 2 Amarillo, TX Metro Area                      Abilene, TX Metro Area        295   15396 0.0192
 3 Anchorage, AK Metro Area                     Abilene, TX Metro Area         49   15396 0.00318
 4 Asheville, NC Metro Area                     Abilene, TX Metro Area          0   15396 0
 5 Atlanta-Sandy Springs-Roswell, GA Metro Area Abilene, TX Metro Area         95   15396 0.00617
 6 Austin-Round Rock, TX Metro Area             Abilene, TX Metro Area        414   15396 0.0269
```

With two distinct networks: *pct95.ig* and *number689.ig* we could compare and contrast our analysis.

REFERENCES:

https://www.census.gov/topics/population/migration/guidance/metro-to-metro-migration-flows.html
https://www.sci.unich.it/~francesc/teaching/network/transitivity.html
https://www.sci.unich.it/~francesc/teaching/network/eigenvector.html
https://cs.brynmawr.edu/Courses/cs380/spring2013/section02/slides/05_Centrality.pdf