Climate Change Policy in Network Science

Robert Sunderhaft, Siddharth Engineer, Harshi Kosaraju

Motivation

Climate change is one of the most pressing problems the world faces today. The first step to finding a solution to mitigate the problem is to figure out which countries are falling behind in contributing to the fight against climate change. Our project aimed to find answers to the questions listed below by analyzing the policies put forth by various nations around the world using network science concepts.

The questions we were trying to answer included questions such as:

- Which countries have the most influence in creating climate change policies?
- Which countries tend to make policies together?
- Are all countries contributing or are there any free riders?

With answers to the questions above, we can make recommendations for actions that need to occur by specific players in public policy.

Introduction to Our Project

Our project specifically focused on creating graphs that would map out patterns in international climate policy. We did so by first collecting the necessary data on climate change policies and then choosing specific policy types to focus on. After collecting the data, we performed various analyses on the data such as temporal analysis on the butterfly-caterpillar ratio for fossil fuels, change point detection on the number of countries implementing green bonds and etc. The analyses led to the discovery of many interesting results as we will show in the following pages.

Methodology

This section details each task we completed as well as the technical methodology we used to complete each task. Each task also has references to papers which we based our project off of.

1. Data

The data we used came from the International Monetary Fund or IMF for short. The IMF hosts a "Climate Change Dashboard" that is home to an extensive number of datasets related to different types of climate policies. We specifically used the IMF data because the data was easy to use and didn't require much preprocessing.

2. Graph Building

Due to the data being easy to use, building graphs to represent it, didn't take much effort to build the graphs. We built the graphs in Python using the NetworkX package. We used the NetworkX package because it is commonly used to conduct network analyses and contains many built-in methods to analyze graphs.

3. Clustering Coefficient

We want to use this section to highlight one of the key metrics we used in evaluating the clustering capabilities of our bipartite model. Note that the general 'clustering coefficient' value will return 0 in a bipartite setting. This is because a bipartite graph cannot contain triangles. Instead, we found a new way of defining instances (Kolda et al).

4. Temporal Analysis

While we have determined that a bipartite graph is able to model our country-policy pairings quite well, we still have a large amount of data to look at. Recall that we are working with 4 different graphs (fossil fuel subsidies, green bonds, environmental taxes, and environmental protection expenditures). Therefore finding the connections between years and the four datasets from before was the best way to see how countries' policies changed over time.

5. Influence Passivity

We recognize that the network we are working with is a form of a social network. Countries implement policies in coordination with one another. We wish to trace the impact of countries participating in the global effort to mitigate climate change. This boils down to both the immediate impact of one's policies, as well as the social influence on others to implement similar policies. We decided to implement the influence and passivity algorithm from Kolda et al.

The idea of this algorithm is to decompose the reactions that people make on Twitter. Instead of basing influence off of something like followers, they would base influence on reusing a link that was tweeted by a user. In the case of our project, we treat implementing a policy that a country has previously implemented as "being influenced" by the first country. This was done in order to determine which countries have the most influential role and which countries have the most passive role in influencing climate change policies.

6. Change Point Detection:

The first task that we completed was change point detection, which is a process that identifies what years marked a turning point in policy usage. The team identified this as an important task because if we are able to identify what years there is change, then we can identify who is causing this change.

Our implementation for change point detection was based on MILA's *Laplacian Change Point Detection for Dynamic Graphs* paper. The authors accomplished change point detection by a simple _ step process. A summary of the algorithm that was implemented is below:

1. For each year, construct the Laplacian of the graph.
    a. NOTE: Although our graph is a bipartite graph, meaning out Laplacian will be sparse, we still used the normal Laplacian because we found it to have better results.
2. Perform the singular value decomposition of each Laplacian to obtain the singular values for each year. These singular values essentially are the fingerprint of a given year. Two sets of singular values that are very similar mean those two corresponding years are very similar. We will call the set of singular values for each year the signature of the year.
3. An arbitrary year is chosen and this is considered present day. The algorithm then calculates the average signature from the previous W years before present day, this is considered our baseline. The variable W is a hyperparameter that denotes how many years we will average over when calculating our baseline.
4. The cosine similarity is then calculated between the baseline and present day. A high cosine similarity denotes a given year has changed much compared to the last W years and a low cosine similarity denotes a large change.

This algorithm was implemented from the paper with Python because the original code the paper provided was in C and MatLab.

7. Changes in Country Policy:

After we complete change point detection, we will be able to find what years had the most change. But this leads to another question of what caused this change. To answer this query, it would be helpful to find what country made the most changes in their policy implementation over the years. This problem is very analogous to change point detection for a given year, but we are applying it at the individual country level instead.

The process for calculating the change of a country is simpler than the overall change in years. First we construct a 1-D feature vector for each country given a year that contains the information on what policies they implemented. For example, if there are two policies, the US in 2020 would have a feature vector that might look like $[0.45, 0.65]^T$ where 0.45 is the weight the US puts on policy 1 and 0.65 is the weight that the US puts on policy 2.

Once we get the feature vectors for each year and country, we simply compute the Euclidean distance between a given year and its previous W years, where W is again a hyperparameter denoting the averaging window of the baseline years to compare to.

The country that has the highest distance between a given year and its baseline is the country that has changed the most over time. We then infer that this country is the main country contributing to the overall change in that year. It should be noted that the hyperparameter for W should match that of the change point detection for overall years, as different values of W can lead to different results.

8. Clustering:

Another key task that must be accomplished to find which countries are the most influential in enacting change is by discovering the clusters of countries that make change together.

Currently, we have built out the framework for identifying clusters for a given year but we have not expanded upon this work for finding how clusters change over the years.

For finding clusters, we performed a form of spectral clustering using BINE embeddings, which are embeddings that are better situated for vertex representation in a bipartite network. We used the paper and code from Learning Vertex Representations for Bipartite Networks to compute the BINE embeddings for each vertex in our graph.

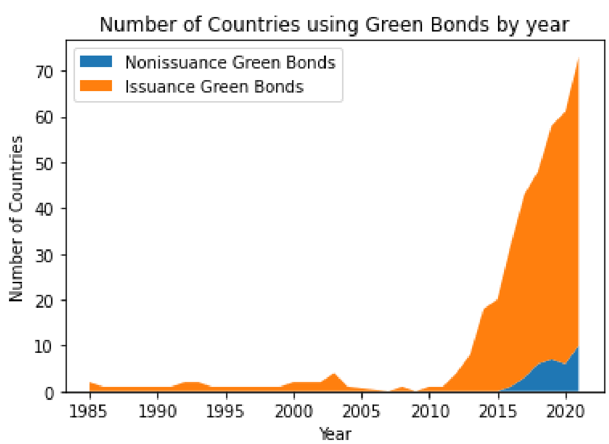
After we computed the BINE embeddings for each node, we performed K-Nearest Neighbors to cluster similar countries together. To find the optimal K, we iteratively tested what value for K would give the highest silhouette score.

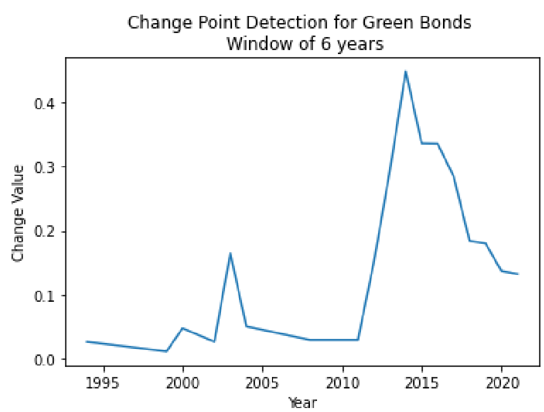
Results

1. Change Point Detection:

The team completed change point detection for 3 out of the 4 major types of policies. Here are the results:

- **Green Bonds:**

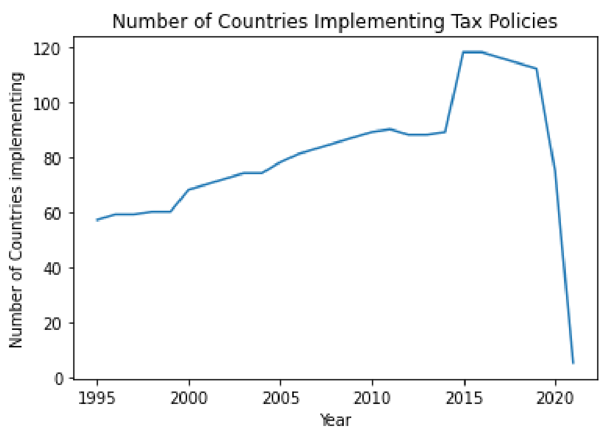


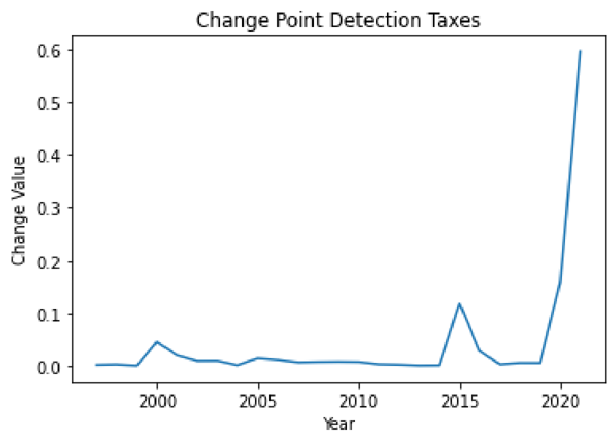


The green bonds data was the perfect data to start on because there were only two green bond policies, meaning we could easily predict how the network was going to change. This allowed us to experiment with how change point detection worked and to tune it to work on the data properly.

The leftmost graph denotes the number of countries that used green bonds by year. This graph clearly shows that there is a significant increase in green bond usage around 2012, so we hoped our change point detection would pick up on this obvious change. The rightmost graph is our change point detection graph showing how much predicted change each year had. It has clear spikes around 2014 and 2003. This is great because it correctly predicted the observed change around 2012 and even picked up on the minor change in 2003.

- **Environmental Taxes:**
  We then test change point detection on the Environmental tax dataset which is a much more challenging dataset because it has roughly 10 different policies, making overall change prediction presumably more challenging.

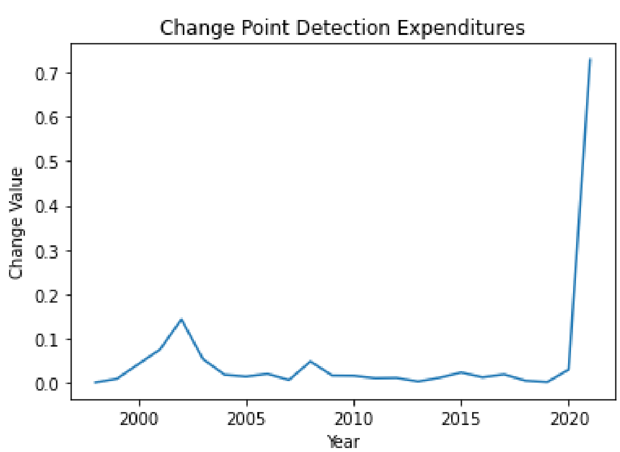




  After we ran the change point detection algorithm, we realized that it is really only observing the very macro changes in the network, like the number of edges in the graph. We can illustrate this fact by looking at the graph depicting the total number of countries implementing tax policies. We can see that there are 2 major changes. One right before 2015 and one right before 2020. We can see that the change point detection clearly shows these changes, as well as the small changes in 2000.

  It should be noted that the large change in 2020 is because the data for 2020 and beyond hasn't been fully entered yet.

- **Environmental Protection Expenditure:**

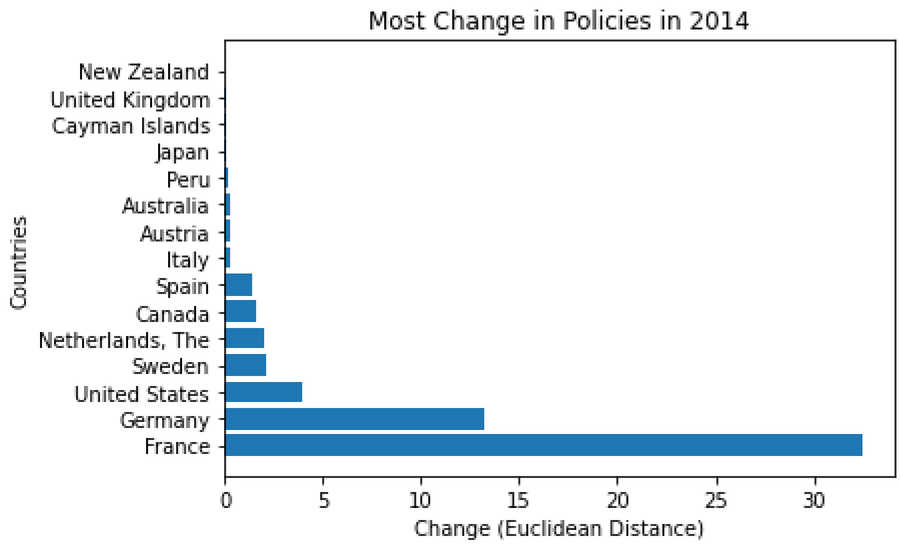


Lastly when we run change point detection for the environmental protection expenditures, we see a similar graph to the environmental tax graph, one with two minor changes and a large change around 2020. We can infer from our research that this change is also related to delayed data entry issues as well.

- **Change Point Summary**
  In total it seems that the change point detection graphs only describe very macro scale changes like new edge connections. This is great for easily detecting when a new group of countries implements a policy, but it doesn't really give us novel insights as one could easily derive these conclusions from other graphs as well.

2. Changes in Country Policy:

The code for gauging what country had the most change in a year was written during the finals period, so with these time constraints we were only able to run the code on the green bonds data. The team chose to run our algorithm on the green bonds data first because of its simplicity and ease to assess. We first decided to test the change in the 2013-2015 range, which was about the time the large amount of change was detected.

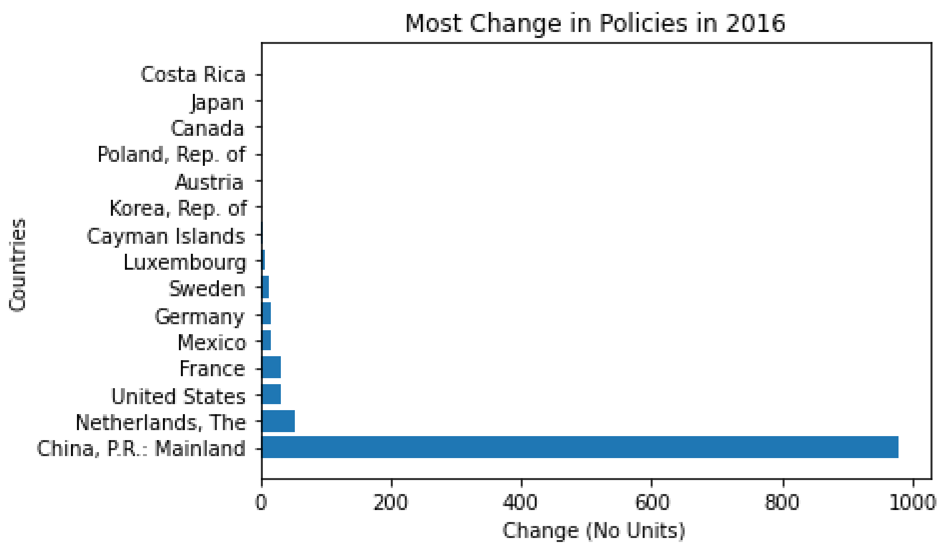


The graph above denotes the change from 2014 to the previous 6 years. As seen above, France had the greatest change in policy with Germany and the US following right behind. These three countries make sense as they generally are at the forefront of climate policy. The table below denotes the amount each country invested in green bonds during the given time span:

Country	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
France	0.650	2.202	6.581	9.234	6.299	25.438
Germany	0	0.340	3.779	7.088	4.436	8.463
US	0	0.5	2.144	8.263	6.075	5.506

As seen from above, we can verify that France had the largest increase. But an interesting point arises from this table. It is that all countries on this list decreased their investments in Green Bonds in 2016, yet this will still show up as a change in graph. So it should be noted that a large change doesn't always denote a good change, but in the year 2014, it seems that the biggest changes at a country level were positive changes.

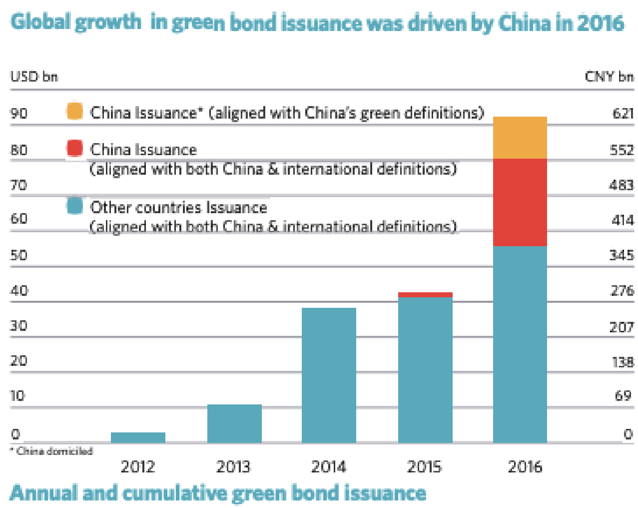
Another very interesting graph is during 2016. While France, Germany, and the US all decreased their investments, another global superpower increased their investments, significantly. The following graph depicts this change.



As seen above, China had an incredible change in 2016. When we look at the data we will see that it was a positive change as well.

Country	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
China	0	0	0.157	0.994	31.303	31.259

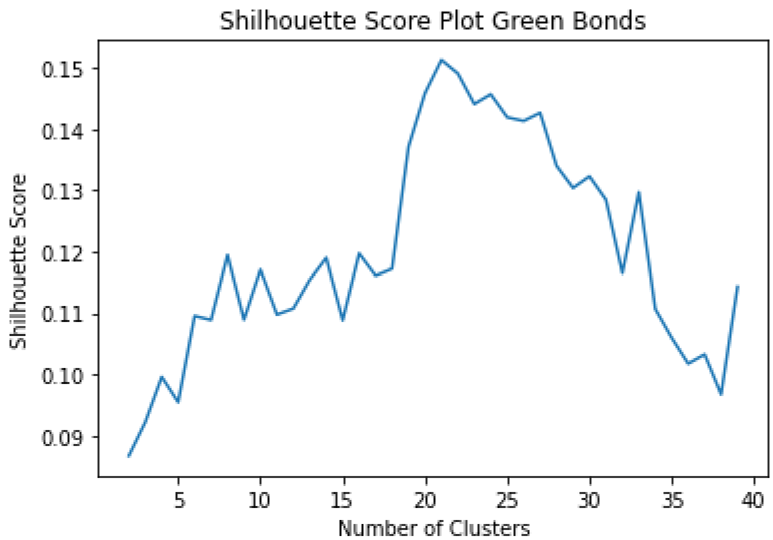
China increased its investment by 31 times in a single year while other countries decreased their investment. After a little bit of research, the team found that this was because green bonds were being issued more frequently by large banks in China. Furthermore, these banks provided guidance and policies for investing which made it easier to put money in green bonds. The following graph depicts the rapid growth of Chinese green bonds compared to global green bonds.



https://www.climatebonds.net/files/files/SotM-2016-Final-WEB-A4.pdf

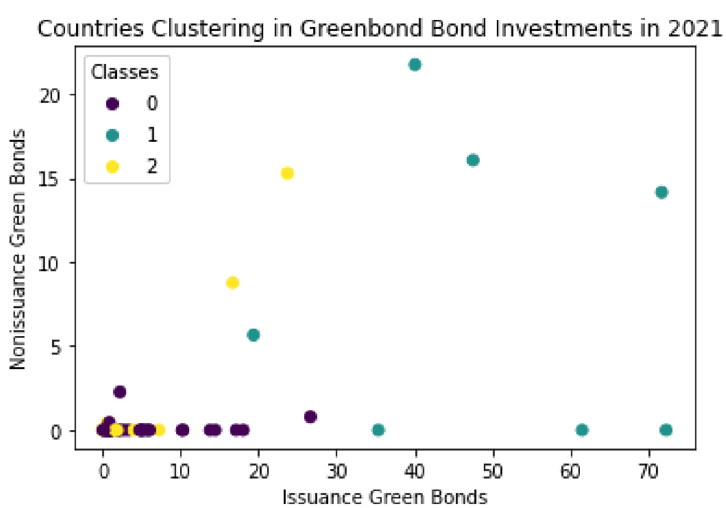
3. Clustering

The team just scratched the surface of clustering as we have only built out the framework for clustering. Like analyzing the change for a given country over a time frame, we only ran our clustering algorithm on the green bonds dataset. We ran two types of clusters. One with the BINE embeddings and one with the vector representation of policy investments by country. But before we did the clusterings we calculated the silhouette scores for each number of clusters:

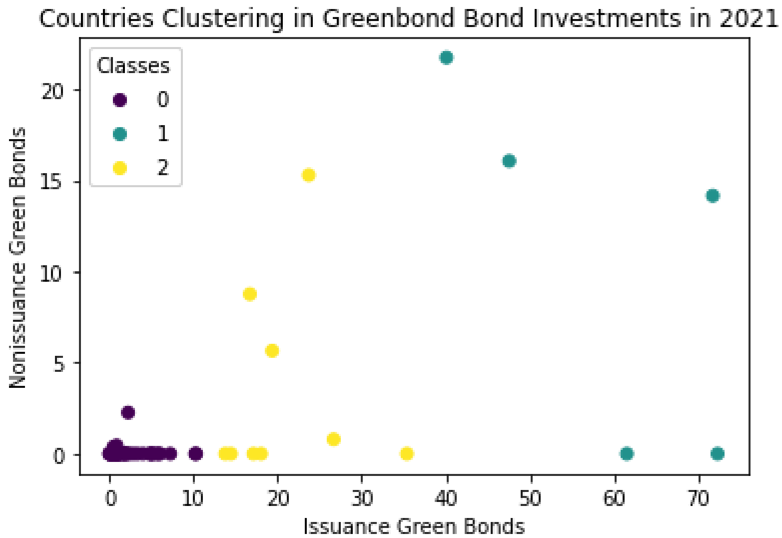


As seen in the graph the optimal amount of clusters was around 22, which means about 3-4 countries per cluster in the green bonds data set. When graphing these clusters we found it to be very difficult to actually make out what each cluster was so we decided to visually try 3 clusters first. We will test the optimal 22 clusters when we complete the clustering for all policies and years.

- **BINE**



- **Normal Vector of Green Bond Policies**

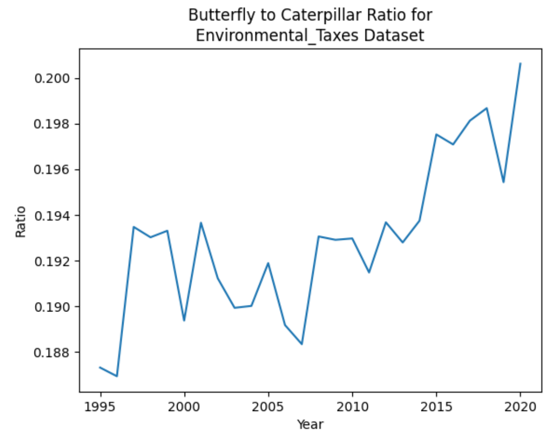


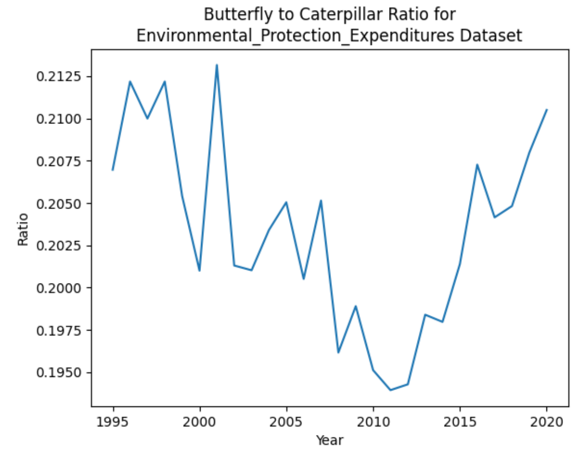
Overall both clusters seem to do a good job. The BINE clustering will be particularly useful when clustering on higher dimensional data, which is policies with many subcategories, as it will allow us to cast the original data to a lower dimensional feature space and complete clustering there instead of the original high dimensional space.

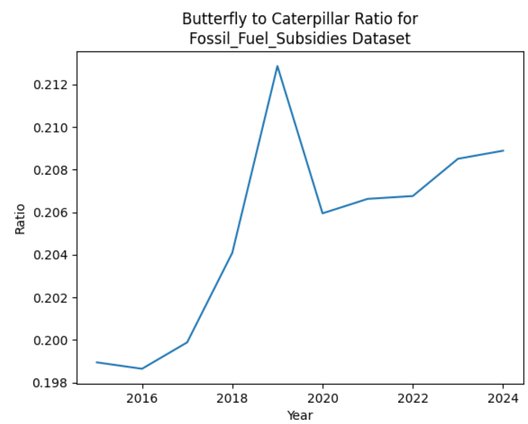
As of now we just have the clustering for the year 2021, but we hope to obtain clusters for all years and policies and see how the clusters change overtime.

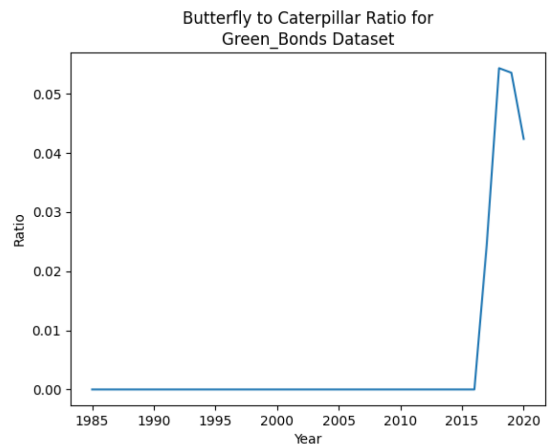
4. Temporal Analysis

We generated our temporal metrics for each of our four graphs. Below you can see our clustering coefficient for bipartite graphs over time in each of them.







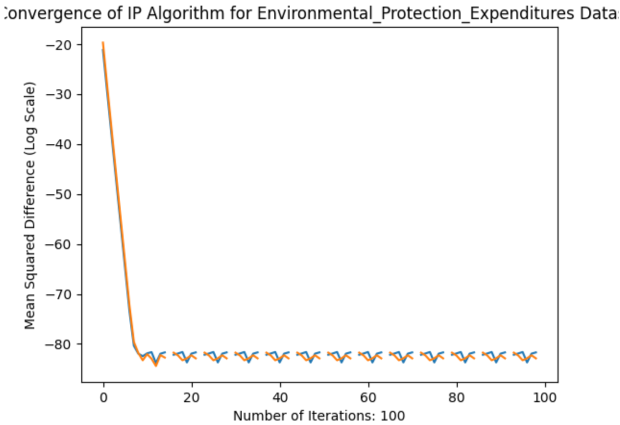


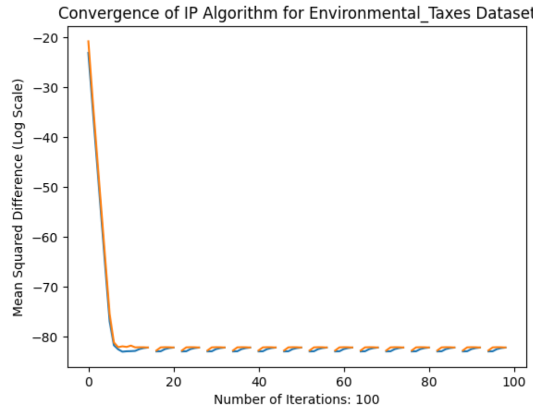
This was the most significant metric that we could find to show overall trends.

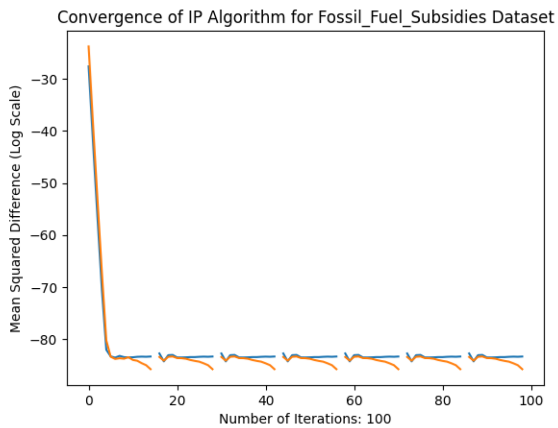
5. Influence and Passivity

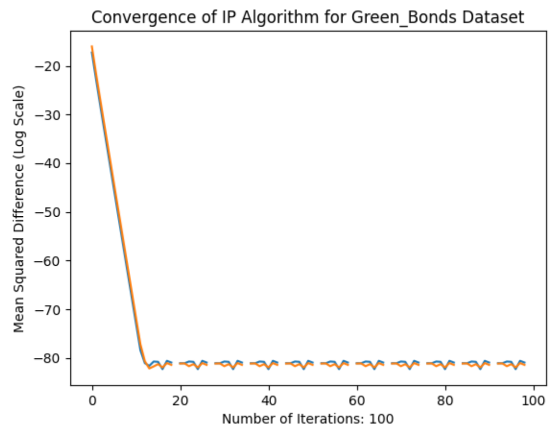
Now let's dive into the results of our IP analysis. Due to time constraint, the amount of fine-tuning of parameters was limited causing our insights to be small. However, some of the initial findings are consistent with what we would have expected, for example the participation of France as an influencer in various policies.

First, let's take a look at our IP algorithm's ability to converge on specific values for the influence and passivity of each country. As you can see in each of the graphs below, our algorithm converges in less than 20 iterations. This is quite promising as it indicates that our algorithm worked as expected by the paper.









Second, the most valuable metrics which we can pull from these influence - passivity values would be a ranking of the top 10 influencers for each policy category. The results are posted below:

Environmental Protection Expenditures -

Country	Influence
Kuwait	0.0187432
Iran, Islamic Republic of	0.01665724
Oman	0.01452929
Burundi	0.01307858
Bangladesh	0.01301114

Germany	0.01250165
Kenya	0.0124231
Estonia	0.01233171
Slovenia	0.01224845
Austria	0.01224845

Environmental Taxes -

Country	Influence
Serbia	0.01397006
Poland	0.01244476
Lithuania	0.0122698
China	0.01224986
Cape Verde	0.01215871
C√¥te d'Ivoire	0.01208435
Belgium	0.01200151
Netherlands	0.01200151
Slovenia	0.01200151
Estonia	0.01200151

Fossil Fuel Subsidies -

Country	Influence
Estonia	0.00961008
Hungary	0.00902625
Kyrgyzstan	0.00873135
France	0.00865393

Italy	0.00860223
Moldova, Republic of	0.00856131
Luxembourg	0.00850807
Mexico	0.00849748
India	0.0084194
Romania	0.00839889

Green Bonds -

Country	Influence
Norway	0.037423
Estonia	0.02856209
New Zealand	0.02828968
United Kingdom	0.02505858
Sweden	0.02390566
Korea, Republic of	0.02368861
United States	0.02325972
Panama	0.02296011
France	0.02186995
Virgin Islands, British	0.02174406

The most significant insight we could find from these charts is that France is very frequently among the top 10 for influencers. This is very promising as we know that France is a global leader in climate change efforts. Note that there are only 2 policies in our Green Bonds graph which makes it difficult for the IP algorithm (and our clustering coefficient) to accurately make distinctions.

Conclusion

Overall we were pleased with the analysis we were able to complete. With that being said we are nowhere near done on this project. As of now, we have built out the framework for the analysis, but our next step is to leverage these algorithms to draw actionable insights. In our short analysis of Influence and Passivity, change point detection, and clustering, we were able to get a small taste of what insights we can find, but we are just scratching the surface.

Next Steps

1. Apply algorithms to all datasets
2. Rewrite code to be object oriented so it can easily work on other graphs
3. Analysis the change point detection on country at a deeper level
4. Run temporal cluster analysis to see how clusters form and change over time
5. Get an answer to our original question: What country is the most influential and causes the most change across every policy.

Works Cited

Kolda, Tamara G., et al. "Measuring and Modeling Bipartite Graphs with Community Structure." *Measuring and Modeling Bipartite Graphs with Community Structure. (Technical Report) | OSTI.GOV*, 1 July 2016, https://www.osti.gov/servlets/purl/1561802.

Romero, D.M., Galuba, W., Asur, S., Huberman, B.A. (2011). Influence and Passivity in Social Media. In: Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2011. Lecture Notes in Computer Science(), vol 6913. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23808-6_2