Data Mining: Evolving Geopolitical Alignments

A Computational Analysis of UN General Assembly Voting Patterns (1946–Present)

Wail Ameur

December 28, 2025

1 Introduction

The United Nations General Assembly (UNGA) serves as a unique global forum where every member state, regardless of its economic or military stature, possesses an equal vote. While the resolutions passed by the General Assembly are generally non-binding, they represent the "collective conscience" of the international community. For decades, these votes have served as a proxy for diplomatic alliances, reflecting the underlying tensions and coalitions that define the global order.

This report uses a data-driven approach to map the evolution of these international alignments from the aftermath of World War II to the current era of renewed great power competition. By applying unsupervised machine learning techniques to nearly eight decades of voting records, we aim to visualize how the "political space" of the UN has shifted, expanded, and polarized.

1.1 Objective

The primary goal of this analysis is to identify and classify "voting blocs" within the UNGA and track their structural changes over time. By using clustering and network graphs, we seek to uncover:

- The rigidity of the bipolar world order during the Cold War.
- The impact of Decolonization and the emergence of the Non-Aligned Movement.
- The realignment of the international system following the dissolution of the Soviet Union.
- The current state of global polarization in the 21st century.

1.2 Data Source and Attribution

The foundation of this report is built upon the comprehensive UN voting dataset compiled from historical records. The data processing pipeline relies on the following attribution:

> **Copyright:** David Robinson, 2016
> **Based on:** Erik Voeten, "Data and Analyses of Voting in the UN General Assembly," *Routledge Handbook of International Organization*, edited by Bob Reinalda (2013).
> **Source Repository:** `https://github.com/Rackam06/un-voting-behaviors`

The dataset consists of three primary components: `un_votes.csv` (the raw voting records), `un_roll_calls.csv` (resolution metadata), and `un_roll_call_issues.csv` (thematic classifications). These files allow us to not only see *how* countries voted, but also the specific geopolitical issues (such as nuclear weapons, human rights, or the Palestinian conflict) that drove those decisions but we won't detail those for the clustering or divide by theme here.

1.3 Methodological Overview

To process this high-dimensional data, we employ several computational strategies focusing on unsupervised learning. First, we construct a normalized voting matrix where categorical votes are converted to numerical values $(1, 0, -1)$. The primary classification is performed using **K-Means** clustering to mathematically define the boundaries of diplomatic communities based on high-dimensional similarity. To supplement this, **Bipartite Network Projections** are used to measure the specific "centrality" and bilateral agreement between nations. Finally, **Principal Component Analysis (PCA)** is utilized strictly as a visualization tool to project the multi-dimensional cluster results into a 2D plane for human interpretation.

The project can be accessed here:
GitHub Repository: `https://github.com/dgrtwo/unvotes`

2 Methodology

To transform raw diplomatic records into a longitudinal study of international relations, we employed a multi-stage computational pipeline. This section details the data preprocessing, the mathematical transformation of voting records, and the machine learning algorithms used for clustering and network analysis.

2.1 Data Normalization and Matrix Construction

The primary challenge in analyzing UN General Assembly (UNGA) data is the categorical nature of voting. Each record in the `un_votes.csv` dataset, which was converted from .rda, represents a specific interaction between a member state and a resolution. As implemented in our `build_matrix.py` script, we transformed these categorical entries into a numerical format suitable for linear algebra operations.

The voting behavior was mapped using a tri-modal scale:

- **Yes (1):** Direct support for the resolution.
- **No (-1):** Direct opposition to the resolution.
- **Abstain/Absent (0):** Neutrality or lack of participation.

While some political analyses treat "Abstain" and "Absent" differently, for the purposes of structural alignment, both were coded as 0 to represent a lack of positive or negative alignment with the motion. Vetos were represented as -1. We then constructed a *Voting Matrix* A, where rows i represent countries and columns j represent unique Resolution IDs (`rcid`).

2.2 Graphical Mapping: Principal Component Analysis (PCA)

Because the voting matrix A contains thousands of dimensions (one for each resolution), it is impossible to visualize the cluster distributions directly. To address this, we apply PCA as a secondary visualization step.

It is important to note that PCA is used here solely to provide a two-dimensional "map" of the data points. While the axes (Principal Components) represent the directions of maximum variance, the analytical grouping of countries is determined by the K-Means algorithm in the original high-dimensional space. The PCA plots allow us to observe the relative distances between the clusters identified by the machine learning model.

In the context of the UNGA:

- **PC1** typically captures the most significant global cleavage (such as the East-West divide during the Cold War or the North-South development divide).

- **PC2** often captures secondary tensions, such as regional conflicts or specific thematic disagreements (such as Middle Eastern issues).

2.3 Unsupervised Clustering: K-Means

To move beyond visual interpretation and achieve mathematical classification, we used the K-Means clustering algorithm, as seen in `cluster_kmeans.py`.

The algorithm partitions n countries into k clusters, where each country belongs to the cluster with the nearest mean. To ensure countries with different participation rates were comparable, we normalized the feature vectors to unit length before clustering:

$$\hat{x}_i = \frac{x_i}{\|x_i\|}$$

This prevents countries that vote frequently from being mathematically distanced from allies who might be absent more often. The optimal number of clusters (k) was determined using Silhouette Scores, which measure how similar an object is to its own cluster compared to other clusters.

The Silhouette Score $s(i)$ for an individual data point i is calculated as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Where:

- $a(i)$ is the **mean intra-cluster distance**: the average distance between point i and all other data points within the same cluster.
- $b(i)$ is the **mean nearest-cluster distance**: the average distance between point i and all points in the closest neighboring cluster of which i is not a member.

The value of $s(i)$ ranges from -1 to $+1$. A score near $+1$ indicates that the country is perfectly assigned to its cluster, while a score near 0 suggests the country lies on the boundary between two clusters. By calculating the mean $s(i)$ for various values of k, we selected the cluster count that maximized global cohesion.

2.4 Bipartite Network Projection

Finally, to understand the "strength" of international ties, we used the `bipartite_projection.py` script. We modeled the UNGA as a bipartite graph $B = (U, V, E)$, where U is the set of countries and V is the set of resolutions. An edge exists if a country voted "Yes" on a resolution.

We then projected this onto a weighted unipartite graph G, where edges connect countries u_1 and u_2. The weight of the edge $w(u_1, u_2)$ represents the number of shared "Yes" votes. This allows us to use community detection algorithms (specifically Greedy Modularity Maximization) to find tightly-close diplomatic groups that transcend simple geography.

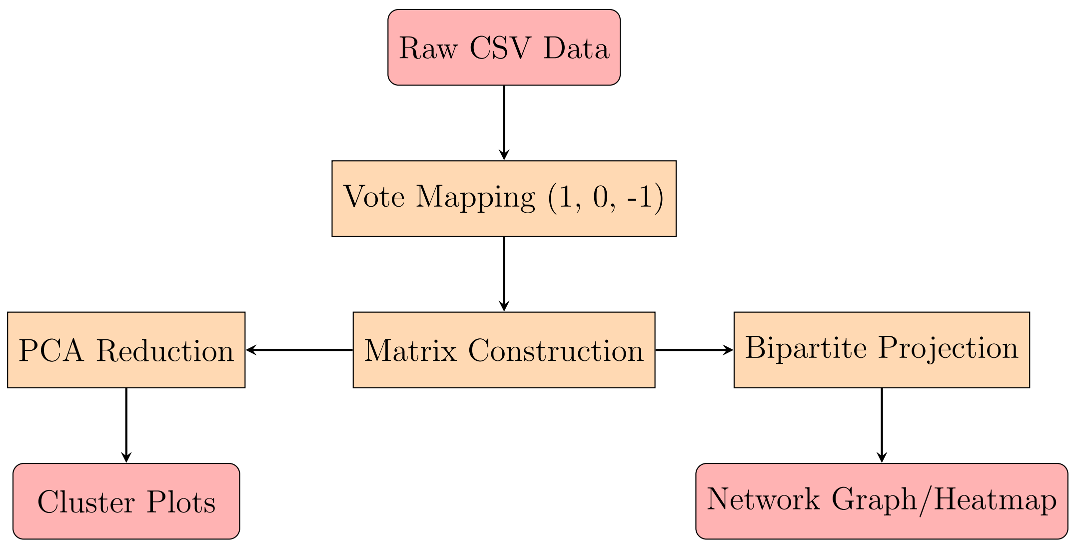

Figure 1: Methodological workflow: From voting records to multidimensional analysis.

3 The Cold War Era: Bipolarity and Decolonization (1946–1989)

The first four decades of the United Nations were defined by the overarching shadow of the Cold War. Our computational analysis of this period reveals a transition from a strictly bipolar world to a more complex, multi-polar landscape driven by the end of European empires.

3.1 The Early Bipolar Order (1946–1959)

In the immediate aftermath of World War II, the UNGA was characterized by a rigid ideological divide. The PCA results for this era typically show two clusters at opposite ends of the primary axis (PC1). We could divide the 2 clusters as Western against Communist and Third World Countries.

During this period, voting was highly predictable, with high modularity scores in our network analysis.

Figure 2: K-Means Clustering and PCA Projection (1950–1959). Note the clear separation between the Western and Soviet spheres.

3.2 The Impact of Decolonization (1960–1979)

The 1960s and 1970s marked a demographic explosion within the UN. As decolonization accelerated across Africa and Asia, the number of member states grew rapidly. Our clustering analysis for these decades shows the emergence of a third, massive cluster: the Non-Aligned Movement (NAM). Meaning that the NAM bloc (even though part of the second cluster) could be a cluster in itself since it's relatively long distance on the PC1 axis. Here it's still k=2 because (and it will be the same aftermath as well) the communist bloc isn't numerous enough to form a cluster.

The PCA projection for the 1970s illustrates a shift where PC2 begins to capture "North-South" issues, specifically economic development, sovereignty over natural resources, and the Palestinian conflict. While the Soviet and Western poles remained, the "central" cluster of newly independent nations often voted as a cohesive bloc on issues of anti-colonialism, often aligning with the Soviet bloc against the former colonial powers of the West.

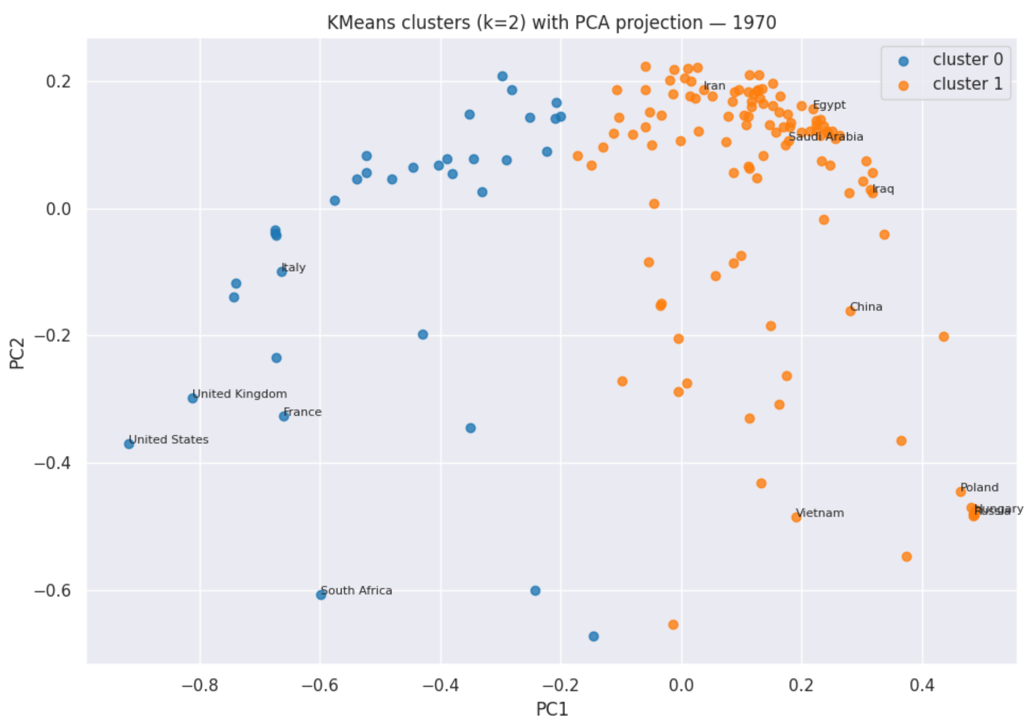

Figure 3: Geopolitical Realignment (1970–1979). The rise of the Non-Aligned Movement creates a central bridge between the two ideological poles.

3.3 The Late Cold War and Proxy Tensions (1980–1989)

By the 1980s, voting patterns reflected a "Second Cold War" characterized by renewed tensions. However, the data also shows a "fragmentation" within the Global South even though they are still close. Our `bipartite_projection.py` script identifies that during this decade, sub-communities began to form within the Non-Aligned cluster, often based on regional interests in Latin America and the Middle East.

The "Yes" vote agreement between the United States and the rest of the Assembly reached a historical low during this decade, as the U.S. frequently found itself in the minority on resolutions related to nuclear disarmament and the South African apartheid regime.

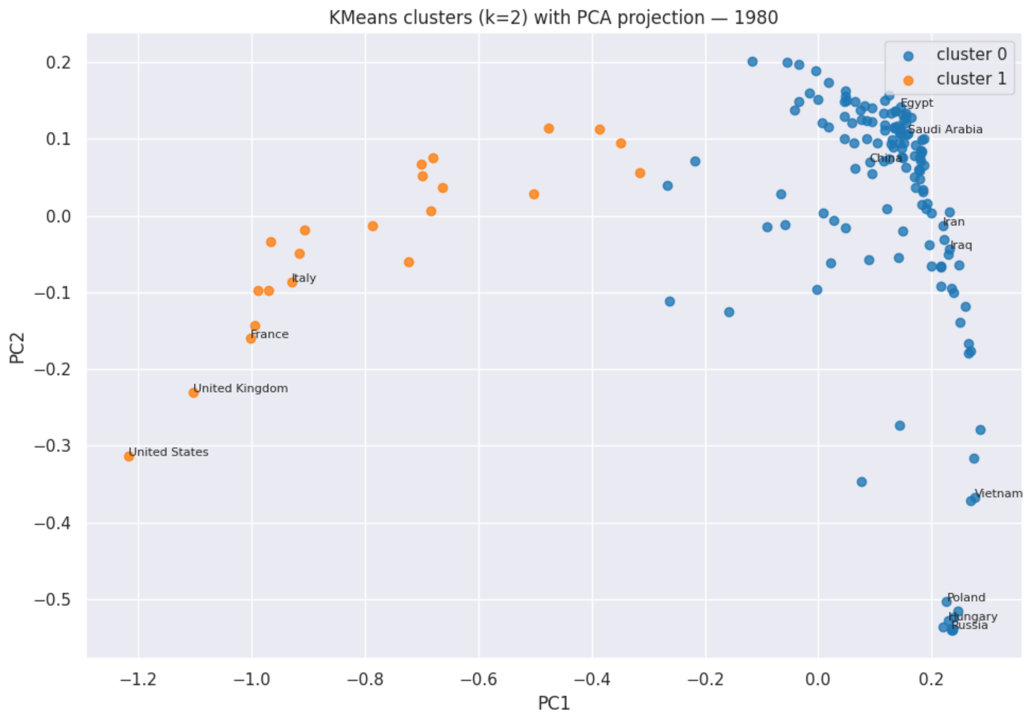

Figure 4: Clustering results for the 1980s, showing a highly isolated Western pole on specific thematic issues.

4 The Post-Cold War Transition (1990–2005)

The dissolution of the Soviet Union in 1991 fundamentally altered the topology of the United Nations. The rigid bipolarity of the previous four decades collapsed, replaced by what political scientists often term the "Unipolar Moment," where the United States and its allies exerted significant influence over the Assembly's agenda.

4.1 The 1990s: Eastern European Realignment

Our analysis for the 1990s reveals a dramatic migration of countries across the political space. Nations that were previously core members of the Soviet-aligned cluster, such as Poland, Hungary, and the Baltic states, shifted rapidly toward the Western cluster. This is represented mathematically by a relative reduction in the Euclidean distance between these nations and the "Western" pole in our K-Means model. We can also note that Russia seems now perfectly centered as USSR collapsed and Iraq is becoming more of a pariah.

During this decade, the clusters identified by `cluster_kmeans.py` become less distinct. The "ideological distance" between the former East and West narrowed as many resolutions achieved broad consensus, particularly those regarding the transition to market economies and international peacekeeping in the Balkans.

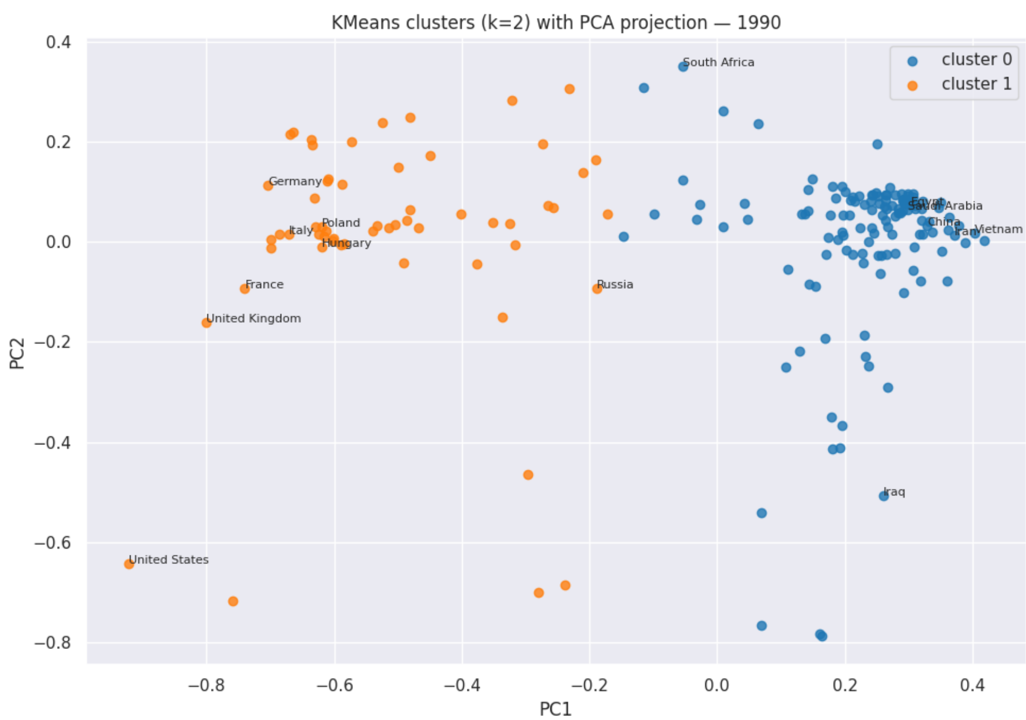

Figure 5: The Unipolar Shift (1990–1999). Note the disintegration of the Soviet cluster and the eastward expansion of the Western-aligned group.

4.2 The Early 2000s: Fragmentation and the War on Terror

By the turn of the millennium, the consensus of the 1990s began to fracture. The 2000s were marked by the "War on Terror" and the 2003 invasion of Iraq, which introduced new dimensions of disagreement within the UNGA.

Our clustering results for the early 2000s show a fragmentation of the Western cluster itself. While European nations and the United States remained generally aligned on economic issues, specific resolutions regarding pre-emptive military force and human rights monitoring created visible "sub-clusters." Additionally, this period saw the re-emergence of a cohesive "Global South" bloc rejoined by Russia (see graph below), frequently voting together on issues of global trade and patent laws for essential medicines. The USA are more distant than ever with the rest of the western countries.

5 The Modern Era and Multi-Polarity (2006–2019)

The contemporary era in this dataset is defined by the re-emergence of great power competition and the rise of "swing states" in the Global South. Our analysis identifies a return to a more polarized "political space," though the current poles differ structurally from the Cold War era.

5.1 The Rise of the BRICS and Regional Blocs

Since 2006, the PCA projections have increasingly shown the formation of a counter-pole to the Western-aligned cluster. This pole is anchored by China and Russia, but it increasingly draws in emerging economies. The `bipartite_projection.py` script indicates that these nations have high "edge weights" with one another, suggesting a consistent voting alliance that challenges Western-backed resolutions on sovereignty and non-intervention.

5.2 Thematic Divergence: Human Rights vs. Development

Using the `un_roll_call_issues.csv` data, we can see that polarization is not uniform. On issues of "Economic Development," there is often a broad consensus. However, on "Human Rights" and "Colonialism-related" issues, the Assembly remains deeply split. The Western cluster often votes "Yes" on specific human rights monitors, while the majority of the Assembly (the "Global South" cluster) votes "No" or abstains, citing concerns over national sovereignty.

5.3 Conclusion of the Longitudinal Analysis

As our data reaches 2019, we observe that the "middle ground" is shrinking. The modularity scores from our network analysis have increased over the last decade, suggesting that countries are increasingly retreating into stable voting blocs. While the 1990s suggested a move toward a "Global Village," the 2010s data suggests a return to a "Bilateral" world order.

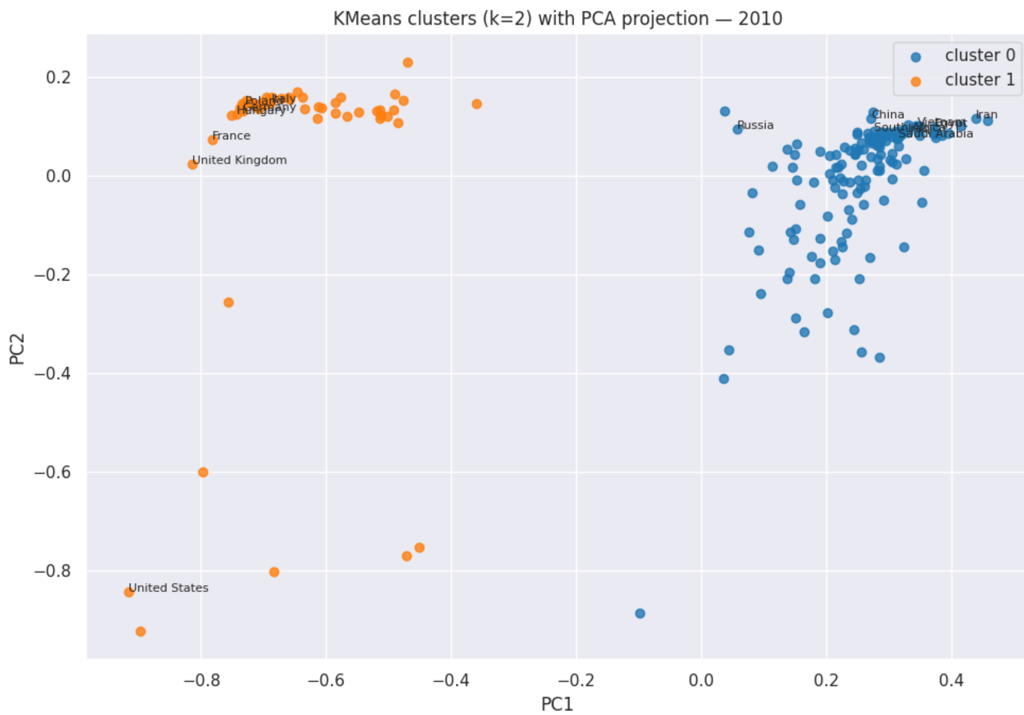

Figure 6: K-Means Clustering (2010–2019). The plot illustrates a two-bloc system: a Western pole and a China-Russia pole plus Global South group.

6 Network Topology and Alliance Analysis

While the clustering analysis in previous sections provided a snapshot of global groupings, this section employs graph theory to quantify the precise strength of bilateral relationships. By projecting the bipartite relationship between countries and resolutions into a weighted unipartite network, we can move beyond simple classification to measure the "diplomatic density" of the international system.

6.1 Quantifying Diplomatic Alignment via Shared Votes

The updated methodology in `bipartite_projection.py` focuses on the frequency of shared "Yes" votes between a selection of "Important Countries." These nations represent a cross-section of global power, including the NATO members, regional leaders in the Middle East, and former Eastern Bloc states.

The resulting heatmap (Figure 7) provides a high-resolution view of diplomatic agreement. In this matrix, the cell value $w(i, j)$ represents the total number of resolutions where both country i and country j cast an affirmative vote. This metric is a powerful indicator of shared interests, as it ignores passive agreement (abstentions) and focuses on active alignment.

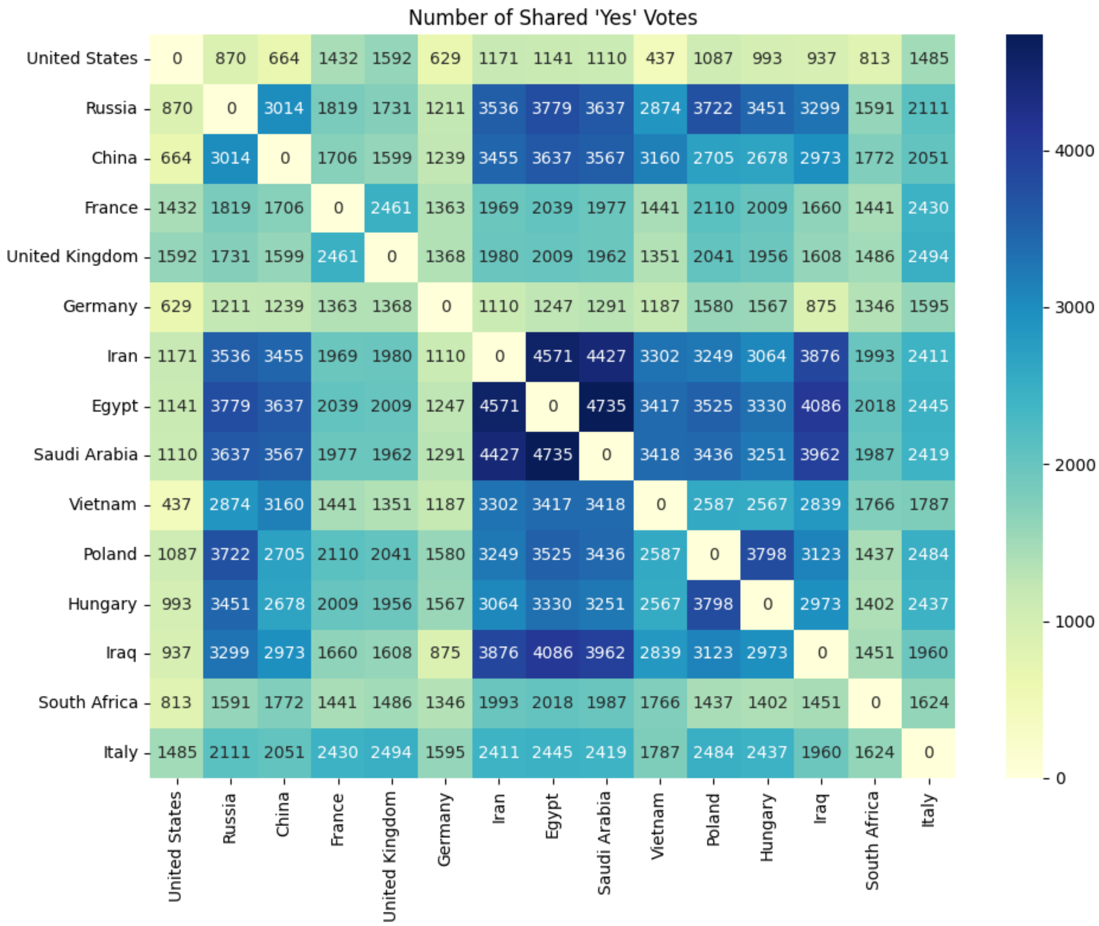

Figure 7: Heatmap of Shared "Yes" Votes among 15 Significant Powers. Warmer colors indicate a higher frequency of voting agreement, while cooler colors highlight diplomatic isolation or opposition.

6.2 Core-Periphery Dynamics and the "Isolation of the West"

Our network projection reveals a significant core-periphery structure. A striking finding in the heatmap and the associated `important_countries_matrix.csv` is the relative isolation of the United States and the United Kingdom compared to the "Global South" consensus.

For instance, countries like **Egypt**, **Saudi Arabia**, and **Iran** often display extremely high edge weights with one another (often exceeding 3,000 shared votes), forming a dense "diplomatic core." In contrast, the US frequently occupy the periphery of this network, with significantly lower shared vote counts relative to the Assembly majority. This suggests that while the US maintains a cohesive cluster of allies (as seen in Section 4), it is frequently at odds with the broader "Global South" consensus on the majority of UNGA resolutions.

6.3 Top Allies and Community Detection

The `important_countries_top_allies.csv` output allows us to identify the heart of each power's diplomatic sphere. By calculating the top 10 allies for each important country, we can observe the following patterns:

- **Legacy Alliances:** Saudi Arabia and Egypt remain each other's primary allies, reflecting a high level of economic interests and shared security priorities.
- **The Russia-China Axis:** These two powers appear as new top allies, especially since recent years, indicating a successful consolidation of a counter-Western voting bloc.
- **Transitionary Powers:** Countries like Poland and Hungary show a mix of Western and regional allies, reflecting their unique position as former Eastern Bloc nations now integrated into the European Union.

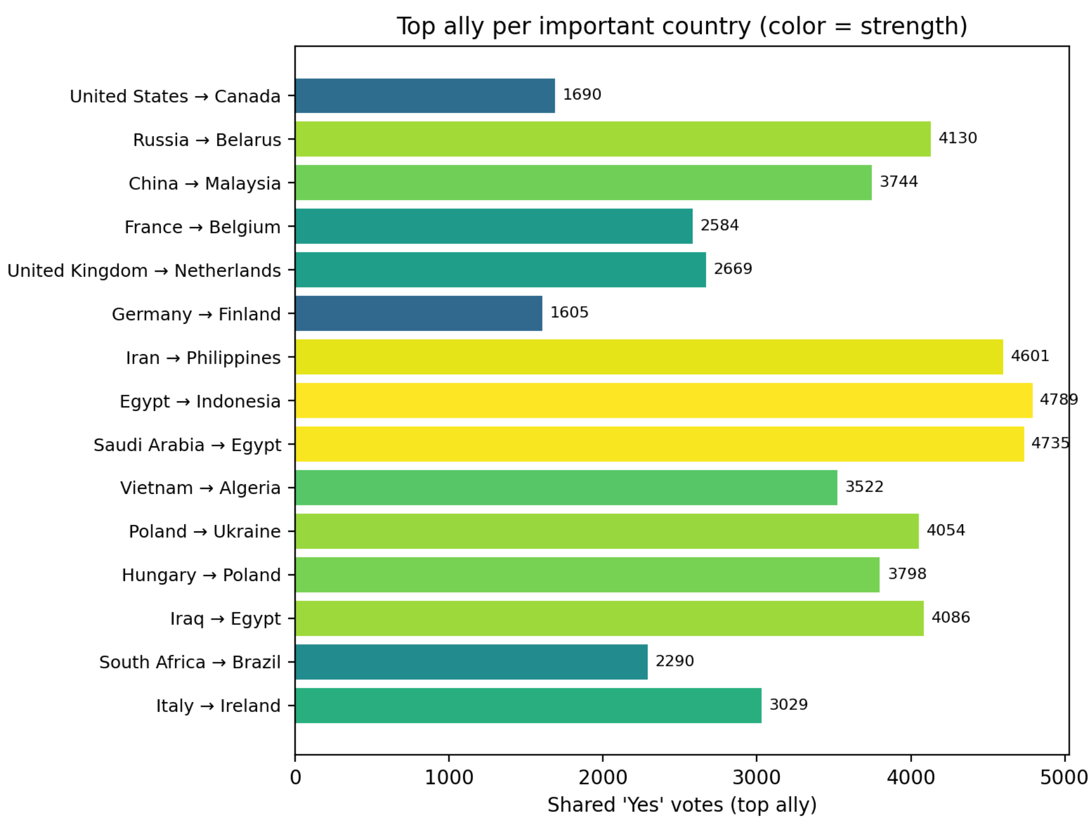

Figure 8: Projection of important countries and their top allies at the UN General Assembly.

6.4 Visualization of the Weighted Projection

Finally, the network visualization (Figure 9) uses a Spring Layout algorithm, where the distance between nodes is inversely proportional to their voting agreement.

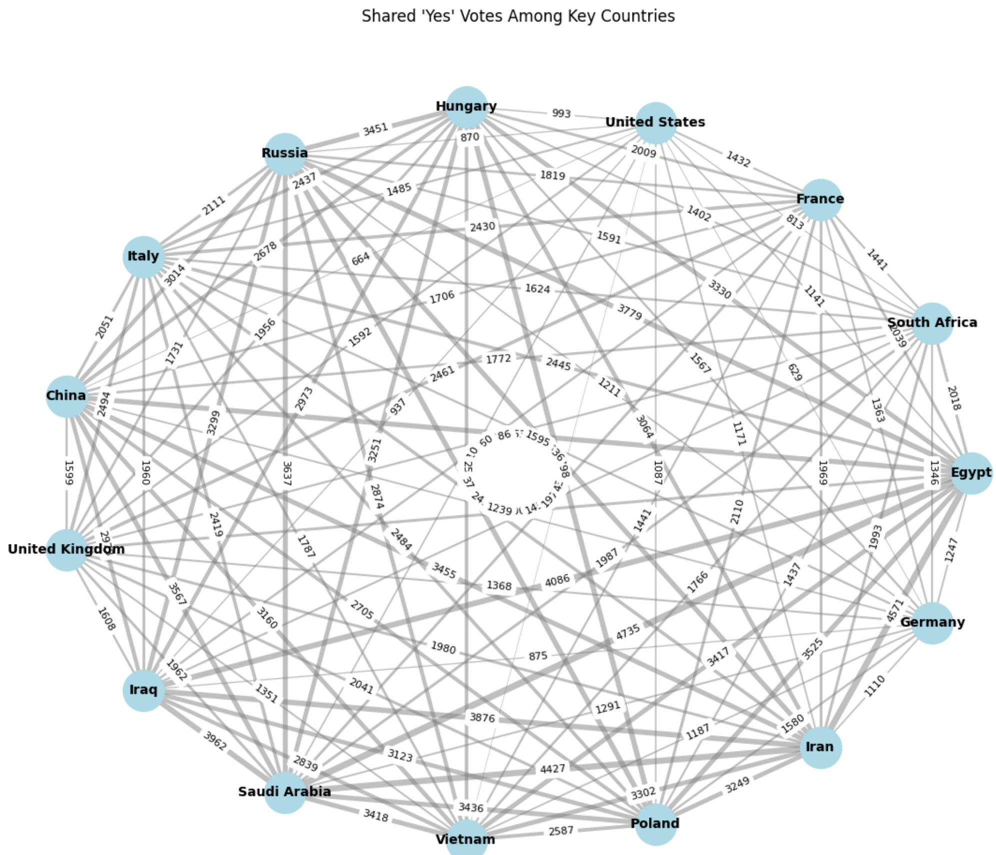

Figure 9: Weighted Network Projection of important countries from the UN General Assembly.

This visualization clearly demonstrates that the "International Community" is not a monolith but a collection of overlapping spheres. The dense cluster at the center represents the high-agreement majority of the Global South, while the smaller, more dispersed satellites represent the Western bloc and isolated states.

7 Conclusion

This computational analysis of nearly 8,000 UN General Assembly resolutions provides a longitudinal map of the shifting tides of global diplomacy. By applying Bipartite Projection and K-Means clustering to the voting matrix, we have moved beyond qualitative historical accounts to provide a quantitative verification of geopolitical structural changes.

7.1 Summary of Findings

Our research highlights three distinct eras of international relations. The **Cold War era** was defined by a mathematical bipolarity, where the variance in voting was almost entirely captured by a single axis of East-West tension. The **Post-Cold War transition** (1990–2005) saw a temporary collapse of this binary, characterized by a "unipolar" clustering around Western norms. Finally, the **Modern era** (2006–2019) has seen the emergence of a multi-polar network that tends to return to bipalor. As evidenced by our bipartite network projection and shared-vote heatmaps, while a Western core remains, a secondary pole led by China and Russia has solidified, drawing a significant portion of the Global South into its orbit.

7.2 Methodological Limitations

While the `bipartite_projection.py` and clustering scripts offer deep insights, they are subject to certain limitations:

- **Vote Weighting:** Our model treats every resolution equally. In reality, a vote on a procedural amendment is less significant than a vote on nuclear non-proliferation.
- **Strategic Abstention:** As noted in the 2010s analysis, many nations use "Abstain" or "Absent" as a deliberate diplomatic tool. By coding these as 0, we treat them as neutrality, which may occasionally mask a "soft" opposition.
- **The "Empty" Chair:** Smaller nations with limited resources often fail to vote due to lack of representation rather than political choice, which can skew cluster density.

Plus some results seems at odd with non-mathematical geopolitical analysis such as the proximity between Saudi Arabia and Iran knowing their legendary feud since the late 70s in the Middle-East. This can be explained here by the fact that they often vote on same resolutions even when they don't think them, this is plain example of geopolitical hypocrisy.

7.3 Final Remarks

The United Nations General Assembly remains a vital data source for understanding the global order. As the world enters a period of renewed fragmentation, computational tools like K-Means can be useful for diplomats and scholars to identify emerging ”bridge” nations and predict the formation of new alliances. Future work could improve this model by a lot of means to analyze the text of resolutions, allowing the model to weight ”Important Votes” more heavily than what we have done here.