2013 IEEE 13th International Conference on Data Mining Workshops

Analysis of UN Voting Patterns via Diffusion Geometry and Thematic Clustering

Minh-Tam Le[1] ,Mathew Lawlor[3]
1. Dept. of Computer Science
Yale University New Haven, CT 06520-8285
Email: minhtam.le@yale.edu

Bruce M. Russett[2]
2. Dept. of Political Science
Yale University
New Haven, CT 06520-8301

John Sweeney[1], Steven W. Zucker[1,3]
3. Program in Applied Mathematics
Yale University
New Haven, CT 06520-8285

***Abstract*—We apply a range of data mining techniques to analyze voting patterns in the United Nations. We begin with non-linear dimensionality reduction, showing that diffusion geometry reveals an historically relevant organization of countries based on their UN voting patterns. Key historical events can be "read out" from these embeddings, such as de Gaulle's influence on France and the breakup of the Soviet Union. These events are not apparent in other (e.g., PCA) embeddings. We then switch to an organization of resolutions, revealing dominant themes during different political epochs. Formally themes are introduced as summaries (eigenfunctions) within a modified hierarcical clustering algorithm.**

I. Introduction

We seek to discover the implicit structures underlying the relationship among social actors in social datasets. In this work, we mainly focus on the network of countries in IR (International Relation) datasets such as trade, IGO (Intergovernmental Organization) membership and the UN (United Nations) GA (General Assembly) voting records. Understanding the structures of these networks may help shed light on theoretical questions in IR such as (i) Do IGOs have any influence on armed conflicts? (ii) What do votes in the UN GA reveal? (iii) Does trade help reduce conflicts?

Our goal is twofold: (i) to motivate, develop, and illustrate a diffusion-based approach [1] to embedding high-dimensional UN voting data; and (ii) to cluster the resolutions "driving" these embeddings to reveal thematic threads running across time. In the end we show that there is enormous richness in these data, sufficient to reveal major historical events directly. That is to say, one could learn much about history using only these data.

To start, we review diffusion geometry, a non-linear dimensionality reduction technique based on the concept of diffusion distance, which considers not only direct dyadic connections between social actors, but also all indirect paths of diffusion through intermediate neighbors. This is important in political science because influence accumulates in a manner than may not be revealed by linear techniques.

We apply this technique to socio-political databases, such as IGO membership [2] and UN voting [3]. While these have received significant attention from scholars of international relations [4]–[8] we do not believe that they have previously been analyzed by techniques such as ours.

Research supported by Air Force Office of Scientific Research.

Secondly, we develop a hierachical clustering algorithm to identify themes running through the voting patterns. This is, in effect, the complement to the above, because it reveals structure among resolutions rather than countries. Taken together both techniques reveal how much structure is implicit within UN General Assembly voting patterns.

II. Diffusion distance

We approach the dimensionality reduction problem by means of the social network model: Consider $G(V, W)$ as a network whose vertices $i \in V$ are the countries and kernel function W_{ij} represents the similarity between countries i and j (e.g. $W_{ij} = e^{-\frac{r_{ij}^2}{10^8}}$ where r_{ij} is geographical distance between capital cities of countries i and j).

Social phenomena and trade, unlike geography, follow a different distance measure. Goods and social capital *diffuse* from one place to another, perhaps through an intermediate country. Thus nearby countries matter more than distant ones. Since classical techniques preserve all pairwise Euclidean distances between data points, we argue that not all distances should be preserved uniformly. Instead, *only short distances shoud be maintained, perhaps even attenuated to preserve the local structure, while long distances can be discounted*. The argument is illustrated in Fig. 1. In political terms, we see a polarization in which two camps (B, C) closely communicate, but (A, B) barely interact with each other except through intermediary contacts located in the neck. An embedding that highlights this polarization should tighten the clusters' girth (thus *attenuating short distances*) and stretch the neck's length (thus *loosening long distances* and separating the two clusters from each other). Those are the characteristics of *diffusion* and they differ from gravitational potentials [9].

Think of a substance (e.g. money, population, or political influence) diffusing from a source point out to neighboring points in amounts proportional to the neighbors' similarity to the source. The substance continues to diffuse to the neighbors of those neighbors, etc. Assuming a fixed amount of substance in the network, we can define $p_t(k|i)$ as the density of substance, originating from source point i, at point k at time t. Thus $p_t(k|i)$ would be high if there are many paths of length $\leq t$ connecting i to k, and low otherwise. If we take point $i = B$ on the right of Fig. 1 as the source, after t time steps, most of the substance originated from B should end up at points like $k = C$ on the right cluster, and only a small fraction at points like $k = A$ on the left; there are significantly more

978-0-7695-5109-8/13 $31.00 © 2013 IEEE
DOI 10.1109/ICDMW.2013.81

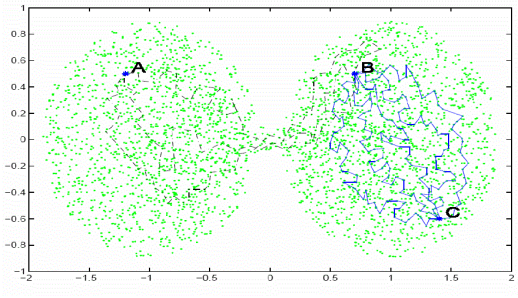


Fig. 1: *Two tight clusters separated by a narrow path. It is obvious that there are many paths between any pair of nodes from the same cluster (B and C), while there are significantly fewer paths between any pair of nodes from different clusters (A and B).*

paths from B to C than to A. The intuitive *diffusion distance* [1] between points i and j is a weighted difference between the two probability density functions:

$$\begin{aligned} D_t^2(i,j) &= \|p_t(k|i) - p_t(k|j)\|_\omega^2 \\ &= \sum_k (p_t(k|i) - p_t(k|j))^2 \omega(k) \end{aligned} \tag{1}$$

where $\omega(\bullet)$ is the weight function that normalizes the distance according to the density estimate of each vertex.

International trade can also be viewed as a diffusion process in which money diffuses from country to country. The polarization in Fig. 1 can be described in terms of trade during the Cold War. Assuming the trade pattern stays constant, the money will diffuse out to the two sources' trading partners, like 'bumps' of heat diffusing through a graph. Thus $p_t(\bullet|USA)$ will be high in the West, and low in the East, while $p_t(\bullet|USSR)$ behaves in the opposite direction. The function $p_t(\bullet|USA)$ provides a notion of "trading sphere" of the USA. Therefore, the diffusion distance between the USA and the USSR can be defined as the difference between their corresponding spheres $p_t(\bullet|USA)$ and $p_t(\bullet|USSR)$, as described by Eq. 1.

III. Random walk

To compute the diffusion distance $D^t(i,j)$, which takes into account *all paths* (of length t) between i and j, begin by considering a random walk for a traveler in country network $G(V,W)$. The transition probability is $M = D^{-1}W$, where D is a diagonal matrix $D_{ii} = d_i = \sum_j W_{ij}$, the degree matrix. The matrix $\widetilde{M} = D^{1/2}MD^{-1/2} = D^{-1/2}WD^{-1/2}$ is symmetric and has the same spectrum as M. If $p_t(i)$ denotes the probability the traveler appears in country i at time t, then

$$p_{t+1}^T = p_t^T M = p_t^T D^{-1} W \tag{2}$$

Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of $\widetilde{M}$ and $\{\upsilon_k\}$ their corresponding orthonormal eigenvectors:

$$\widetilde{M} = \Upsilon \Lambda \Upsilon^T \tag{3}$$

where Λ is the diagonal matrix with $\{\lambda_k\}$ on its diagonal, and Υ is a matrix whose columns are the corresponding eigenvectors $\{\upsilon_k\}$. Therefore

$$M = D^{-1/2}\widetilde{M}D^{1/2} = D^{-1/2}\Upsilon\Lambda\Upsilon^T D^{1/2} = \Psi\Lambda\Phi^T \tag{4}$$

where $\phi_k = D^{1/2}\upsilon_k$ and $\psi_k = D^{-1/2}\upsilon_k$, which implies that $\{\phi_k\}$ and $\{\psi_k\}$ are the left and right eigenvectors of M corresponding to eigenvalues $\{\lambda_k\}$. Since $\{\upsilon_k\}$ are orthonormal vectors, ϕ_i and ψ_j are bi-orthonormal: $\phi_i^T\psi_j = \delta_{ij}$. We can also verify that

$$\begin{aligned} \widetilde{M}d^{1/2} &= D^{-1/2}WD^{-1/2}d^{1/2} \\ &= D^{-1/2}W\mathbb{1} = D^{-1/2}d = d^{1/2} \end{aligned} \tag{5}$$

Therefore $d^{1/2}$ is an eigenvector of $\widetilde{M}$ with eigenvalue 1, and hence $\forall k\ |\lambda_k| \leq 1$ [10]. Thus $\lambda_1 = 1$. In fact, if G is connected (so that M represents an irreducible and aperiodic Markov chain) then $\forall k > 1\ |\lambda_k| < 1 = \lambda_1$. We also have $\upsilon_1 = \frac{d^{1/2}}{\|d^{1/2}\|}$, which leads to $\phi_1 = \frac{d}{\|d^{1/2}\|}$ and $\psi_1 = \frac{\mathbf{1}}{\|d^{1/2}\|}$. That means ψ_1 is a constant vector, while $\phi_1(i) = \frac{d_i}{\sqrt{\sum_k d_k}}$.

Let $p_t(j|i)$ be the probability that the traveler starts walking from country i and appears in country j at time t. It follows from Eq. 2:

$$p_t(j|i) = e_i^T M^t = e_i^T \Psi\Lambda^t\Phi^T = \sum_k \psi_k(i)\lambda_k^t\phi_k(j) \tag{6}$$

where e_i is a vector whose entry $e_i(k) = \delta_{ik}$. Therefore, if G is connected, the following limit holds, regardless of the initial starting point:

$$lim_{t\to\infty} p_t(j|i) = \psi_1(i)\phi_1(j) = \frac{d_j}{\|d^{1/2}\|^2} = \frac{d_j}{\sum_k d_k} \tag{7}$$

The first eigenvector ϕ_1 serves as the stationary distribution of the random walk M. It can also be considered a density estimate, which tells us of how frequently our walker passes by a particular country. In social network terminology, it is the centrality vector.

IV. Diffusion Maps

For each country i, suppose the diffusion process starts with an initial distribution $p_0(j|i) = \delta_{ij}$. After t steps, this distribution diffuses out to the neighborhood of i with the landscape described by $p_t(j|i)$. The walker is more likely to end up in states close to i than those far away. The diffusion distance $D_t^2(i,j)$ can be measured by Eq. 1, with the weight function $\omega(k) = \frac{1}{d_k}$ that normalizes the distance by the centrality measure of each node. $D_t^2(i,j)$ can be seen as the weighted difference between the two distributions of concentrations after t steps of two random walks starting from nodes i and j.

Define the diffusion map Ψ_t between the original data space onto the first κ left eigenvectors of M:

$$\Psi_t(i) = (\lambda_1^t\psi_1(i), \lambda_2^t\psi_2(i), \ldots, \lambda_\kappa^t\psi_\kappa(i)) \tag{8}$$

The diffusion distance in Eq. 1 is equal to Euclidean distance in the diffusion map space:

$$
\begin{aligned}
D_t^2(i,j) &= \sum_l^n \left(\sum_k^\kappa \lambda_k^t (\psi_k(i) - \psi_k(j)) \phi_k(l) \right)^2 \frac{1}{d_l} \\
&= \sum_{k_1,k_2}^\kappa \lambda_{k_1}^t \Big(\psi_{k_1}(i) - \psi_{k_1}(j) \Big) \lambda_{k_2}^t \Big(\psi_{k_2}(i) - \psi_{k_2}(j) \Big) \delta_{k_1 k_2} \\
&= \sum_k^\kappa \lambda_k^{2t} \Big(\psi_k(i) - \psi_k(j) \Big)^2 \\
&= \| \Psi_t(i) - \Psi_t(j) \|^2
\end{aligned}
\tag{9}
$$

Practically, only the last $(\kappa - 1)$ coordinates are to be considered because ψ_1 is constant. Importantly, since $\forall k \ |\lambda_k| < 1$, components $\lambda_k^t \psi_k(i)$ in Eq. 8 corresponding to smaller values of λ_k vanish rapidly as t increases, achieving nonlinear dimensionality reduction [11].

V. Experimental Results

We present several examples of diffusion maps applied to geopolitical databases. Dynamic visualizations are available at http://www.cs.yale.edu/homes/vision/zucker/embeddings.html.

For illustrative purposes, we define a distance-ratio function $DR(c,T) = \frac{\text{dist(c,T)}}{\text{diam(T)}}$ as the the distance between country c and a group of countries T; $DR(c,c',T) = \frac{\text{dist(c,c')}}{\text{diam(T)}}$ as the distance between countries c and c', both ratios are normalized by the diameter of group T.

A. Geographical map - A physical perspective: Fig. 2 provides an experiment with geographical embedding of national capitals [12], with the kernel $W_{ij} = e^{-\frac{r_{ij}^2}{10^8}}$. The resulting embedding approximates global positions.

B. Intergovernmental organization (IGO) membership pattern: Fig. 3 reveals how various countries are positioned, given their IGO memberships [2] in the year 2000. The diffusion maps were derived using the correlation of joint membership as the kernel function. The maps show that IGO membership pattern tends to correlate with regional geographical positions.

C. UN vote pattern - de Gaulle's France: Using the Pearson product correlation kernel, we embed the UN member nations in three dimensions, according to their votes in the UN General Assembly for various years (Fig. 4). Fig. 5 plots the ratios of embedding distance in the period 1965-2000. The plots of different distance measures show how the diffusion method amplifies the connections between highly connected actors, while increasing separation between distant parties.

France's self-isolation under de Gaulle's presidency is apparent from the diffusion maps. In 1957 (Fig. 4a), France (cyan star, upper left corner) was close to the USA, UK, Belgium, Luxembourg (blue markers). By 1959, France under de Gaulle began to withdraw from the NATO military command and completed that process in 1966. Viewing the diffusion maps as time proceeds, we see France slowly moving to the edge of the (blue) Western group in 1960 (Fig. 4b), gradually edging further away by 1963 (Fig. 4c), planting itself in a distant position from that of the West in 1967 (Fig. 4e). The distance ratio plot in Fig. 5a shows the blue line (FRN-EU) started at around 0.8, the green line (FRN-UK) reaching its peak at 9 in 1967-1968, while the red line (UK-EU) lying low initially, indicating France's isolated position from that of the Western countries (and UK) at the time. After de Gaulle left office in 1969, the blue line began to decline steeply, moving in tandem with the red line, implying a reverse course in France's foreign policy, gradually edging closer to that of the rest of West. Indeed, Fig. 4f shows France (cyan star, bottom left) moving back toward integration in NATO, its position in 1972-1973 (Fig. 4g-4h) approaching UK (blue triangle, top left) (FRN opened up from its self-isolation, allowing UK to join EC in 1973). By 1975 (Fig. 4i) France again stood close to the Western bloc. In the 80s until the end of the Cold War, the distance ratios FRN-EU and UK-EU (blue & red lines, Fig. 5a) ascended slightly, due to the absorption of new members into the EU. The green line (FRNK-UK), however, remains low throughout the 1980s, showing how close FRN and UK's policies were to each other during that period.

The diffusion maps reveal the inherently low dimensional structure among countries, in agreement with prior analysis [13], [14]. PCA fails to reveal a pattern in the movements of countries in the network, while diffusion distance uncovers the

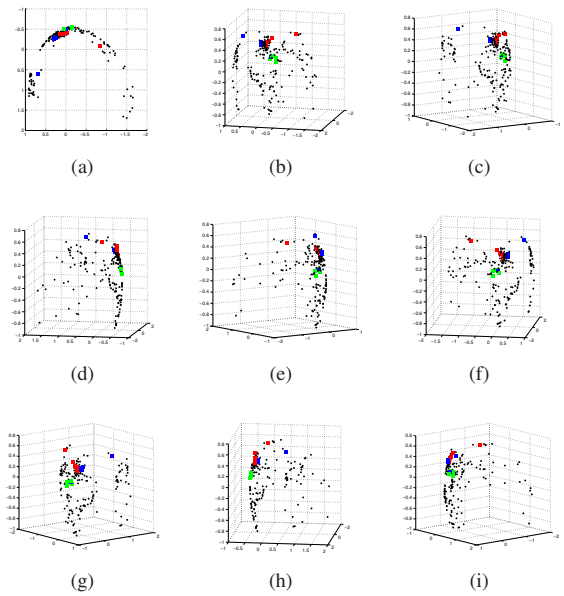
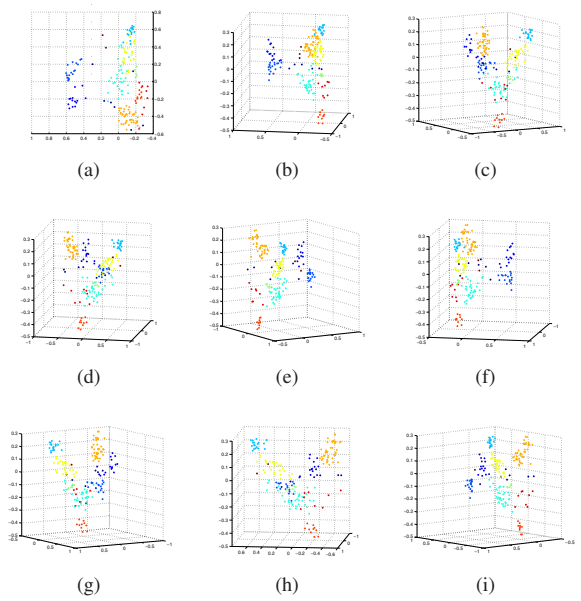
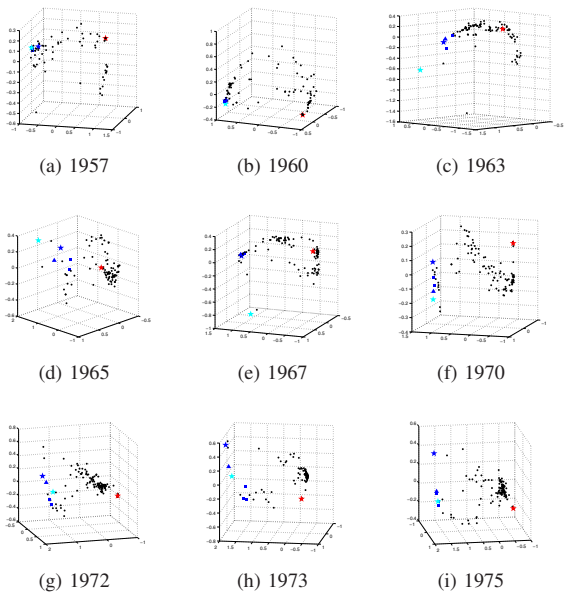


Fig. 2: *Geographical embedding of national capitals in 3-dimensional space, using the 2nd, 3rd, and 4th vectors of the diffusion map. The edge weight function is defined as* $W_{ij} = e^{-\frac{r_{ij}^2}{10^8}}$ *where* r_{ij} *is geographical distance between captitals of nations* i *and* j*. Figure (a) provides a top down view, while (b)-(i) show side views of the embedding from different angles, turning from west to east (counterclockwise). Several countries are marked with colored squares for easy identification:* ■ *(USA, UK, FRN, BEL, ISR),* ■ *(USSR/RUS, CHN, POL, HUN, BLR),* ■ *(EGY, SYR, LEB, SAU, KUW).*



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 3: *Diffusion map of countries, given their IGO membership in 2000, using the 2^{nd}, 3^{rd}, and 4^{th} vectors. (a) provides a top-down perspective, while (b)-(i) show side views from different angles, in counterclockwise rotation. The countries are manually colored according to their geographical locations, which shows again that IGO's aligning influence is mostly regional. Legend (with respect to (a)): Caribean (dark blue, upper left); Central & South American (medium blue, lower left); Western European (light blue, upper right); former Soviet states & ISR (yellow, upper right); North African (light red, middle far right); African (light orange, lower right); Middle East (dark orange, middle right); USA & CAN (dark red, middle).*

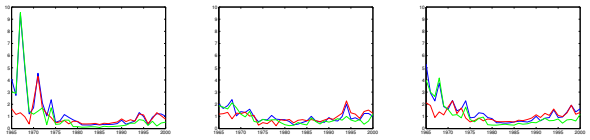


(a) 1957 (b) 1960 (c) 1963 (d) 1965 (e) 1967 (f) 1970 (g) 1972 (h) 1973 (i) 1975

Fig. 4: *De Gaulle's France: Diffusion maps of UN voting pattern 1957-1975. Several countries are marked for case study identification: ★(USA), ▲ (UK), ★ (FRN), ■ (BEL, LUX, GFR), ★ (USSR/RUS). These maps show France started out close to the Allies in 1957. Then in 1960, France, under de Gaulle's presidency, distanced itself from the West. The 70s saw France coming back toward the Western fold, once de Gaulle had left.*

same pattern as the global Hamming distance. The spectrum given by PCA decays very slowly: it requires 20-30 dimensions to describe all variances in the voting data. The diffusion method, on the other hand, requires only 5-7 dimensions to describe the voting patterns [14]. The diffusion method performs better in amplifying significant events in its distance plot (e.g. the period from 1957-1967 in which France isolated itself). However, the diffusion distance in Fig. 5 is computed from only 5 dimensions, whereas the Hamming distance is the aggregated result of votes on all UN resolutions in a particular year.

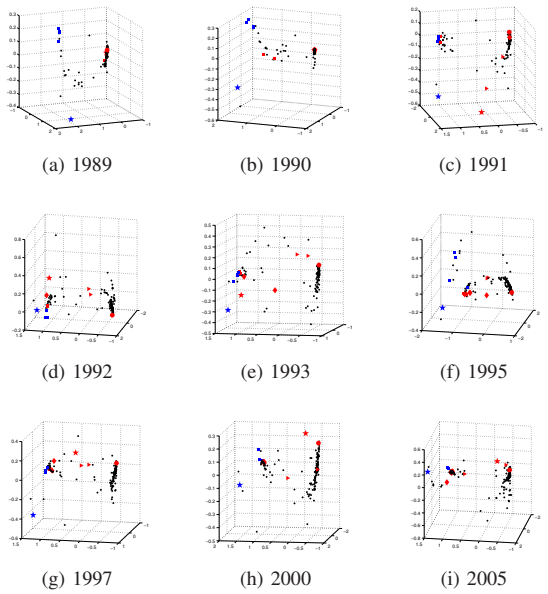
D. UN vote pattern - The collapse of the Soviet Union: Fig. 6 shows diffusion embeddings of countries according to their UN voting patterns during 1989-2005. Countries are closer to each other if they voted similarly and apart if they did not. Fig. 7 compares 3 distance metrics: (a) diffusion distance by our method (which shall be defined in more detail later in this article), (b) PCA embedded distance (Euclidean distance between data points embedded by a Principal Component Analysis projection), and (c) Hamming distance (normalized number of resolutions that countries voted differently from one another.) Each subfigure plots the ratios of embedding distances in the period 1965-2000.



(a) Diffusion distance (b) Distance of PCA embedding (c) Hamming distance

Fig. 5: *Embedding distance ratios $DR(FRN, EU*)$, $DR(UK, EU*)$, $DR(FRN, UK, EU*)$ in 1965-2000. Here EU* is defined as the states of the European Community, excluding FRN & UK. These plots show how relations between France, UK and the rest of the Western European states changed over time, with France standing far apart during the 60s, and coming back to the fold afterward.*

The 1989 diffusion map is polarized with the Western bloc (blue) on the left and the Eastern bloc (red) on the right of Fig. 6a. The distance ratio plots (Fig. 7a) show the green line (POL-EU) trailing the red line (USSR/RUS-EU) prior to 1989, indicating Poland's policy completely dominated by that of the Soviet Union. However, in 1990 (Fig. 6a), Poland and Hungary (red squares) switched to the left, followed quickly by Czecholovakia, Bulgaria, and then the three newly independent Baltic republics. Fig. 7a clearly reveals a break between the green line and the red line from 1989, showing different trends

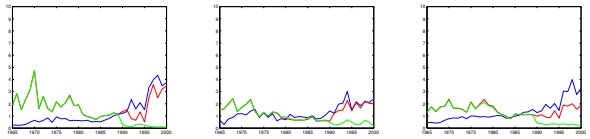


(a) 1989 (b) 1990 (c) 1991

(d) 1992 (e) 1993 (f) 1995

(g) 1997 (h) 2000 (i) 2005

Fig. 6: *The collapse of the Soviet Union: Diffusion maps of UN voting pattern 1989-2005. Several countries are marked for case study identification:* ★ *(USA),* ■ *(UK, FRN, BEL, LUX),* ★ *(USSR/RUS),* ♦ *(YUG),* ► *(UKR, BLR),* ■ *(POL, HUN),* ● *(CHN).*



(a) Diffusion distance (b) Distance of PCA embedding (c) Hamming distance

Fig. 7: *Embedding distance ratios* $DR(USA, EU)$, $DR(USSR/RUS, EU)$, $DR(POL, EU)$ *in 1965-2000. Here EU is defined as the states of the European Community. These plots show relations among the USA, USSR/RUS, Poland, and Western European states changed over time, with Poland tied to the USSR/RUS until 1989, after which it completely aligned itself with the West.*

in Poland and the USSR/RUS's policies from then on. By 1991 (Fig. 6c), the Soviet Union (red star), Belarus, and Ukraine followed suit, as they (the 2 red triangles) moved toward the center. In 1992, after the Soviet bloc fully disintegrated (Fig. 6d), its members had all migrated to the left, with Ukraine and Belarus hanging in the middle, leaving China (red circle) on the right, close to the Arabs and the third world. Figs. 6d- 6f depict Russia's effort to get close to the West, as Yeltsin vied for Western support for admission to NATO or the EU. The downward trend of the red line during 1992-1995 in Fig. 7a indicates Russia's similar lack of success seeking eventual membership in the EU. After Yeltsin's second election in 1996 and his inability to integrate his country into Western

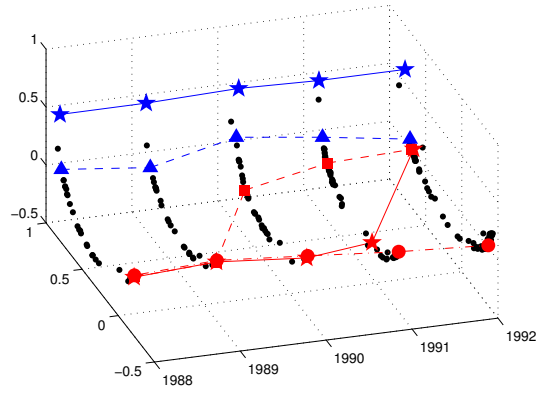


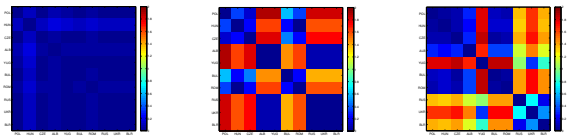
Fig. 8: *The disintegration of the Soviet Union (1988-1992): The evolution of 2-dimensional diffusion maps of nations according to their voting patterns in the UN Assembly. Each dot denotes the global position of a country in a particular year. Special markers are drawn to denote:* ★ *(USA),* ▲ *(UK),* ★ *(USSR/RUS),* ■ *(POL),* ● *(CHN). Several lines are also plotted connecting the "paths" of these countries over time. Note how USA and UK stayed relatively steady at their positions, while the paths of Communist states started to diverge since 1989. POL was the first to move out of the camp in 1990, followed by USSR/RUS, whereas CHN remained in their original position throughout the whole period.*

institutions (Fig. 6g), Russia moved to the right of the map. Fig. 7a records a sharp ascent of the red line after 1996, implying Russia's abandonment of its westward movement. A further shift eastward occurred after Putin replaced Yeltsin in 2000 (Fig. 6h), as Russia moved further to the right and close to China.

The collapse is even more evident in Fig. 8, which provides a time-evolution by stringing the 2-dimensional structures of the alignments in Fig. 6 along the time dimension. It is apparent from the figure that:

- USA and UK stood close to each other in the 2-dimensional alignment, and their distance remain relatively stable throughout the 5-year period.
- The break-up of the Soviet bloc is shown in the diverging lines of the USSR/RUS, POL and CHN. The bloc stayed intact until 1990, when POL moved away, toward the other side of the map. In 1991, the Soviet Union inched apart from CHN and the third-world countries, and then Russia moved completely out by 1992.

For further analysis, we consider the group of Communist countries in the years 1989-1991. Fig. 9 shows the diffusion distances among these countries in 1989-1991. The group was tight in 1989 and quickly disintegrated in 1990 and 1991, as the diffusion distances suddenly spiked up in these two years.



(a) 1989 (b) 1990 (c) 1991

Fig. 9: *Diffusion distances among the countries in the Soviet Bloc (POL, HUN, CZE, ALB, YUG, BUL, ROM, USSR/RUS, UKR, BLR) in 1989-1991. The colors denote distance value from low (cool, blue color) to high (hot, red color).*

	1989	1990	1991
Middle East	2	7	6
Weapon Nonproliferation	2	6	5
Anti-Apartheid & Human Rights	6	2	2
Territory & Sovereignty	5	5	6
Others	5	0	1

TABLE I: *Topical breakdown of the 20 highest-variance resolutions according to the votes of Eastern Bloc members (POL, HUN, CZE, ALB, YUG, BUL, ROM, USSR/RUS, UKR, BLR) during 1989-1991.*

VI. Themes across Resolutions

We now switch emphasis to inferring implicit structure among resolutions. Since voting patterns are responsible for the global embedding, further insight can be obtained by looking at those resolutions that have the highest variance among clusters of countries. In essence we are asking: among nearby countries, which topics are most controversial; i.e., on which neighbors vote differently. We focus, in particular, on the Soviet bloc of Eastern European countries.

Numerical values are assigned to votes: Against (-1), Abstain (0), For (1) so we can compute the variances of the votes of the Eastern Bloc for every UN resolution in the three years around the breakup of the Bloc.

Table I shows a topical breakdown of the 20 highest-variance resolutions among these countries votes in 1989-1991. During the first year most of the attention remained focused on old Cold War issues and matters of development, anti-colonialism, and human rights in the global south. The Soviet bloc had commonly sided with less developed countries against the developed north and west. But by 1990 and 1991 those divisive issues faded, and Middle Eastern issues became dominant. On those issues the US and Israel were in a minority even among other western states. Consequently they became, and have remained, apart from the Assembly majority.

It is clear from this example that there are currents in the resolutions. Our next goal is to discover them automatically. To avoid preconceived notions, we adapt a hierarchical clustering algorithm and an eigenfunction summary method.

VII. Building hierarchical clustering trees

We now seek to organize the resolutions according to how countries voted on them, with the goal of uncovering themes that summarize them. Given the lack of a prior on themes among resolutions – how many there are or, even, whether any exist – we adapt a hierarical clustering algorithm.

For each cluster in the hierarchy, we seek a set of "summary questions" that best approximate large groups of questions underlying the embeddings. This has two advantages: (i) it reduces the dimension of the data set; and (ii) if the summary questions are combinations of small numbers of questions, they are more interpretable. We stress that our approach is in contrast to factor analysis, which leads to factors that are linear combinations of all questions.

Any pair of resolutions are related if they are highly correlated either positively or negatively. For example, during the Cold War period, a UN resolution condemning Israel in Middle East issues will most likely be rejected by the West and supported by the Arabs; however, another UN resolution in support of Israel would lead to the exact opposite voting pattern. Therefore, we study the absolute value of data correlation as a topical similarity function. More formally, this leads to a relatively standard objective function that only depends on dot products. It can be modified using the kernel trick to incorporate non-linearities, in particular those that arise with our diffusion kernel.

We treat each resolution as a vector of responses $\boldsymbol{q}_i$ normalized so $\sum_j \boldsymbol{q}_i(j) = 0$ and $\|\boldsymbol{q}_i\| = 1$. We denote $Q = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$, the set of votes to all resolutions. On the way to designing an objective function, we first seek to find a set of "summary questions" $S = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k\}$ and a clustering $C = \{c_1, \ldots, c_k\}$ of questions with summary questions with the following properties:

$$\bigcup_{i=1}^{k} c_i = Q \tag{10}$$

$$c_i \cap c_j = \varnothing, i \neq j \tag{11}$$

$$\|\boldsymbol{s}_i\| = 1 \tag{12}$$

Equations (10) and (11) guarantee that each question is assigned to a single cluster. We now maximize the similarity between each question and the summary question to which it is assigned. The objective function is: $\phi(C, S) = \sum_{i=1}^{k} \sum_{\boldsymbol{q}_j \in c_i} |\langle \boldsymbol{q}_j | \boldsymbol{s}_i \rangle|^2$

In bioinformatics this is called the diametric clustering objective function [15], and it has an equivalent metric clustering minimization problem. Using the fact that $|\langle \boldsymbol{q}_j | \boldsymbol{s}_i \rangle|^2 \leq 1$

$$\begin{aligned} \arg\max_{C,S} \phi(C, S) &= \arg\min_{C,S} \{n - \phi(C, S)\} \\ &= \arg\min_{C,S} \sum_{i=1}^{k} \sum_{\boldsymbol{q}_j \in c_i} d(q_j, s_i)^2 \end{aligned}$$

where $d(\boldsymbol{v}, \boldsymbol{w}) = \sqrt{1 - |\langle \boldsymbol{v} | \boldsymbol{w} \rangle|^2}$. $d(\cdot, \cdot)$ is a pseudometric, which is to say (i) $d(\boldsymbol{v}, \boldsymbol{v}) = 0$; (ii) $d(\boldsymbol{v}, \boldsymbol{w}) = d(\boldsymbol{w}, \boldsymbol{v})$; (iii) $d(\boldsymbol{u}, \boldsymbol{v}) + d(\boldsymbol{v}, \boldsymbol{w}) \geq d(\boldsymbol{u}, \boldsymbol{w})$. (i) and (ii) are trivial. Proof of (iii) is technical, and is omitted for space reasons.

The maximization version of this problem suggests one heuristic, while the minimization problem suggests another. The first is a modification of Lloyd's algorithm.

procedure ModifiedLloyd($\{q_1, \ldots, q_n\}$)

```
    cluster = initialclustering()
    while φ_old ≠ φ_new do
        φ_old = φ_new
        for i = 1 to k do
            V = concat(q ∈ c_i)                ▷ V = [q_c1|···|q_cm]
            v_i = SVD(V)                       ▷ v_i is largest left sing. vect.
        end for
    end while
    for j = 1 to n do
        put q_j in the cluster that maximizes |⟨v_i|q_j⟩|
    end for
    recompute φ_new
end procedure
```

This algorithm increases the objective function ϕ at each stage. In fact, each for-loop increases ϕ.

Proof: The second for-loop is straightforward, as each question is assigned to the cluster that maximizes the objective. Therefore if any questions change cluster, the objective function will increase.

Let V be defined as above. Then $V = U * D * W^T$ where U and W are unitary and D is a diagonal matrix of singular vectors. Then

$$\sum_{\boldsymbol{q_j} \in c_i} |\langle \boldsymbol{q}_j | \boldsymbol{s} \rangle|^2 = \|\boldsymbol{s}^T V\|^2$$
$$= \sum_i D_{ii}^2 \langle \boldsymbol{u_i} | \boldsymbol{s} \rangle$$

where $\boldsymbol{u_i}$ are the columns of U. This is maximized by setting $\boldsymbol{s}$ to be equal to the largest singular vector $\boldsymbol{u}_1$

Therefore each stage of the algorithm increases ϕ. Since there are a finite number of clusterings, and hence values for ϕ and each stage of the algorithm increases ϕ, it converges, though possibly not to the global optimum. ■

A. Toward Thematic Hierarical Clustering

Although Lloyd's algorithm guarantees a local maximum in the objective function, for our application we seek a related – but in a local sense, slightly different – condition: we guarantee that the absolute correlation distance cannot exceed a threshold.

We start with n individiual singleton clusters of entities E and a data matrix D of m countries (rows) and n resolutions (columns) (e.g. Table II). We also have a correlation threshold $\theta \in (0, 1)$ and a cooldown ratio $\alpha \in (0, 1)$. We repeatedly iterate through the following steps, merging clusters until only one remains:

```
procedure GREEDYCLUSTER(D, θ)
    unallocated = D
    for c in unallocated do
        remove c from unallocated
        for q in unallocated do
            if abs(corr(c, q) < θ) then
                remove q from unallocated
                assign q to cluster c
            end if
        end for
    end for
    reassign questions to most correlated cluster center
    return clusters
end procedure
procedure GREEDYTREE(D, θ, α)
    while numclusters > 1 do
        clusters = GreedyCluster(D, θ)
        set D to largest singular vector of each cluster
        θ = θα
    end while
end procedure
```

Performance is very similiar to the Lloyd algorithm, which could in effect be inserted into the first procedure.

	#3508	#3510	#3515	#3538	#3570
USA	-1	-1	-1	-1	-1
UK	-1	-1	-1	0	0
USSR/RUS	1	1	1	1	1
POL	0	0	0	0	0
CHN	1	1	0	1	1

TABLE II: *An excerpt from the UN voting data [3] of 5 countries (USA, UK, USSR/RUS, POL, CHN) in 1990 on 5 issues, denoted by their roll call id's (RCID): #3508 (Dissemination of information on decolonization) #3510 (Observer status of national liberation movements recognized by the OAU and/or by the League of Arab States) #3515 (Cessation of all nuclear test explosions) #3538 (Calls upon Israel to become party to the Treaty on the Non-Proliferation of Nuclear Weapons) #3570 (Status of the International Convention on the Suppression and Punishment of the crime of Apartheid). The votes are represented by numbers: 1 (Yes), 0 (Abstain), -1 (No).*

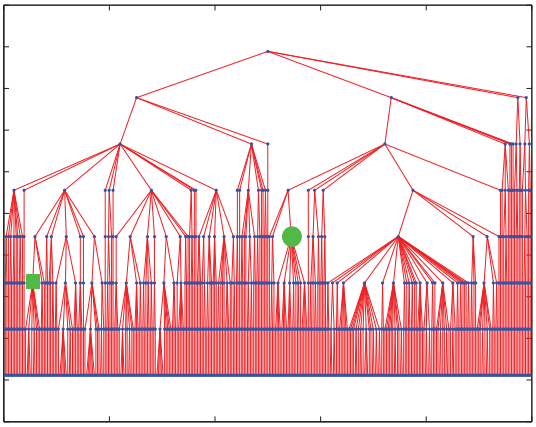


Fig. 10: *Clustering result of UN resolutions during the period 1998-2002. Two individual clusters are marked with • and ■ symbols for demonstration.*

B. Results on UN Resolutions

We applied the clustering algrithm on the set of UN resolutions during the period 1998-2002 [3], with $\theta = 0.95$ and $\alpha = 0.8$. Fig. 10 shows the clustering hierarchy with two clusters • and ■. The resolutions in cluster • pertain only to Middle East-related resolutions, while cluster ■ comprises resolutions from two topics (Human Rights and Nuclear Disarmaments).

We take a more detailed look at the resolutions during the breakup of the Soviet bloc in Figs.11 - 13.

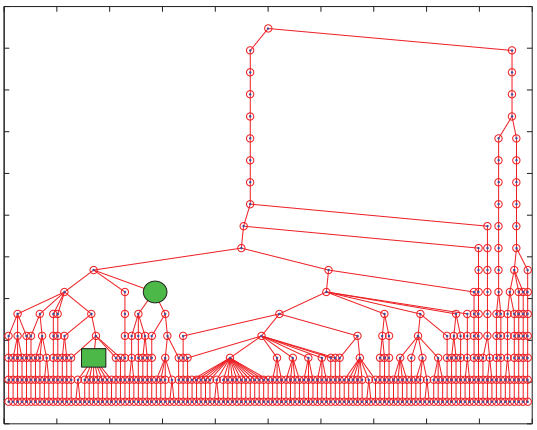


Fig. 11: *Thematic clustering of UN Resolutions 1989. The ■ cluster is about Middle East issues, while the ● is about disarmament and nuclear weapons. The variance in voting patters across Eastern Bloc countries on these issues is virtually 0.*

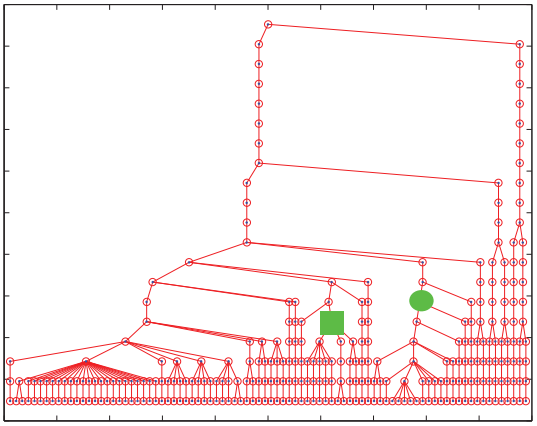


Fig. 12: *Thematic clustering of UN Resolutions 1990. The variance across clusters starts to increase, indicating political change. The cluster ● on Middle East issues is growing larger, while others (e.g. ■ remain fixed on nuclear weapons issues.*

VIII. Discussion

In this paper we developed a diffusion-based approach to embedding high-dimensional UN voting data and showed how to cluster the resolutions "driving" these embeddings. Organization among countries revealed political relationship, and cluster analysis revealed thematic threads running across time. In effect we showed that much of the historical record can be "read out" from UN voting patterns.

References

[1] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps," *PNAS*, vol. 102, no. 21, pp. 7426–7431, 2005.

[2] J. C. Pevehouse, T. Nordstrom, and K. Warnke, "The COW-2 international organizations dataset version 2.0," *Conflict Management and Peace Science*, vol. 21, no. 2, pp. 101–119, 2004.

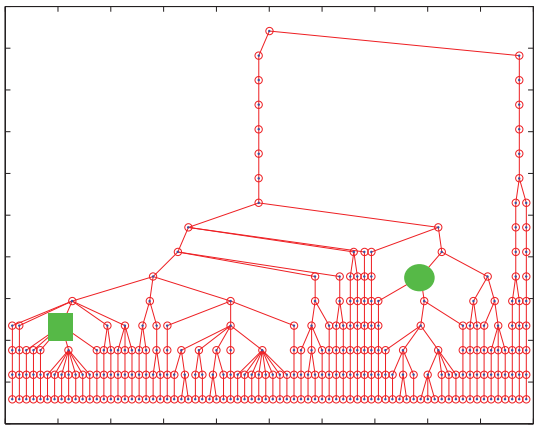


Fig. 13: *Thematic clustering of UN Resolutions 1991. Again the Middle East ● cluster remains while the nuclear weapons cluster ■ enlarges to include economic and other testing issues.*

[3] A. Strezhnev and E. Voeten, "United nations general assembly voting data," 2012. [Online]. Available: http://hdl.handle.net/1902.1/12379

[4] M. A. Porter, P. J. Mucha, M. E. J. Newman, and C. M. Warmbrand, "A network analysis of committees in the u.s. house of representatives," *PNAS*, vol. 102, no. 20, pp. 7057–7062, 2005.

[5] H. Alker and B. Russett, "Discovering voting groups in the general assembly," *American Political Science Review*, vol. 60, no. 1, pp. 327–339, 1966.

[6] B. Kinne, "Multilateral trade and militarized conflict: Centrality, openness, and asymmetry in the global trade network," *Journal of Politics*, vol. 74, pp. 308–322, 2012.

[7] E. Voeten, "Clashes in the assembly," *International Organization*, vol. 54, no. 2, pp. 185–215, 2000.

[8] Z. Maoz, *Networks of Nations: The Evolution, Structure, and Impact of International Networks, 1816-2001*. New York: Cambridge University Press, 2011.

[9] J. Q. Stewart, "Demographic gravitation: Evidence and applications," *Sociometry*, vol. 11, no. 1/2, pp. 31–58, 1948.

[10] C. D. Mayer, *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000, ch. 8, pp. 661–704.

[11] R. R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 21, pp. 5–30, 2006.

[12] K. S. Gleditsch and M. D. Ward, "Measuring space: A minimum-distance database and applications to international studies," *J. of Peace Research*, vol. 38, no. 6, pp. 739–758, 2001.

[13] H. R. Alker, "Dimensions of conflict in the general assembly," *The American Political Science Review*, vol. 58, no. 3, pp. 642–657, 1964.

[14] S. Y. Kim and B. M. Russett, "Voting alignments in the general assembly," in *The Once and Future Security Council*, B. M. Russett, Ed. New York: St. Martin's Press, 1997, pp. 29–57.

[15] I. S. Dhillon, E. M. Marcotte, and U. Roshan, "Diametrical clustering for identifying anti-correlated gene clusters," *Bioinformatics*, vol. 19, no. 13, pp. 1612–1619, 2003.