

# Candidate Mapping: Finding Your Place Amongst the Candidates

Justin Donaldson, William Hazlewood  
 Indiana University School of Informatics  
 jjdonald@indiana.edu, whazlewo@indiana.edu

## Abstract

*The process of voting for a candidate involves selecting an individual who best matches a personal system of values and beliefs. Typically, voters must select a candidate whom they believe fits their issue stances best by determining their approximate similarity to the candidates on the issues, and cognitively positioning themselves amongst the candidates. We show in the context of our candidate position data that the intrinsic dimensionality of candidate similarity in our data can be sufficiently expressed algorithmically in two dimensions using Gower similarity and Sammon mapping. A participant study analyzes how voters choose to position themselves on this low dimensional representation, and how this positioning is related to the position dictated by their actual responses to issues, as well as to their general political stance.*

*Keywords—Political, Visualization, Map*

## 1 Introduction

Political theorists have attempted to explain the behavior of voters by using spatial models of voter choice. These spatial models have their basis in work by Howard Hotelling who studied the equilibrium of spatial economic competition in the late 1920s. This theory was brought into the political sphere first by Downs [1], who proposed spatial voting theory as a model for explaining electoral choice on the basis of voter to candidate (or party) *proximity*. According to Downs, this proximity was determined by a minimal distance between positions on one or more ordinal ‘policy continuums’ (e.g., a continuous set of income tax rates) held by both the voter and potential candidate. A simple Euclidean or Manhattan distance calculation expresses the voter-candidate proximity, which then predicts the vote for a given voter (A full discussion of spatial voting theory is beyond the scope of this article, please see [2] for more information).

Spatial voting theory works well at expressing a voter to candidate proximity in many political contexts. However, since voters typically only select one candidate in a conventional election, the validation information for the model is sparse. It would be far better to solicit more ‘proximity’

information from the voter in terms of candidates whom they felt also matched their interest. However, in elections with larger fields of candidates, collecting this information may be prohibitively time consuming.

Fortunately, in many political situations, many candidates are quite similar in their stances, and their position variance is further limited through expectations carried by party affiliations. It should be possible in some instances to express the variety of candidate position similarity in a lower dimension map, and solicit voter-candidate proximity in a simpler and more direct fashion by asking voters to simply select their position on a map of candidates. In this fashion, more voter-candidate proximity data could be collected at once.

Our work focuses on comparing voter perception of candidate proximity in this fashion by using a participant study. Rather than comparing the voter behavior (the casting of a vote for a single candidate out of many) to the prediction of the model, we analyze how a potential voter chooses to position themselves *implicitly* through the comparison of candidate and participant issue stances, as well as *explicitly* through an analysis of their chosen positioning in the midst of a two dimensional ‘map’ arrangement of candidates. We also focus on exploring differences in political stance (individuals who identify themselves as ‘very liberal’ to ‘very conservative.’)

## 2 Data Collection

In order to generate the low dimensional representation of candidate positions, a wide range of candidate issue stances were gathered and verified from two online sources, <http://www.2decide.com>, and <http://www.ontheissues.org>. The 2decide.com website provided a consistent classification of topics, candidates, and responses for the candidates, while ontheissues.org site provided further validation and extended information on the candidates themselves. Our data includes responses from the following list of ten candidates: Hillary Clinton, John Edwards, Rudy Giuliani, Mike Gravel, Mike Huckabee, Dennis Kucinich, John McCain, Barack Obama, Ron Paul, and Mitt Romney.

Furthermore, changes in candidate positions were de-

tected and saved where possible. This resulted in two candidates holding two positions in the map; One before and after they changed their mind on a given topic. Clinton and Edwards exhibited a shift in issue stance<sup>1</sup>. The candidates' responses were collected for the following issues:

1. Roe v. Wade
2. Death Penalty
3. Education: No Child Left Behind
4. Embryonic Stem Cells: Legalization of Research
5. Energy & Oil: Pursue ANWR Drilling
6. Energy & Oil: Adopt Kyoto Protocol
7. Guns: Assault Weapons Ban
8. Guns: Background Checks for Handguns
9. Homeland Security: Patriot Act
10. Homeland Security: Guantanamo
11. Homeland Security: Waterboarding (torture)
12. Homeland Security: Domestic Wiretapping
13. Immigration: Citizenship Path for Illegals
14. Immigration: Border Fence
15. Internet Neutrality
16. Iran: Sanctions
17. Iran: Military Action as Option
18. Iraq: Initial Invasion Justified
19. Iraq: Troop Surge
20. Iraq: Withdrawal
21. Minimum Wage Increase
22. Same-Sex: Marriage
23. Same-Sex: Civil Union
24. Same-Sex: Constitutional Ban
25. Universal Healthcare

The issue response types were coded as ordinal variables of different types, corresponding to the nature of response required by the question. The ranking of the variables in each type reflect an increasing difference of stance on the given issue. There were two special types used for coding Iraq war positions and same sex marriage positions. All others use the 'Default' type. The following is list of the ordinal variable types:

- **Default:** Supports < Mixed Opinion < Opposes.
- **Iraq war withdrawal:** Immediate Withdrawal < Supports Phased Withdrawal < Opposes.
- **Same sex marriage/union:** Supports < Supports but believes the issues should be left to the states. < Mixed opinion < Opposes but believes the issue should be left to the states < Opposes.

<sup>1</sup>Edwards and Clinton both changed their 'Iraq War: Initial Invasion Justified' stance from 'Oppose' to 'Support'.

<sup>2</sup>For instance, the existence of the 'leftist' democratic party position on the left hand side of the plot is completely (and conveniently) coincidental.

### 3 Data Preparation

The resulting mixed ordinal data are then processed into a Gower dissimilarity matrix. Gower's ordinal dissimilarity [3, 4] is defined as :

$$s_{ijk} = 1 - \frac{|x_{ik} - x_{jk}|}{r_k} \quad (1)$$

Where  $r_k$  is the range of the values for the  $k^{th}$  variable, depending on the validity/existence of the variable under comparison. Any missing information was not included in the calculation of the Gower similarity. The Gower dissimilarity was performed for each candidate pair, which created a dissimilarity matrix  $D$  that expresses the degree to which each candidate is related to every other candidate according to their similarity in position stances.

### 4 Dimensional Candidate Analysis

Sammon mapping was applied to the Gower dissimilarity matrix in order to express the candidate relational information in two dimensions[5]. Sammon maps optimize an error function  $E$  by following a steepest-descent iterative process:

$$E = \frac{1}{\sum_{i < j} [d_{ij}^*]} \sum_{i < j}^N \frac{[d_{ij}^* - d_{ij}]^2}{d_{ij}^*} \quad (2)$$

Where  $d_{ij}^*$  is each element of the Gower dissimilarity matrix  $D$  from above, and  $d_{ij}$  is the *embedding* of each candidate in the two dimensional output space. When embedded in two dimensions, the candidate dissimilarity data had an error of 0.011 ( $\sim 1\%$ ) well below the error tolerance for dimensional scaling. Furthermore, only 16 out of 250 possible candidate responses were missing, and the layout was consistent.

Figure 1 shows the top two dimensions of dissimilarity as the x and y axis, respectively. Candidate positions are labeled with their last names, and previously held positions (by Clinton and Edwards) are colored grey. The distribution of candidates exposes the expected separation between Democratic and Republican party stances. Huckabee, McCain, and other Republican candidates are clustered towards the right side of the plot, while the democratic candidates are all positioned to the left.

It is worth noting that the orientation of the plot axes are somewhat arbitrary. Subsequent versions of the plot of Figure 1 may be inverted through either axis. Therefore, the candidate orientation (whether they show on the left or right side) is arbitrary, but their position relative to the rest of the candidates will not change<sup>2</sup>.

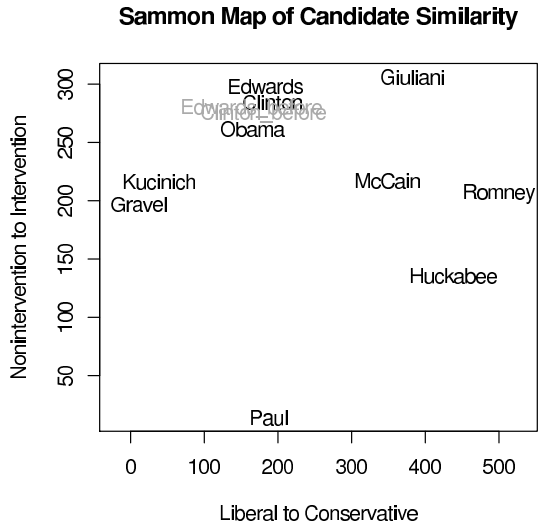


Figure 1: Sammon map of candidate issue similarity.

The second (y) dimension of Figure 1 is based on a level of ‘governmental intervention’ in the form of federal legislation. Ron Paul, on the bottom of this dimension, exemplifies the non-interventionist stance, opposing Federal legislation on nearly every issue – preferring instead to leave many issues to the state governments. Rudolph Giuliani lies on the other end of the dimension, supporting a larger variety of federal legislation. The unit of distance shown in the axes is a shifted and scaled version of the raw distance information. This was done in order to fit a roughly 500 by 300 pixel display applet. All distances referenced in this paper are expressed in this proportionally scaled form for consistency.

A call for participation was announced via e-mail and on various online political message boards. Participants were directed to the study website located at <http://www.candidatemapper2008.net>. In total, around 150 individuals participated and submitted their responses along with their demographic information. After providing the demographic information, they were instructed to fill out a questionnaire survey, measuring their stance on the same twenty-five issues that were gathered from the presidential candidates.

A portion of this questionnaire interface is shown in Figure 2. In this interface, acceptable answers are given in a drop down list under ‘Personal Stance’. ‘Personal Relevance’ drop down selections indicate the importance of the issue to the individual. However, relevance information was not directly included in any of the results presented in this study. Issue topic titles are hyperlinks to relevant wikipedia entries on the subject.

Issue	Personal Stance	Personal Relevance
<a href="#">Roe v. Wade</a>	Oppose	3
<a href="#">Death Penalty</a>	Support	4
<a href="#">Education</a> <small>No Child Left Behind</small>	Mixed Opinion	6
<a href="#">Embryonic Stem Cells</a> <small>Localization of Research</small>	Support	1
<a href="#">Energy &amp; Oil</a> <small>Pursue Oil/US Oil Res.</small>	Oppose	2
<a href="#">Energy &amp; Oil</a> <small>Not Support Nuclear</small>	Mixed Opinion	0

Figure 2: Partial participant questionnaire interface.

## 5 User Study

Once the participants have selected their personal stances on the issues and indicated their personal relevance, their results are saved and added to the responses for the candidates. Then, the same Gower similarity and Sammon mapping process was run on their responses plus the original candidate responses. In order to ensure the similarity and mapping processes could be completed successfully, participants were required to answer all but 3 questions of their choosing.

After successfully completing the questionnaire, participants were then shown the results of the mapping process, with their own calculated location hidden. This interface is shown in Figure 3. They were then asked to position an indicator icon (the large orange circular icon in the upper left corner) near the candidate whom they thought best matched their stance on the issues.

The participants were not given any information on how the candidates were positioned, nor what the dimensions represented. Several candidates dropped out of the race during the course of the study, but the participants were instructed to consider these candidates as active even if they were no longer actually viable in the election. After dragging and dropping the indicator icon and confirming their position, they were shown their ‘actual’ position calculated by the mapping process shown in Figure 4.

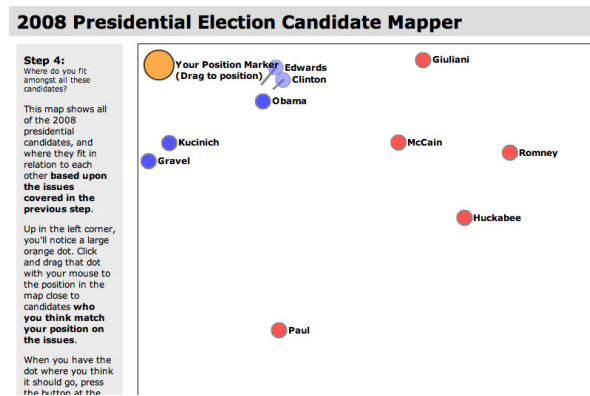


Figure 3: Partial participant candidate mapping interface.

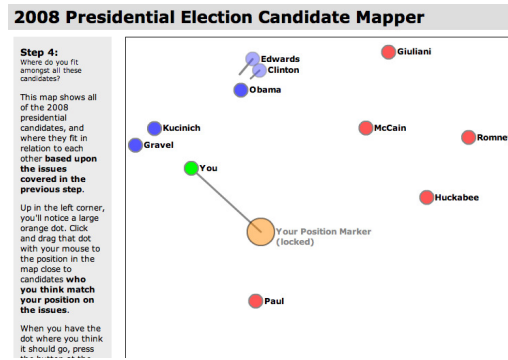


Figure 4: Interface after selecting personal position.

Figure 4 shows the ‘calculated’ personal position in green and labeled ‘You’. The original large indicator icon indicated that the selected position is locked and can no longer be moved. The participants were then encouraged to click and explore the map. By clicking on each candidate, as well as their own selected position, they could compare their position and responses to the candidates using the candidate information panel (which is now visible at this point) shown in Figure 5. Participants were also able to click on links to visit the candidate’s web site, or explore the candidate positions in more depth using the ontheissues.org website.

## 6 Participant Study Analysis

The calculated and selected sets of participant positions were gathered and analyzed. The distance between these positions indicate a form of ‘error’ in their judgement of position. In our study, the magnitude of this distance was related to their demographic information, most notably their main political leaning, which is referred to as a general ‘participant stance’ (from very liberal to very conservative) in the following results. Tables 1 and 2 show the summary statistics and inter-group t-test significance levels respectively. Unfortunately, there were not enough ‘very conservative’ individuals to consider this group. Figure 6 shows the mean distributions of distances (with error bars of length  $2 \times \text{stderr}$ ) for each general participant stance.

Several of the seven participant political stance groups have non-overlapping error bars. Additionally, the t-test values indicate that some political stances are different with at least a 90% confidence level, with others approaching significance.

The data seem to suggest that liberals have a better awareness of their candidate similarity since their error distances are lower. However, this is not necessarily true. The democratic candidates had fewer variations in their stances than the republican candidates, and this lead to tighter relative clustering in the map of all candidates, making it easier for liberal participants to guess their position.



Figure 5: Interface panel showing the list of issue positions.

Table 1: Summary statistics for distance error vs. participant stances (vy = very, sw = somewhat, sl = slightly)

	N	mean	stddev	stderr
vy liberal	42	86.93	39.70	6.13
sw liberal	49	99.03	47.08	6.73
sl liberal	20	98.75	66.46	14.86
neither	20	112.29	60.50	13.53
sl conserv.	8	136.40	65.09	23.01
sw conserv.	22	113.27	68.47	14.60
vy conserv.	2	169.90	73.31	51.83

Table 2: Pairwise t-test values for participant stances (l=liberal,c=conservative)

	vl	swl	sl	n	slc	swc
sw liberal	0.19					
sl liberal	0.47	0.99				
neither	*0.10	0.39	0.50			
sl consrv.	*0.07	0.16	0.19	0.38		
sw consrv.	*0.10	0.38	0.49	0.96	0.41	
vy consrv.	0.35	0.40	0.39	0.46	0.63	0.46

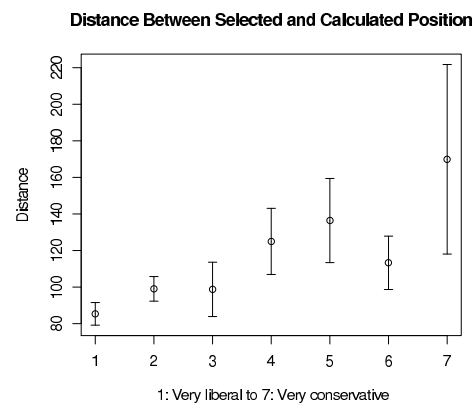


Figure 6: Mean distance distribution between selected and calculated distances, by participant political stance.

Whereas the liberal participants had consistently smaller distance errors, they also had more consistency with the *direction* of their error, particularly on the liberal-to-conservative x axis. After separating out the distances into their x and y components, it is possible to characterize the nature of the distance error in terms of the dimensions of the map (the liberal-to-conservative dimension, and the governmental intervention dimension). Figure 7 shows a boxplot of the distribution of differences along the x ‘liberal-to-conservative’ axis. The majority (the portion of distribution in the box covering the first through third quartiles) of most liberal groups had a distance error in the x dimension, and this error was almost always negative (meaning that they should’ve positioned themselves further to the left, towards more liberal candidates like Kucinich or Gravel). The liberal participants did not show the same consistency in their y dimensional error, with many of the quartile boxes centered on or overlapping 0. The conservative participants did not have a consistent direction of x or y dimension error (except for the marginal ‘very conservative’ group).

Figure 9 superimposes all participant positions onto the original candidate map. Each point in the plot is a selected position by a study participant. An arrow line extends from the point to their position calculated through the Sammon mapping method. The line is colored blue to red according to the political stance indicated by the participant. Each candidate is indicated by a text label over their general position. Since each participant has a candidate map generated especially for their responses, very small differences in candidate positioning occurred. This was not enough to affect the general layout of the candidates in a meaningful way, and we use text labels for candidate positions instead of precise points because of this.

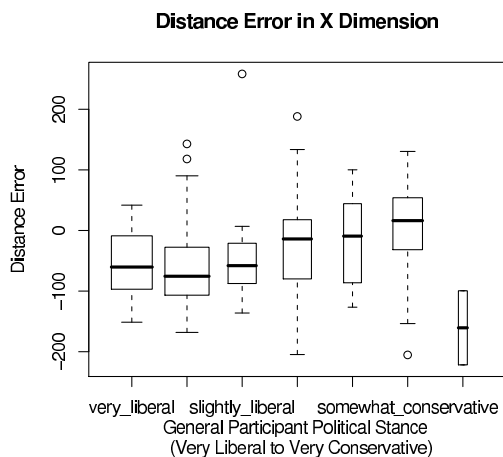


Figure 7: Difference in x dimension

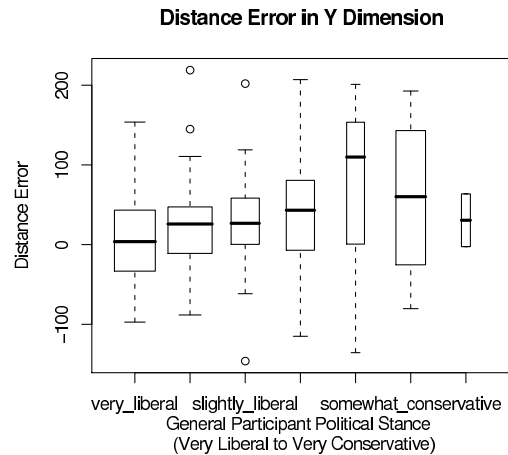


Figure 8: Difference in y dimension

## 7 Conclusions

The candidate mapping process provides a method of directly comparing spatial voting methods to a cognitive approach using explicit positioning. It has uncovered discrepancies in voter explicit and implicit positioning, and is able to express this error in terms of the predominant political sensibilities (the liberal-to-conservative, and interventionist dimensions of the map).

There is a significant and consistent error in the positioning of liberal participants. The basis for this error in the participants (largely twenty to thirty year old graduate and undergraduate students) point to the popularity of Barack Obama. Several participants were surprised to learn of his stance on same-sex marriage (opposes), and the death penalty (supports). With all issues being equal, and no other concerns to bear, the liberal participants would have found better matches with candidates who had a *more* liberal position on the map.

## 8 Discussion

In general, the idea of ‘who is the more liberal or conservative’ candidates is difficult to answer, because the idea (or relevancy) of what a liberal or conservative is can change. Ron Paul’s platform highlights the changes that have occurred in the Republican party. In terms of candidate similarity, Ron Paul is an ‘outlier,’ and very dissimilar from most other candidates. However, his philosophy towards federal governance reflects the ideals that the party was founded under (to minimize federal involvement in legal matters, and let states decide things for themselves). His distance from his peers reflects the change in party attitudes.

Expressing candidate-voter similarity in the manner described here helps not only to describe the predominant political sensibilities, but also helps to better understand

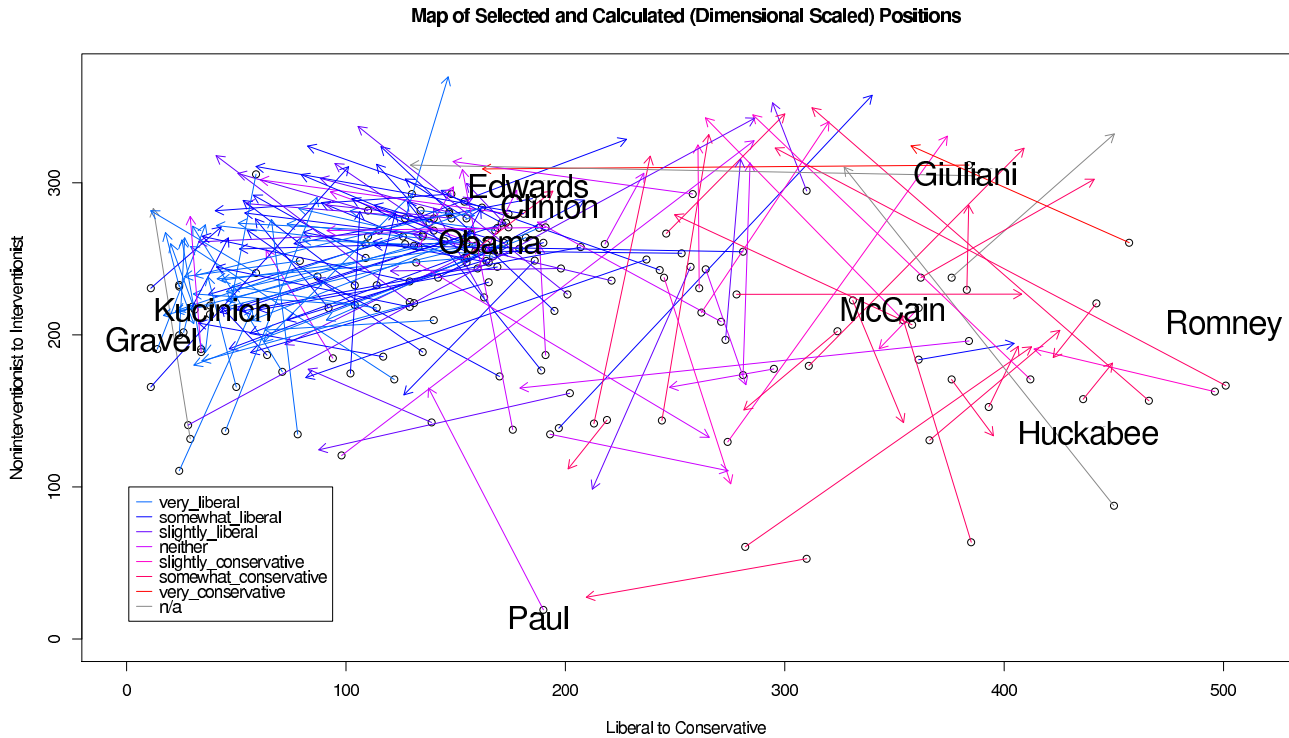


Figure 9: Position of all participants who provided both their issue position stances and their indicated position.

how voters perceive themselves and their potential candidates. Furthermore, the method helps expose any discrepancies that may exist in the voter’s awareness of candidate stances, or in the characterization of their behavior according to spatial voting theory.

## 9 Future Work

The technique described in this work only touches on one of the spatial voting approaches. The other predominant form, known as ‘directional’ spatial voting addresses the concept of ‘issue relevance weights’, which emphasizes the importance of certain issues over others for a given voter [6]. This approach was not discussed here because it was not conducive towards large scale visualization, but is an area for future work for this method.

Furthermore, this approach is not necessarily limited to the political sphere. Other tasks involving personal discernment of a cultural sensibility could be viable domains for this approach, given the underlying basis of the information is conducive for quantitative judgements of similarity and low dimensional representations for visualization.

## 10 Acknowledgements

The authors would like to thank Katy Börner for her discussions and guidance, and Jenya Iuzzini for her support.

This work was sponsored in part by CAREER: Visualizing Knowledge Domains NSF IIS-0238261.

## References

- [1] A. Downs. An economic theory of political action in a democracy. *The Journal of Political Economy*, 65:135–150, 1957.
- [2] J. M. Enelow and M. J. Hinich. *Advances in the Spatial Theory of Voting*. Cambridge University Press, 1990.
- [3] A. D. Gordon. Constructing dissimilarity measures. *Journal of Classification*, 7:257–269, 1990.
- [4] J. C. Gower. A general coefficient of similarity and some of its properties. *Biometrics*, 27:857–871, 1971.
- [5] J. W. Sammon Jr. A nonlinear mapping for data structure analysis. *Computers, IEEE Transactions on*, 100:401–409, 1969.
- [6] G. Rabinowitz and S. E. Macdonald. A directional theory of issue voting. *The American Political Science Review*, 83:93–121, 1989.