Math 168 Final Project

Charles Stoksik, Matthew Dalton, Matthew Reyes

December 2020

Abstract

In our paper we attempt both an empirical, bottom-up approach to community detection between the world's countries, and a top down analysis in the form of a MRQAP and an analysis of modularity and assortativity coefficients. Our bottom-up empirical analysis is based on Yahoo! emails and Twitter followers between countries. Using the amount of people who communicate between countries we were able to find communities that match well with previously established communities among countries. In our top down approach we analyze continent affiliation and economic forum affiliation as possible groupings, but find with both our MRQAP analysis and our assortativity coefficient analysis, that these two characterizations were not statistically significant in community detection.

1 Introduction

The world has changed a lot over the past 40 years. Many people, myself included, think about these times for what major world events occurred, such as the falling of the Berlin Wall, the ending of Apartheid, the 9/11 terrorist attacks, the assassination of Osama Bin Laden, the list goes on. But one major event sets this time period apart from the rest: The invention of the internet. The internet has given information a platform to travel around the globe, altering global dynamics and allowing any individual with internet access to retrieve information about any location on Earth. The internet also has some negative effects on society, however in our study we will be analyzing current communication between citizens of different countries using the internet. To analyze this communication we will be using a data-set retrieved from State et. al. [11].

2 Background

2.1 Describing the Data

The data we are working with in this project is a weighted graph, where nodes are countries and edges are message connections. A message connection is represented as a bi-directed communication between two people in different countries. These bi-directed communications are are obtained from two social media data-sets: Yahoo! mail and Twitter. In the Yahoo! mail data-set an edge is formed between two people if both people send at least one email to each other. In the Twitter data-set an edge is formed between two people if they follow each other. The geolocation of the Yahoo! emails was found according to the procedure in [13]. The geolocation of the Twitter followers was found according to the [12]. Yahoo! data was managed onsite by [11], sometime around or before May 2015. The Twitter data-set was obtained from [6] in 2012. The results of comparison between the Yahoo! and Twitter data-sets were comparable [11], so they were combined into one.

We now have a weighted graph of bi-directed connections between countries. However, there are a few issues with this data-set. The first and main issue, Yahoo! and Twitter are not commonly used around the globe. Some countries have more people on the internet using these services and in 2012 countries had very different levels of internet access to begin with. Therefore State attempted to normalize these edge weights through the following procedure. Denoting T_{ij} as the raw number of bi-directed communications, we will denote U_i as the predicted number of internet users in country i (obtained from internetworldstats.com), and

S_i as the number of subscribers in country i with either communication medium (provided by the relevant data-set). Therefore we can recompute T'_{ij} as:

$$T'_{ij} = T_{ij}\left(\frac{U_i}{S_i}\frac{U_j}{S_j}\right) \tag{1}$$

The normalized ties between countries in (1) are then transformed to be positive or negative based on how many ties these countries are expected to have. This is done similarly to the configuration model. Keeping degrees constant, State randomly permuted the edges such that a random graph is formed. By performing this procedure many times, an expected number of ties under randomness $\hat{T}_{ij}$ is formed. We achieve a messaging density according to the following formula:

$$D_{ij} = \log\left(\frac{T'_{ij}}{\hat{T}_{ij}}\right) \tag{2}$$

It is stated that the positive and negative weights obtained from (2) are preferable for an MRQAP model. Therefore the weighted graph we will be working with is one of nodes representing countries and edge weights with the value D_{ij} in (2).

3 Detecting Communities

This section acts as an attempt to mimic the hypothesized geopolitical communities from Huntington in [5] by working with different community detection algorithms in network science. In State's paper [11], he was able to verify Huntington's predictions by achieving a Rand Index scoring of 59% of pairings occurring in the same community. In our findings in the following section, we were unable to reproduce the same high Rand Index scoring for any method we attempted.

3.1 Restructuring Data

In order to use community detection algorithms from the third party library igraph [2], we needed to make all edge weights positive. Therefore, we decided to add a new column to the data-set specified as 'positive_Dij', according to the formula below:

$$D^{+}_{ij} = \exp\{D_{ij}\} \tag{3}$$

Equation (3) allowed us to preserve proportionality between edge weights and keep them positive at the same time. This is likely equal to the ratio of actual bi-directed ties to predicted bi-directed ties under randomness, however State never specified what base logarithm he took to produce his data-set.

Similarly to State, we experienced a lot of poor results when working with all the countries provided in the data-set, therefore we followed the same protocol and removed countries with less than 5 million predicted internet users in 2012 [11].

3.2 Huntington's Communities and Previous Results

Huntington predicted that after the Cold War ended, international relations would revert to cultural divisions that existed throughout history [5]. Therefore, in Huntington's analysis there are nine distinct communities according to the following map:

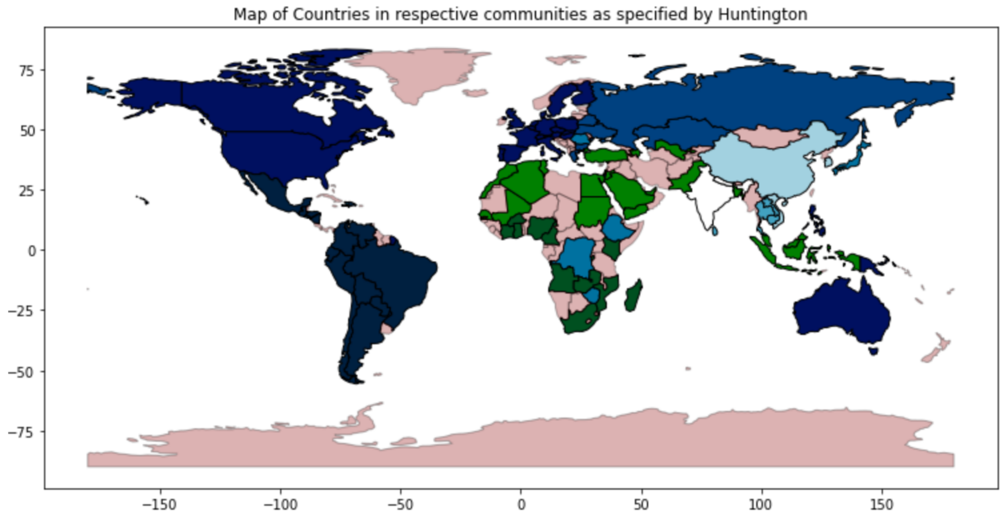

Figure 1: Huntington's Divisions on a Map. List of divisions in Appendix A

In 1 we notice clear alliances among Western Countries, Latin American Countries and Arabic Countries. We see all non-Arabic African countries with more than 5 million internet users grouped together, as well at the Orthodox European countries. It should be noted that some countries such as Japan, Romania and Zimbabwe were not classified by Huntington.

3.3 Community Detection Algorithms

For the bottom up community detection we will be using the two community detection algorithms used in the paper: the Fastgreedy Algorithm and the Spinglass Algorithm. We will also be using the Louvain Algorithm and the Leading Eigenvector Algorithm for further analysis.

Louvain and Fastgreedy are both techniques that locally optimize the modularity with each step. They are similar in that they both start with all nodes in different communities and at each step merge communities that are locally optimal for increasing modularity. The difference is that after all nodes are examined in the Louvain algorithm, the communities create become nodes in a new network. Therefore, we will expect to see similar results between the Louvain Algorithm and the Fastgreedy Algorithm, but not the same results [8] [1].

Leading Eigenvalue also maximizes the modularity function but through a top-down approach where all nodes start in the same community and are split in such a way that locally optimizes the modularity at each split. Since we were able to specify the number of communities we want in this algorithm, we expect the early stopping will cause different results to the other algorithms. The algorithms splits the network by calculating the leading eigenvector of the modularity matrix below [9]:

$$B_{ij} = A_{ij} - \frac{k_i k_j}{2m} \tag{4}$$

At each step a new modularity matrix will be created for each community and split according to the number of splits specified. In our experiments we use algorithms with splits of 2, 3, 4, and 8.

Lastly, the Spinglass algorithm was only introduced in an attempt to mirror the results of State. It does not work on modularity, which was the main focus of our optimization tasks. It is only calculated for completeness purposes [10].

3.4 Results

For the Spinglass, Fastgreedy, and Louvain algorithms our results were empirical and did not change for each simulation. Our Leading Eigenvector results did change though, so visualizations below are based on the most commonly occurring outcomes of those algorithms. Firstly we will compare the original two algorithms to the results achieved by State in [11]. Below visuallizes the communities detected using the three previously stated algorithms:

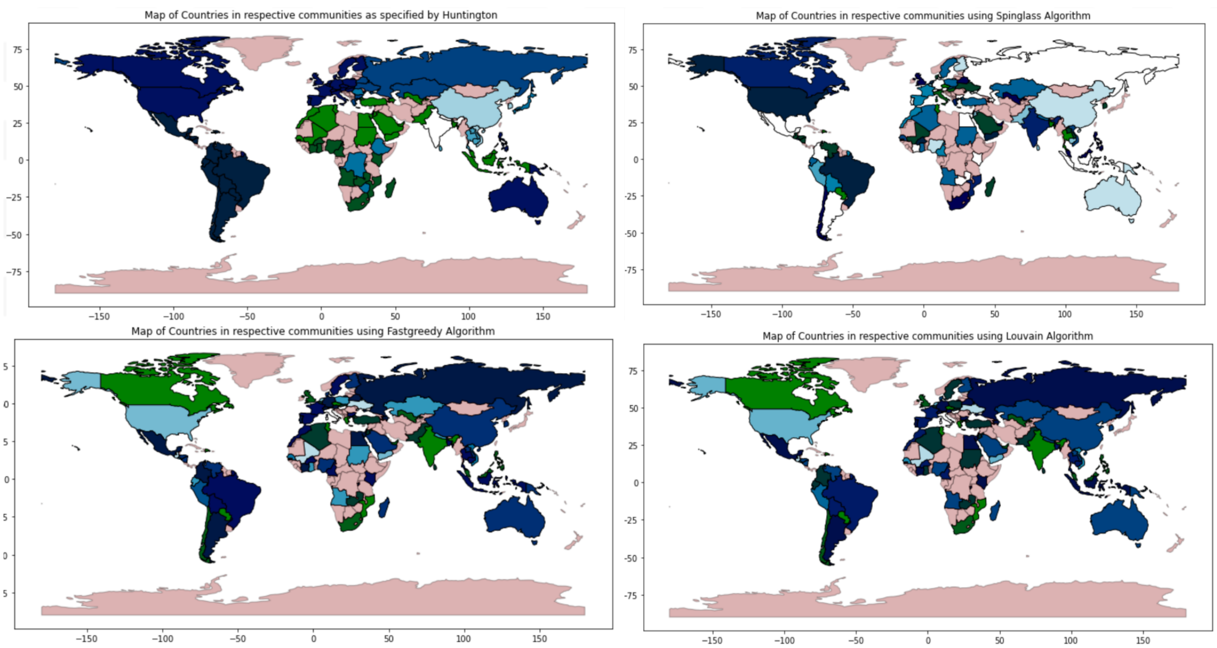

Figure 2: Top Left: Huntington, Top Right: Spinglass, Bottom Left: Fastgreedy, Bottom Right: Louvain

There are some notable commonalities among the three algorithms. They all group Australia with China, and India with Canada, despite these not being in Huntington's conclusions. However, like Huntington predicted Mexico and Argentina are in the same community along with Spain and France. There is a lot of overlap in the three graphs. Also, as predicted, there is a lot of overlap in the Louvain and Fastgreedy algorithms. They actually share a Rand Index scoring of .925, meaning that 92.5% of all pairs of nodes were both put in the same community in both results or in different communities by both results [4]. We will be using the Rand Index to quantify the changes in our community structure.

This being said, there is significant correlation between these three models and the communities proposed by Huntington according to Rand Index Scores. In fact as we will see in **??**, the Rand indexes between these three algorithms predicted the community structure proposed by Huntington better than the forced leading Eigenvector techniques.

Further experiments with limiting the number of categories tended to lend slightly more favorable results in finding communities. With the Leading Eigenvector community detection we were able to find two, three, four, and eight communities. Our results from these respective algorithms are visualized below.

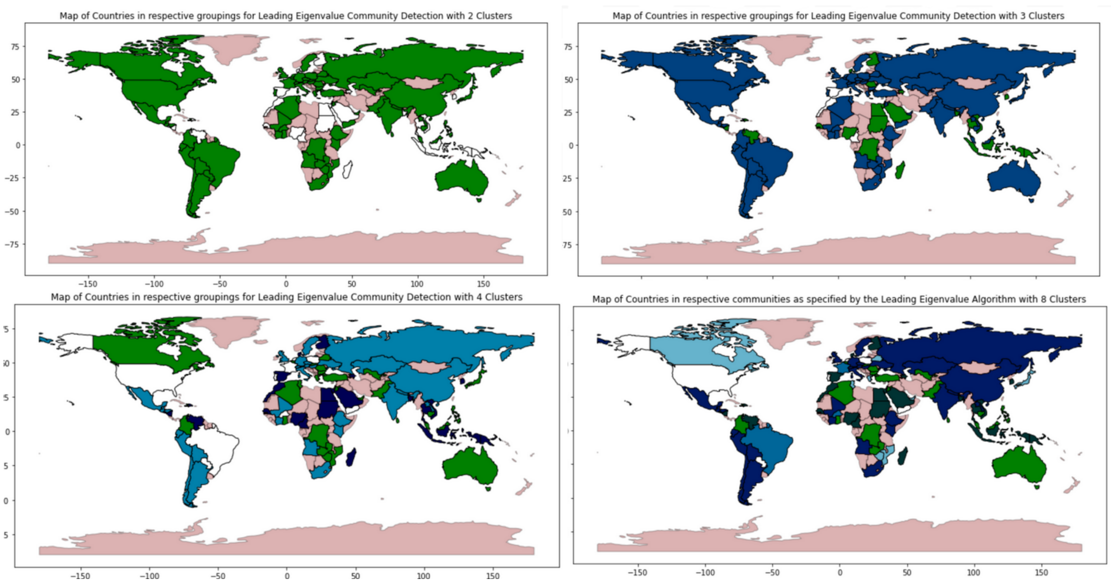

Figure 3: Top Left: 2 Clusters, Top Right: 3 Clusters, Bottom Left: 4 Clusters, Bottom Right: 8 Clusters

The maps in 3 were not the only results we got from this algorithm. There appears to be a stochastic aspect, but these were the mappings that resulted in the best Rand Index when compared to Huntington's communities. Like the previous mappings, we can postulate different reasons for certain connections. For instance, more developed countries appear green in the 2 clusters, and blue in the 3 clusters, but it is clear that the groups do not reflect that of Huntington's predictions. The results from forcing eight communities performed very well relative to the other Eigenvector algorithms. It should also be noted that the 3 Cluster and 4 Cluster algorithms resulted in effectively the same communities with one community being split into two parts. As previously promised, below are the Rand Indices ([similar pairings + dissimilar pairings] / total possible pairings) of the different models [4]:

	Huntington	Fastgreedy	Spinglass	Louvain	L.E. 2 Clusters	L.E. 3 Clusters	L.E. 4 Clusters	L.E. 8 Clusters
Huntington	1.0	0.762	0.78	0.749	0.43	0.448	0.448	0.693
Fastgreedy	0.762	1.0	0.91	0.925	0.428	0.445	0.445	0.742
Spinglass	0.78	0.91	1.0	0.916	0.437	0.46	0.46	0.739
Louvain	0.749	0.925	0.916	1.0	0.428	0.452	0.452	0.733
L.E. 2 Clusters	0.43	0.428	0.437	0.428	1.0	0.976	0.976	0.608
L.E. 3 Clusters	0.448	0.445	0.46	0.452	0.976	1.0	1.0	0.604
L.E. 4 Clusters	0.448	0.445	0.46	0.452	0.976	1.0	1.0	0.604
L.E. 8 Clusters	0.693	0.742	0.739	0.733	0.608	0.604	0.604	1.0

Table 1: Rand Index Scores rounded to nearest thousandths place

4 MRQAP Model

In contrast to the previous section, which employed a bottom-up approach for community detection, this section and its successor analyze the existence and influence of international communities via a top-down approach. That is, prior to any statistical analysis, we decide upon known international communities whose influences we want to measure with respect to the communication densities observed. Specifically, this section uses Multiple Regression Quadratic Assignment Procedures (MRQAP). We first provide some background on the procedure before describing the international communities we analyze and finally providing our results.

4.1 What is an MRQAP?

It can be extremely useful to attempt to model collected data as a function for myriad reasons, among them being able to concisely describe the distribution observed. A popular statistical technique for doing so is regression. For example, suppose we wanted to create a model to accurately predict the party affiliations of politicians. We could collect data on which party a certain politician belongs to, as well as on which other politicians they collaborate with. We could make the reasonable hypothesis that politicians who collaborate are more likely to belong to the same party, and thus we could run a linear regression to better understand the predictive power of collaboration.

We could model each variable by a matrix. Let N be the number of politicians. Then, we could create an $N \times N$ matrix for collaboration, where the $(i, j)^{th}$ and $(j, i)^{th}$ entries are 1 if and only if politician i collaborates with politician j, and -1 otherwise. Call this matrix X. We could execute a similar process for party affiliations. Call this matrix Y. Then, we essentially have two N-node networks, where matrix entries are edge weights.

After running linear regression, we will obtain an affine equation $Y' = \beta_0 + \beta_1 X$ that best fits the data. The issue is, observations in this context are *not* independent of each other; if politician a collaborates with politician b, and politician b collaborates with politician c, then politicians a and c are more likely to collaborate. Moreover, observations in a single row or a single column tend to be highly correlated, meaning that errors are not random [3].

To combat this issue, one must create a null hypothesis against which to compare observations. This is where a QAP comes into play. Recall that these matrices can alternatively be interpreted as networks. The QAP idea is simple: randomly swap labels on vertices in the independent variable network to create a random sample and then perform linear regression. After several iterations on this procedure, we will have obtained a sample distribution against which we can test if our original linear regression model is different enough to conclude that our model indeed has predictive power. In terms of matrices, ”swapping the labels” amounts to randomly permuting the rows and columns, making sure that the indices are consistent across rows and columns (i.e. if the i-th row moves to the k-th spot, then so does the i-th column), otherwise the resulting adjacency matrix would be nonsense.

An MRQAP is analogous to the QAP described above for more than one independent variable. That is, rather than performing linear regression, it performs multiple regression. In order to implement this procedure in our project, we leveraged the `mrqap-python` package created by Lisette Espin.

4.2 Communities

The purpose of our MRQAP is to discover the influence, if any, that particular international communities have over the communication densities between nations. Specifically, we are interested in how geographic location and economic status affect how much nations communicate with each other. To capture geographic proximity, we say that the weight of an edge between two countries is 1 if they are situated on the same continent, and -1 if not. To capture economic status, we used the G20 and G33 international alliances as proxies. The G20 includes the world’s leading industrial states, while the G33 includes developing countries. As before, the weight of an edge between two countries is 1 if they participate in the same forum, and -1 if not.

Furthermore, we only consider the 90 countries with a population greater than 5 million in an attempt to avoid statistical skews or errors due to lack of data.

4.3 MRQAP Results

In order to get a good grasp on how our independent variables affect communication densities, we perform an MRQAP on three different models: (1) just continents, (2) just G20 and G30 affiliations, and (3) both.

Table 2 displays our results.

	Model 1	Model 2	Model 3
Intercept	-4.2607	-4.4064	-4.2678
Continents	0.2485		0.2486
G20 and G33		-0.0044	-0.0089

Table 2: MRQAP beta coefficients

Note that the coefficients for both *Continents* and *G20 and G33* are both very near zero, and that neither of their presences has much of an affect on the intercept. This leads us to accept the null hypothesis that neither sharing a continent nor membership in G20 or G33 has an effect on the communication density between densities. Perhaps the most important detail, however, is that none of these results are statistically significant. That is, we accept the null hypothesis that shared continent and shared international forum have no influence over the communication densities between countries.

We would like to point out that there are other independent variables that would, in theory, be a better predictor for communication densities: shared language, shared religion, shared border, etc. However, [11] does a good job of covering these more obvious predictors; the purpose of our research was to check the contributions of more peripheral commonalities.

5 Assortative Mixing

Similar to the previous section, we will be analyzing the existence and influence of international communities via a top-down approach. We will use assortative mixing to study network structure, using modularity and the attribute assortativity coefficient.

5.1 Modularity

Let G be a graph modelling our network. Modularity measures the extent to which similar nodes are connected to each other. It is defined as

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

where m is the total weight of all edges, A is the adjacency matrix of G, k_i is the sum of the weights connected to node i, and $\delta(c_i, c_j)$ is the Kronecker delta function, taking a value of 1 if $i = j$, and 0 otherwise. If modularity is positive, then the network demonstrates a tendency to connect to similar nodes. If it's negative, then the network demonstrates a tendency to connect to different nodes. If it's close to zero, the network demonstrates little to no such tendency. We will be partitioning countries based on two factors: (1) continent, and (2) G20 and G33 affiliations. Furthermore, like the previous section, we will only consider countries with a population greater than 5 million.

5.2 Attribute Assortativity Coefficient

The attribute assortativity coefficient is a measure of correlation between connected nodes based on similar attributes. It is defined as [7]

$$r = \frac{\sum_i (M_{ii}) - \sum_i (a_i b_i)}{1 - \sum_i (a_i b_i)}$$

$$a_i = \sum_j M_{ij} \qquad b_j = \sum_i M_{ij}$$

where M is the mixing matrix for the specified attribute, with M_{ij} representing the fraction of edge weight connecting nodes of type i to type j. The attribute assortativity coefficient r only takes values from -1 to 1. An r-value of 1 indicates a perfectly assortative network, one that has edges between nodes of the same attribute. An r-value of -1 indicates a perfectly dissortative network, one that has edges between nodes of different attributes. An r-value close to 0 indicates that there may be no correlation at all based on the selected attributes. Here, we are using the same as the attributes used to compute modularity: (1) continents, and (2) G20 and G33 affiliations.

5.3 Assortative Mixing Results

We will be using $D^+_{ij} = exp\{D_{ij}\}$ to indicate messaging densities. That way, edge values indicate a multiplier for how much higher or lower the normalized data was than in the random graph. Like the previous part, we only consider the countries with a population greater than 5 million. We compute modularity and attribute assortativity coefficient based on two attributes: (1) continents, and (2) G20 and G33 affiliations. Table 3 displays our results:

	Modularity	AAC
Continents	-0.0287	-0.0115
G20 and G33	0.0515	-0.0115

Table 3: Modularity and Attribute Assortativity Coefficient results

For both *Continents* and *G20 and G33*, both measures of modularity and attribute assortativity coefficient are very close to zero, meaning neither of their presences affects any assortative mixing across the network. Like in the previous section, we come to the conclusion that neither sharing a continent nor membership in G20 or G33 has an effect on the communication density between countries.

6 Conclusions

We were able to recreate State's [11] bottom up analysis of Huntington's communities based on empirical communication data. We were also able to show that Continent affiliation and economic forum affiliation were not significant predictors of community structure in the geopolitical messaging network. To see our code please visit: www.github.com/mattdalton/meshOfCivilizations.

Appendix A: Communities of Countries according to the Bottom-Up Algorithms:

Huntington

1. Morocco, Egypt, Turkey, Mali, Jordan, Azerbaijan, Algeria, Uzbekistan, Yemen, Saudi Arabia, Sudan, Senegal, Tunisia, Indonesia, Bangladesh, United Arab Emirates, Malaysia, Pakistan
2. Mozambique, Cameroon, Ghana, Madagascar, South Africa, Nigeria, Zambia, Burundi, Cote D'Ivoire, Kenya, Angola, Burkina Faso
3. Ecuador, Bolivia, Nicaragua, Guatemala, El Salvador, Colombia, Dominican Republic, Brazil, Paraguay, Peru, Honduras, Mexico, Chile, Argentina, Venezuela
4. Canada, France, Austria, Netherlands, Belgium, Denmark, Papua New Guinea, Slovakia, Portugal, Sweden, Spain, Italy, Philippines, Germany, Australia, United States, Israel, Poland, Hungary, Czech Republic, Switzerland, Finland, United Kingdom
5. Ukraine, Greece, Bulgaria, Russia, Kazakhstan, Belarus
6. Haiti, Zimbabwe, Romania, Japan, Ethiopia, The Democratic Republic Of Congo
7. Cambodia, Sri Lanka, Thailand, Singapore, Laos
8. China, Republic Of Korea, Vietnam
9. India, Nepal

Fastgreedy

1. Kenya, Cambodia, Morocco, Bolivia, France, Romania, Cameroon, Bulgaria, Israel, Brazil, Spain, Sweden, Thailand, Indonesia, Bangladesh, The Democratic Republic Of Congo, Burkina Faso
2. Zimbabwe, Nicaragua, China, Republic Of Korea, Australia, Madagascar, Saudi Arabia, Nigeria, Finland, Honduras, Denmark, Venezuela, Papua New Guinea, Laos
3. Canada, India, Greece, Dominican Republic, Czech Republic, Paraguay, Tunisia, Mozambique
4. Portugal, Egypt, Austria, Cote D'Ivoire, Colombia, Japan, Germany, Azerbaijan, Russia, Switzerland, Mexico, Burundi, Argentina, Belarus
5. Philippines, Uzbekistan, South Africa, Chile, United Arab Emirates, Malaysia, United Kingdom
6. Haiti, Turkey, Jordan, Algeria, Pakistan, Sri Lanka, Zambia, Slovakia, Belgium
7. Ecuador, Nepal, Kazakhstan, Sudan, Singapore, Ghana, Poland, Angola
8. Peru, Ethiopia
9. Yemen, Netherlands, United States
10. Hungary, Italy
11. Ukraine, Mali, Guatemala
12. Vietnam, Senegal, El Salvador

Louvain

1. Cambodia, Morocco, Bolivia, France, Romania, Cameroon, Brazil, Spain, Russia, Israel, The Democratic Republic Of Congo, Burkina Faso
2. Zimbabwe, Nicaragua, Nepal, China, Republic Of Korea, Kazakhstan, Australia, Madagascar, Saudi Arabia, Nigeria, Finland, Honduras, Denmark, Venezuela, Singapore, Papua New Guinea, Angola, Laos
3. Canada, India, Greece, Dominican Republic, Czech Republic, Paraguay, Tunisia, Thailand, Bangladesh, Mozambique
4. Portugal, Egypt, Austria, Cote D'Ivoire, Bulgaria, Japan, Germany, Azerbaijan, Switzerland, Mexico, Burundi, Indonesia, Argentina, Kenya, Belarus
5. Philippines, Uzbekistan, South Africa, Chile, United Arab Emirates, Malaysia, United Kingdom
6. Ecuador, Haiti, Sweden, Colombia, Turkey, Jordan, Algeria, Pakistan, Sri Lanka, Sudan, Zambia, Slovakia, Ghana, Poland, Belgium
7. Peru, Ethiopia
8. Yemen, Netherlands, United States
9. Hungary, Italy
10. Ukraine, Mali, Guatemala
11. Vietnam, Senegal, El Salvador

Spinglass

1. Peru, The Democratic Republic Of Congo, Ethiopia, Romania
2. Dominican Republic, Czech Republic, Paraguay, Tunisia, Thailand, Bangladesh, Singapore
3. Portugal, Egypt, Austria, Cote D'Ivoire, Bulgaria, Colombia, Germany, Azerbaijan, Japan, Switzerland, Mexico, Zambia, Burundi, Indonesia, Argentina, Kenya, Belarus
4. Philippines, Uzbekistan, South Africa, Chile, United Arab Emirates, Malaysia, United Kingdom
5. Hungary, Italy
6. Yemen, Netherlands, United States
7. Nicaragua, Republic Of Korea, Madagascar, Saudi Arabia, Honduras, Denmark, Venezuela, Laos
8. Ecuador, Nepal, Sweden, Turkey, Kazakhstan, Algeria, Sudan, Ghana, Poland, Angola, Slovakia
9. Haiti, Cambodia, Brazil, Jordan, Russia, Pakistan, Sri Lanka, Burkina Faso, Belgium
10. Zimbabwe, China, Australia, Nigeria, Finland, Papua New Guinea
11. Vietnam, Senegal, El Salvador
12. Canada, India, Guatemala, Ukraine, Greece, Mali, Mozambique
13. Morocco, Bolivia, France, Cameroon, Spain, Israel

Leading Eigenvalue 2 Clusters

1. Vietnam, Romania, Dominican Republic, Republic Of Korea, Denmark, Bangladesh, Papua New Guinea, Egypt, Spain, Saudi Arabia, Cameroon, Nigeria, Sudan, Senegal, Thailand, Venezuela, Morocco, Nicaragua, El Salvador, Azerbaijan, Madagascar, Finland, Indonesia, The Democratic Republic Of Congo
2. Canada, France, Austria, Guatemala, Brazil, Jordan, Russia, Netherlands, Uzbekistan, Peru, Zambia, Burundi, South Africa, Zimbabwe, United Kingdom, Burkina Faso, Malaysia, Ecuador, Mozambique, Cambodia, Portugal, Slovakia, Bolivia, Sweden, Ukraine, Japan, China, Mali, Algeria, Cote D'Ivoire, Belarus, Haiti, Italy, India, Nepal, Bulgaria, Greece, Germany, Philippines, Kazakhstan, Australia, Paraguay, Turkey, Honduras, Mexico, Tunisia, United States, Chile, Israel, Kenya, Poland, Singapore, Angola, Laos, Pakistan, Hungary, Yemen, Colombia, Czech Republic, Ethiopia, Ghana, Switzerland, Sri Lanka, Argentina, United Arab Emirates, Belgium

Leading Eigenvalue 3 Clusters

1. Morocco, Romania, Cameroon, Spain, The Democratic Republic Of Congo
2. Canada, France, Austria, Guatemala, Brazil, Jordan, Russia, Netherlands, Uzbekistan, Peru, Zambia, Burundi, South Africa, Zimbabwe, United Kingdom, Burkina Faso, Malaysia, Ecuador, Mozambique, Cambodia, Portugal, Slovakia, Bolivia, Sweden, Ukraine, Japan, China, Mali, Algeria, Cote D'Ivoire, Belarus, Haiti, Italy, India, Nepal, Bulgaria, Greece, Germany, Philippines, Kazakhstan, Australia, Paraguay, Turkey, Honduras, Mexico, Tunisia, United States, Chile, Israel, Kenya, Poland, Singapore, Angola, Laos, Pakistan, Hungary, Yemen, Colombia, Czech Republic, Ethiopia, Ghana, Switzerland, Sri Lanka, Argentina, United Arab Emirates, Belgium
3. Vietnam, Dominican Republic, Republic Of Korea, Denmark, Bangladesh, Papua New Guinea, Egypt, Saudi Arabia, Nigeria, Sudan, Senegal, Thailand, Venezuela, Nicaragua, El Salvador, Azerbaijan, Madagascar, Finland, Indonesia

Leading Eigenvalue 4 Clusters

1. Morocco, Romania, Cameroon, Spain, The Democratic Republic Of Congo
2. Canada, France, Austria, Guatemala, Brazil, Jordan, Russia, Netherlands, Uzbekistan, Peru, Zambia, Burundi, South Africa, Zimbabwe, United Kingdom, Burkina Faso, Malaysia, Ecuador, Mozambique, Cambodia, Portugal, Slovakia, Bolivia, Sweden, Ukraine, Japan, China, Mali, Algeria, Cote D'Ivoire, Belarus, Haiti, Italy, India, Nepal, Bulgaria, Greece, Germany, Philippines, Kazakhstan, Australia, Paraguay, Turkey, Honduras, Mexico, Tunisia, United States, Chile, Israel, Kenya, Poland, Singapore, Angola, Laos, Pakistan, Hungary, Yemen, Colombia, Czech Republic, Ethiopia, Ghana, Switzerland, Sri Lanka, Argentina, United Arab Emirates, Belgium
3. Vietnam, Dominican Republic, Republic Of Korea, Denmark, Bangladesh, Papua New Guinea, Egypt, Saudi Arabia, Nigeria, Sudan, Senegal, Thailand, Venezuela, Nicaragua, El Salvador, Azerbaijan, Madagascar, Finland, Indonesia

Leading Eigenvalue 8 Clusters

1. Romania, Jordan, Uzbekistan, Zambia, Malaysia, Slovakia, Cambodia, Algeria, Haiti, Bulgaria, Philippines, Turkey, Australia, Pakistan, Colombia, Sri Lanka, Ghana, The Democratic Republic Of Congo, Belgium
2. Vietnam, Dominican Republic, Republic Of Korea, Denmark, Bangladesh, Papua New Guinea, Egypt, Spain, Saudi Arabia, Cameroon, Nigeria, Sudan, Senegal, Thailand, Venezuela, Morocco, Nicaragua, El Salvador, Azerbaijan, Madagascar, Finland, Indonesia
3. Austria, Guatemala, Russia, Peru, Burundi, South Africa, Ecuador, Bolivia, Sweden, Ukraine, China, Mali, Cote D'Ivoire, Italy, India, Nepal, Germany, Kazakhstan, Honduras, Mexico, Chile, Israel, Kenya, Angola, Laos, Hungary, Czech Republic, Ethiopia, Switzerland, Argentina, United Kingdom
4. Brazil, France, Burkina Faso
5. Zimbabwe, Canada, Mozambique, Japan, Greece, Tunisia, Singapore, United Arab Emirates, Belarus
6. Portugal, Paraguay, Netherlands, Yemen, United States, Poland

References

[1] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, Oct 2008.

[2] G. Csárdi and T. Nepusz. The igraph software package for complex network research. *InterJournal Complex Systems*, page 1695, 2006.

[3] Snijders T Dekker D, Krackhardt D. Multicollinearity robust qap for multiple regression. *CASOS Working Paper*, 2003.

[4] L. Hubert and P. Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.

[5] Samuel P. Huntington. *The Clash Of Civilizations And The Remaking Of World Order*. Simon & Schuster, January 1998.

[6] Juhi Kulshrestha, Farshad Kooti, Ashkan Nikravesh, and Krishna Gummadi. Geographic dissection of the twitter network, 2012.

[7] M. E. J. Newman. Mixing patterns in networks. *Physical Review E*, 67(2), Feb 2003.

[8] M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6), Jun 2004.

[9] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3), Sep 2006.

[10] Jörg Reichardt and Stefan Bornholdt. Statistical mechanics of community detection. *Physical Review E*, 74(1), Jul 2006.

[11] Bogdan State, Patrick Park, Ingmar Weber, and Michael Macy. The mesh of civilizations in the global network of digital communication. *PLOS ONE*, 10(5):1–9, 05 2015.

[12] Bogdan State, Ingmar Weber, and Emilio Zagheni. Studying inter-national mobility through ip geolocation. pages 265–274, 02 2013.

[13] Samuel A Stouffer. Intervening opportunities: a theory relating mobility and distance. *American sociological review*, 5(6):845–867, 1940.