A NETWORK AND AGENT BASED MODEL OF POLITICAL POLARIZATION

By

Daniel R. Tabin

A Dissertation Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Major Subject: COMPUTER SCIENCE

Examining Committee:

Boleslaw K. Szymanski, Dissertation Adviser

Rensselaer Polytechnic Institute
Troy, New York

March 2021
(For Graduation May 2021)

CONTENTS

LIST OF FIGURES iii

ACKNOWLEDGMENT iii

ABSTRACT iv

1. INTRODUCTION 1

2. A REVIEW OF THE LITERATURE 3

3. TIPPING POINT MODEL 9

3.1 Broad Overview 9

3.2 Model History and My Contributions 10

3.2.1 Suggestions for Future Research 13

3.3 Mathematical specifics 15

3.3.1 Initialization 15

3.3.2 Updates 15

3.3.3 Measures of Polarization 18

3.3.4 Exogenous Shock 18

4. INITIAL WORK AND A SEPARATE MODEL 20

4.1 Word2Vec, Modularity, and Political Polarization 20

4.2 Suggestions for Future work 23

5. FUTURE WORK AND CONCLUSION 24

LITERATURE CITED 25

LIST OF FIGURES

4.1 PCA of TopTerm and individual Subjects on Subject Word2Vec Word Embedding Space 21

4.2 The Extremely Dense Naive Bipartite Graph of Bill Subjects and Top Terms at Three Levels of Increasing Magnification 22

4.3 The "51" states 23

ACKNOWLEDGMENT

I would first and foremost like to acknowledge and thank my coauthors on the paper "Polarization and Tipping Points" [1] which we submitted to PNAS: M.W. Macy, M. Ma, J. Gao, and B.K. Szymanski. I learned a lot from all four of my coauthors: both about network science and political polarization, but also academia and research as a whole.

I would also like to thank and acknowledge Professor Szymanski a second time. As my advisor Szymanski, gave me the freedom to focus on the research topic I found most interesting (political polarization) and simultaneously directed me to perform actions that would be the most useful for both our paper and my thesis. Professor Szymanski's constant excitement always reminded me of how interesting and important our research is.

Finally, I would like to thank my mother, Jean Tabin, who helped to motivate and encourage me during the stressful and hectic past year, and especially while I was writing my thesis.

ABSTRACT

In this thesis I discuss the model from the paper "Polarization and Tipping Points" which I coauthored [1]. I go into depth about the decisioning and reasoning behind multiple features of the model and discuss possible future improvements and potential future research. During this, I highlight some of my own specific contributions to the model and paper. In addition to giving background to the model we used to analyze political polarization, I give a background in the current academic literature surrounding political polarization as a whole.

1. INTRODUCTION

Why political polarization? Political polarization is a complex phenomenon that is not easy to define mathematically [2], but can be abstractly described as the political opinions of a collection of individuals becoming more divided. Not only is political polarization an intrinsically interesting phenomenon, but it additionally has a large effect on the efficiency and stability of a nation [3][4]. Moreover, political polarization is on the rise. Many Americans feel as though political polarization has been increasing recently [5], and studies that analyze the voting patters of congress also provide empirical evidence for a growth in political polarization [6][7]. Building up an understanding of the processes governing political polarization will hopefully allow future generations to prevent and reverse trends of growing political polarization. This may become increasingly important if current trends continue.

While the abstract concept of political polarization may seem relatively simple, there is a lot of nuance in the subject. As discussed below, polarization may be measured by a multitude of definitions and formulas, and some of these measurements and definitions may in fact be contradictory [2]. Additionally, there are a number of potentially surprising effects of political polarization. For example, in the United States of America a persons opinion on abortion has a statistically significant correlation with their music tastes. Why? Music and abortion, as far as I know, have no relationship to one another. This interesting phenomenon comes from the fact that as individuals polarize politically they often take up the arbitrary and orthogonal cultural practices of their political allies [8].

Political polarization is, by its very definition, a phenomenon derived by the interaction and relationship between multiple individuals. This can be used to produce a network, with nodes acting as representations of said individuals and the links between those nodes representing the relationship and interactions between those individuals. This is the impetus behind our research. We ask: can we use network science to understand and model political polarization and the dynamics that influence it? Additionally, we want to answer questions about the nature of

polarization itself. Specifically, many networks have the notion of some kind of tipping point where a small change in one feature creates a massive change in a different feature[9][10]. Our lab group wanted to investigate tipping points in the context of political polarization, and to see if small changes to model parameters could generate large changes in the polarization of the model. To answer these questions we created an agent based model, which I will describe in greater detail later on in this thesis.

2. A REVIEW OF THE LITERATURE

As computer scientists, it is important to have an understanding of the domain where you are conducting research. While some of my coauthors, including B.K. Szymanski and especially M. W. Macy, had done previous research in this domain, I am a novice in this field of research. As such, I was tasked with reading literature and previous work done in the field. What follows is an overview and summary of said literature.

While Americans feel that polarization is increasing, in what ways this is happening is often contentious [5]. This is not surprising; polarization is a complex topic and can be mathematically defined in many ways [2]. In Understanding Polarization: Meanings, Measures, and Model Evaluation Bramson et al. discuss this. Mathematically defining polarization and understanding said definition is a crucial prerequisite for its analysis. Bramson et al. lay out nine different overall definitions for polarization, and they note that these definitions are not exhaustive. Additionally, Bramson et al. demonstrate that systems can be considered highly polarized under one definition and not be polarized at all under another definition. The nine definitions laid out by Bramson et al. are as follows:

1. Spread: The range of opinions or the difference between the quantification of the "maximum" and "minimum" opinions.

2. Dispersion: The variance of opinions (although any measurement of variation can be used)

3. Coverage: A measure of how much of the opinion-space is covered (and thus a measure of how concentrated opinions are in the covered area)

4. Regionalization: The length of contiguous empty spaces. Thus, if a system has groups of individuals spaced out across the opinion-space it will be less polarized than a system with equal coverage but all of the empty space existing in a single gap within the center of the opinion-space

5. Community Fracturing: The measurement of how many groups the population can be divided into.

6. Distinctness: The measurement of how distinct and separable groups are

7. Group Divergence: The measure of how far the centers of groups are from one another

8. Group Consensus: The variance of opinions within groups

9. Size Parity: The relative sizes of groups to one another

While these definitions may give similar classifications of polarization at times, they also can give conflicting classifications. Take a population of one hundred opinionated agents that can be divided into groups. In one scenario the population is evenly divided into two groups of fifty, and in another there is a group of fifty agents, and twenty five groups of two agents, for a total of twenty six groups. By definition found in size parity, the first group is more polarized than the second; however, the notion of community fracturing tells us that the second is far more polarized. This, as Bramson et al. state, is one of the reasons why it is crucial for a paper to define the measurement of polarization that it is using. These definitions can also be useful when discussing other papers and comparing their methods, but again, this list is not comprehensive [2].

After providing a background in measuring polarization Bramson et al. describe three families of models used to analyze political polarization. They categorize the families as the Axelrod Family of models, Bounded Confidence and Relative Agreement Models, and Structural Balance Models. Each of these families derives from a common source or theme, but can internally vary greatly.

The Axelrod Family of models derive from the seminal paper The Dissemination of Culture: A Model with Local Convergence and Global Polarization Written by Axelrod in 1997 [11]. All of these models rely on agents which interact with and are modified by their neighbors. In his paper, Axelrod demonstrates that complex social phenomena can be modeled by computers using simple rules. Axelrod divides a grid into many squares with randomized values of "culture" effectively a string of

integer values. These squares are agents which can interact with their neighbors, changing a value of their culture to that of their neighbor's. Importantly, the chance of a square interacting with its neighbor is based on the number of shared cultural values at the beginning of an iteration. At the end of a simulation there are cultural regions that internally have identical cultural values, and whose agents share no values with neighbors not in the region. Thus, no updates are possible as interactions are impossible outside the region, and interactions within the region are meaningless (as both agents by definition have identical cultural values)[11]. Axelrod considers the number of distinct regions that exist once the simulation has come to an end to be the value of polarization. This, as mentioned by Bramson et al., is a measure of community fracturing [2].

Klemm et al. has written a number of papers [12][13][14][15] which analyze the Size Parity of the Axelrod model using a measure called "Giant Size Ratio" which is bounded between zero and one. A Giant Size Ratio of one represents a monoculture. Additionally, Klemm et al. note that the range of cultural traits needed for the model to produce an intermediate Giant Size Ratio is rather small, and call this center of this range q*. This value, q*, is also referred to in the literature as the Klemm threshold [2]. In Homophily, Cultural Drift, and the Co-Evolution of Cultural Groups, Centola et al. present a model of the Axelrod family that uses a threshold comparable to the Klemm threshold [2][16]. Additionally, Centola et al. do not allow agents to interact after they become totally culturally orthogonal, even if the two agents become more similar later in the simulation. [2][16]. Flache and Macy give another model which can be grouped within the Axelrod family in the paper Local Convergence and Global Diversity: The Robustness of Cultural Homophily [17]. Unlike previous models, which only care if cultural values are identical, Flache and Macy allow traits to have some level of intermediate similarity. If agents A, B, and C have a cultural trait with values 1, 2, and 5 respectively, the original Axelrod model will say that the cultural similarity of A and B and A and C are identical: 0 for both AB and AC; however, the Flache and Macy model will see agents A and B as more similar than A and C. These three extensions are more concerned with a sense of size parity than the original Axelrod model [2][17].

Bounded Confidence and Relative Agreement Models are based around agents that exist along an opinion space. At every iteration, agents modify their opinions to become the average opinion of all agents within a certain preset threshold of themselves [2]. This model was described by Hegselmann and Krause in a series of papers [18][19][20]. For small thresholds agents form many small groups, for large thresholds agents form a single group, and for medium thresholds two large groups form. This final scenario is considered the most polarized. Further extensions of this family of models were made by Deffuant et al. [21][22]. Some of the key differences between the Hegselmann-Krause model and Deffuant models include the Deffuant model using a continuous function of influence based on distance rather than agents either failing to influence or fully influencing one another, and the Deffuant model includes a metric of "stubbornness" by allowing agents to have differing thresholds [2].

Structural balance models, the third and final family described by Bramson et al. is an approach based on social networks. Nodes in the network are connected to one another where every link between a pair of connected nodes represents either a friendship or an enmity between those nodes. This model is also known as social balance theory and was introduced as early as 1946 by Fritz Heider in his paper Attitudes and Cognitive Organization [23][2]. Nodes will want to become friends with their friends' friends and their enemies' enemies. Similarly, Nodes want to become enemies with their friends' enemies and their enemies' friends. If two nodes A and B are enemies, but both are friends with a node C, then the system is unstable. Either C will pick a side and become enemies with A or B, or A and B will make amends and all three nodes will end up being friends. On a larger scale a network will go from being unstable to stable. The majority of networks take the form of either "universal harmony" with all links being friendships, or "social mitosis" with exactly two groups where all internal links are friendships, and all links between groups are enmities. If the network is fully connected these are the only possibilities, but if the network is not fully connected a social mitosis resulting in more than two groups is possible [24][25][2]. Polarization can be studied from this family of models by treating social mitosis is a form of community fracturing [2]. These models continue

to be analyzed and extended including by allowing friendliness to lie on a scale between negative one and one rather than being in the binary state of friendship and enmity [26] as well as using similarity in place of friendship in a way similar to that of Axelrod type models [27][28][2].

A model not mentioned by Bramson et al. is the Rice Index, described by Rice in Quantitative Methods in Politics in 1928. Rice's index is simple but powerful and gives the ratio of the difference between the number of agents who hold the majority opinion and the minority opinion and all agents [29].

$$RI = \left| \frac{A - B}{A + B} \right|$$

This index can be extended to American political polarization by comparing the ratios of yes votes on congressional bills [6]. This gives a measure of collaboration, which falls into the category of Distinctness given by Bramson et al. Another model that uses this family of polarization measurements is the one given in Portrait of Political Party Polarization by Moody and Mucha. Moody and Mucha use the modularity of "coalitions" obtained by analyzing the co-voting similarity network of congress. This modularity and the amount of covoting is used to track collaboration over each congress [7].

Other models of social dynamics can be modified to model polarization. In Dynamics of social group competition – modeling the decline of religious affiliation, Abrams, Yaple, and Wiener model the decline of religious affiliation in specific regions of Switzerland, Finland, and the Netherlands. Their model uses a differential equation:

$$\frac{dx}{dt} = yP_{yx}\left(x, u_x\right) - xP_{xy}\left(y, u_y\right)$$

Where y and x represent the proportion of the populations in states X and Y (has religious affiliation and does not have a religious affiliation), u_x and u_y represent the utility an individual gets out of being part of group X and Y respectively, and P_{xy} is the a function giving probability that an individual switches from being a part of group X to being a part of group Y and is based on the size of X and the utility obtained from a member (P_{yx} gives the probability of switching from group Y to

group X). This function is symmetric, meaning $x+y = u_x+u_y = 1$. Abrams, Yaple, and Wiener compare the real data from Switzerland, Finland, and the Netherlands with their differential equation [30]. Lu, Gao, and Szymanski relate this back to polarization replacing X and Y with groups of senators who are willing to work with members of opposite parties and senators who solely vote along party lines. This metric of polarization - what proportion of individuals are willing to collaborate with individuals in separate groups - is a good example of a metric not described by Bramson et al. Lu, Gao, and Szymanski similarly compare their equations to real world data. The proportions x and y are given by the Rice index [6][29]. Every two years members of congress are elected. Generally polarization starts higher just after elections and then trends down as the congress continues. Lu, Gao, and Szymanski fix the utility of polarization and the utility of collaboration for each congress, but allow for its value to change between congresses depending on the best fit of the data. They found that the utility of polarization has been increasing, in turn resulting in increased polarization in each congress [6]. This is consistent with other studies of the American legislative body. [7].

This review is by no means comprehensive, but it should give readers who may be new to this field an understanding of the complexity of both Political polarization and the methods of studying it. Choosing a good model and measure of polarization must be done prior to performing any research on the topic. Rigorously defining measurements of polarization will continue to play a crucial role in polarization research which may become increasingly important and common itself if the trend of growing polarization continues in the United States.

3. TIPPING POINT MODEL

In this section, I will give a broad overview of the model used in our research and how we came to tune it to our specific use cases and my contributions to this tuning. Finally, I will outline the math governing the final rendition of our model.

3.1 Broad Overview

The model used in our research is a network of agents that have N orthogonal political opinions which lie in an N-dimensional space. The initial position of any Agent is random. Agents may interact with their neighbors and update their opinions after these interactions. Importantly, interactions may be positive, resulting in the original agent becoming more similar to its neighbor, or negative, resulting in the original agent moving further away from its neighbor in the opinions space. An interaction is determined to be positive or negative based on the distance of the two interacting agents. If a calculation using the distance of the agents as well as some internal parameters is determined to be less than some threshold, the agents become more similar; however, if the calculation returns a value greater than said threshold the agents become more dissimilar. This can be thought of as the following real world examples: if two agents are close enough in opinion their interactions will be positive, and the two agents may learn from one another. On the other hand, if two agents are politically different to the point where their interaction is strongly negative (such as a heated argument), they may come to dislike the other agent, and their opinions, resulting in the agents modifying their opinions to become more dissimilar [1].

This model runs for a preset number of agent updates, or until there is a convergence where the agents may no longer update and the measured level of polarization may no longer change. In our research, my coauthors and I focused on two separate measures of polarization. While it took some time for us to settle on these measurements, we ended up with what we called extremism and partisan polarization [1]. The first metric is a measure of dispersion, while the second is a

measure of distinctness, as defined by Bramson et. al [2].

A final, but crucial feature of our model is to allow an exogenous shock to modify the model partway through run time. This shock is modeled as a new dimensional feature, of which all agents initially agree upon, being added to the model. This represents an external threat to all individuals, such as war, famine, disease, etc, forcing cooperation and at least temporarily decreasing polarization [1]. This shock is used to test how forgone the system is, in terms of political polarization. Effectively we ask: is there an event-horizon-like tipping point beyond which polarization can no longer be reversed?

3.2 Model History and My Contributions

In this section, as well as the discussion of how our model came to be, I will be discussing my specific contributions to the model. While I was, being the most junior member on a team of five, less influential to the final version of the model than some of my coauthors – especially M.W. Macy and M. Ma – the contributions listed here are by no means exhaustive. As with any well functioning team, most decisions were generated via group discussion. While I played some part in the discussions and resulting decisions, those contributions are largely both too numerous and, more relevantly, too minor to be worth discussing in this section. As such, I will be focusing on a few of the more major contributions as well as ideas that did not make it into the final model that I potentially would want to pursue or see pursued in future political polarization research. When mentioning contributions and ideas that are specifically my own, I will state this outright. When discussing other parts of the model in this section, it can be assumed that they were either generated by group discussion or created exclusively by one of my coauthors. Finally, as context for the reader, we implemented our model three times. This was done as to check for correctness and reproducibility. The implementations were created by M. Ma, M.W Macy, and myself, and I will reference these implementations in this section.

Our model was based on the principles of network science, and thus the degree of each node, as well as the overall structure of the adjacency matrix can in theory vary. In the version of our model that was submitted to PNAS, we view the agents

as making up some legislative body, such as the senate. As such, we both set the number of agents to 100, and set the degree of all nodes to 100, that is to say make the adjacency list a 100x100 list of all ones. Both of these decisions are informed by real legislative bodies. In the senate, an American legislative body, there are 100 senators, and all of these senators presumably interact with one another. In future work involving this model, these parameters may be different. For example, if a researcher wanted to model political polarization in the constituency of a legislative body, rather than the body itself, it would make sense to greatly increase the number of nodes, but to have the overall network be far more sparse. As far as I know, no American is regularly interacting with all of their fellow constituents, who easily number in the millions. In some ways, our decision to use a fully connected network of 100 agents represents a step away from network science, as it limits the number of network science related analyses, such as modularity, that can be performed on the network. These changes provide a more useful model for our purposes and we leave variants of our model using other networks for future research either for our lab or other researchers [1].

Another change which potentially removes some traditional network science analyses is the change from the p-q parameter to the party identity parameter. Originally, our model had a variable p and a corresponding variable $q = 1 - p$. These measure the likely hood of intraparty and interparty interactions respectively. Neither p nor q influence the chance of the interaction to be positive or negative, rather only the chance of the interaction to happen. An agent with $p = 1, q = 0$ will only interact with its fellow party members, an agent with $p = 0.5, q = 0.5$ is equally likely to interact with fellow party members and agents of different parties, and an agent with $p = 0, q = 1$ only interacts with agents whose party membership differs from their own. Originally, p and q represented the probabilities of a link existing between two agents in a random graph [31][32]; however we were forced to change this once we had decided to make the network fully connected. Thus, we initially maintained p and q, but had them work as the chance of an interaction in a tick, rather than the chance of two nodes being connected. As we continued the development of our model we found other issues with the relic p and q parameters.

Both W.M. Macy and I pointed out that while tipping points in political polarization due to changes in p and q did arise, they only happened at extreme values of p and q in what often could be considered unrealistic edge cases. For example, a very high q implying that almost all interactions were happening across party lines. Additionally, I personally took issue with the fact that upon initializing the model there is nothing to differentiate agents by party. In the real world parties form due to shared opinions and political goals; however the only thing shared by same-party agents at the start of a run of the simulation was the static feature –more on this later– which labels them as being part of the party. To make this more realistic, I suggested a parameter which shifted the Gaussian distribution of agents away from one another and the center of the N-dimensional space as said parameter increased. This did not make it into the final model, but helped give the impetus for M.W. Macy to introduce party identity to our model. The finalized version of party identity modifies the distance between two agents. It does this by giving a weight which informs a weighted average between the issue distance and party distance of two agents. If party identity is 0, then the distance between two agents is in fact the euclidean distance of the agents in the N-dimensional opinion space. If party identity is 0.5, then the distance is averaged with 0 (if the agents are of the same party), or 1 (if the agents are of differing parties). Finally, if party distance is 1, then all agents are either 0, the minimum distance, or 1, the maximum distance, from one another depending solely on if they share the same party [1].

As briefly mentioned in the previous paragraph, our model allows for static features, although party membership is currently the only static feature used in the paper submitted to PNAS. Unlike dynamic features, which are regularly updated, static features, as their name implies, do not change when an agent is updated. Rather than representing political beliefs, these static features represent intrinsic and unchanging features of agents that may effect who they feel is part of their in-group and out-group such as, race / ethnicity, gender, religion, and political party. While static features can take any value in the continuous $[-1, 1]$ space, static features are always either -1 or 1 and cannot take intermediate values. Even though our model does not explicitly exclude static features such as religion, gender, race,

etc, the PNAS rendition of our model ended up using only a single static feature: party [1]. As with different network structures, there is potential for future research using this model with static features. Additionally, I proposed to my coauthors, and propose to the reader hear, a modification to static features that I believe may be a more accurate representation of the real world. While some traits may be truly static, some traits like religion and party membership are free to change even if they change rarely. As such I believe a semi-static feature may be more applicable in these situations. Such a feature would be allowed to move, but would move much more readily towards the poles -1 and 1 than the neutral value of 0. Thus a very large influence would be required to shift an agent from one value to the other, and they generally will maintain their current value. I personally believe such a semi-static feature will be more important for modeling constituents than members of a legislative body. Although neither are particularly common, constituents changing their party registrations is much more common and likely than an elected official doing so.

3.2.1 Suggestions for Future Research

As I conclude this section, I will discuss two additional ideas to improve the model that I suggested to my coauthors, but which did not make it into the final rendition of the model.

My first suggestion is performing larger updates in each iteration. Currently, the model updates one agent at every iterations and each update is the interaction between exactly two agents. I believe that a speedup can be gained by allowing multiple agents to update continuously and even be updated by multiple agents. This is not a complex change, it simply requires storing two copies of every agent – a parent and a child generation – and then swapping them at every update. At every generation the updated agents would be stored in the child structure, which then would become the parent of the next iteration of agents. Performing iterations in this way provides a few key enhancements. From a computational perspective, doing iterations this way allows for greater parallelization. With no threat of race conditions, thanks to the constant parent array, multiple threads and

or processes can be run. From a theoretical standpoint, I am of the opinion that this is more reflective of real social dynamics. When an interaction happens both parties are effected and additionally, multiple simultaneous social interactions are possible. Thus, in future versions research using this kind of model, I would suggest performing larger updates per iteration.

The second suggestion I will make to improve the model for future research is to add some amount of additional noise to updates. Originally, the way our model updated each agent was deterministic. The initial positions of the agents and which agents were updated at every iterations were both random; however, once two agents had been selected, the resulting update would be identical no matter how many times it was run. To show the robustness and to generally improve our model, we, at the behest of M.W. Macy, modified the update function to move the influenced agent a random amount towards or away from the influencing agent. Additionally, the chance of a positive or negative interaction, while still based on the distance of the agents, is now probabilistic. I suggested adding an additional (tunable) random epsilon to each opinion dimension at every update which may vary on every dimension and update. This new epsilon would also have tested the robustness of the model, and could even provide an additional tipping point, with larger values of epsilon potentially resulting in a breakdown of polarization (as agents move about more randomly). Additionally, like using larger update sizes, I believe that using an epsilon to add noise to the updates will again make the model more reflective of the real world. In the real world, it is possible that after the interaction of two individuals, one individual then agrees more with the second on a specific topic, but further disagrees on a separate topic. Additionally they may randomly change their own opinions based on new information or self reflection without the need for a social interaction. As there is already a system of random noise in the model, there is less of a need to implement this secondary suggestion. I would still encourage it, and due to some randomness being available in the model, similar effects can be gained by making small tweaks. For example, instead of adding an epsilon, the same effect could be gained by probabilistically allowing for positive or negative interactions on each feature, and allowing the random distance moved

to be rarely be negative, or greater than the distance needed to reach the location it is being pulled towards. This later case is when an agent effectively is not only convinced on an issue, but also becomes more radical on that issue than the agent who convinced it. Finally, it should be noted that with any of these additions, future researchers would need to be careful to keep the opinions of all agents within the $[-1, 1]$ range.

3.3 Mathematical specifics

In this section, I will give some of the mathematical formulas in the model, as seen in our paper submitted to PNAS [1]. These formulas appear there and came from discussions involving all authors, but I will lay them out in order to give greater context to this thesis.

3.3.1 Initialization

Each node is initialized with an $N + 1$ dimensional feature array. One dimension is the party. This feature is static and is randomly selected to be -1 or 1. For all other dimensions Z_i, a random number is generated from a normal distribution that has a standard deviation of 0.25 and a mean of 0. Any values that are not in the range $[-1, 1]$ are set to -1 if negative or 1 if positive [1].

3.3.2 Updates

The first step in any update, as mentioned above, is to select two nodes i and j, where i is the node that will be updated and j is the node that is influencing i. It should be noted that the model allows for the case where $i = j$. In this situation, there are no changes, but one iteration is used up. This can be thought of as a time-step where no social interactions happen [1].

Once we have obtained th two agents i, j, their distances are calculated. Using the variable names from Polarization and Tipping points [1] we calculate the distance to be

$$D_{ij} = D_{ij}^{party} * \beta + D_{ij}^{issues} * (1 - \beta), \tag{3.1}$$

Where D_{ij}^{party} is the party distance of i and j: either 0 if i and j are of the same

party or 1 if they are not. D_{ij}^{issues} is the issue distance, and is the distance of i and j across all $N+1$ dimensions and scaled to be within the range $[0,1]$. Finally, β represents party identity which, as mentioned above, is a set value between 0 an 1. Thus, D_{ij} is a weighted average of party difference and issue difference based on the value of party identity [1]. It should be noted that D_{ij}^{issues} is normally the Euclidean distance, but this is not the case after exogenous shock, which is discussed below and defined mathematically in equation (3.5).

Before actually performing the update, we additionally need to determine if the update is positive or negative. This is done probabilistically so that agents that are of an intermediate distance may have positive or negative interactions. This probability is given by a cumulative logistic function given by the distance calculated above, a variable representing dogmatism, and a variable that effects how steep, and thus deterministic the curve is [1]. The formula for this probability, again taken from Polarization and Tipping Points, is as follows:

$$P_+ = \frac{1}{1 + e^{s(D_{ij} - (1-\alpha))}} \tag{3.2}$$

Where e is Euler's constant, s is a tunable constant that gives the steepness of the curve, D_{ij} is the distance as calculated above, and $(1-\alpha)$ represents how dogmatic agents are. As s increases, the steepness of the curve quickly goes to infinity resulting in nearly all values of P_+ being extremely close to 0 or 1 (and thus almost deterministically generating a positive or negative interaction). On the other hand when s is small, a range of intermediate values can generate positive or negative interactions. The variable α is a measure of intolerance or dogmatism and thus $1-\alpha$ is a measurement of tolerance. I note here that α was originally known as dogmatism, but my coauthors and I changed α to intolerance in order to be more in line with a similar forthcoming model produced by Axelrod. In either case, this variable is a measurement of how different far apart agents may be before the chance of a negative interaction increases. As α gets larger, the two agents must be closer to ensure a positive interactions. When agents are highly dogmatic (α is close to 1) they are likely to have negative interactions with agents that even slightly differ from themselves. When agents are highly tolerant (α is close to 0) they are likely to

have positive interactions with all agents except for those who are the most distant from themselves [1].

Finally, once we have determined if i and j are having a positive or negative interaction, we can update i. There are two equations, both again taken from Polarization and Tipping Points [1]. I will list both equations before explaining them in depth.
If i and j have a positive interaction:

$$Z_{if,t+1} = Z_{if,t} + (Z_{jf,t} - Z_{if,t}) * (1 - D_{ij}) * ran, \tag{3.3}$$

If i and j have a negative interaction:

$$Z_{if,t+1} = Z_{if,t} + \frac{L - Z_{if,t}}{2} * D_{ij} * ran, \tag{3.4}$$

In each of these equations $Z_{kf,t}$ represents the f^{th} dimension of agent k at time step t. Thus, at the text time step, $t+1$, agent i's f^{th} feature is updated by adding some additional value [1].

In the positive case, some amount of the difference between i and j on the f^{th} feature $(Z_{jf,t} - Z_{if,t})$ is added resulting in i and j becoming more similar, and i is not allowed to be influenced to the point to where it passes j on feature f. Additionally, we multiply this difference by the inverse of the distance, $(1 - D_{ij})$, so that agents that are closer have a stronger influence on one another. Finally, we also multiply both of these values by ran, a uniform random integer between 0 and 1. All together this results in i necessarily either becoming closer to j or not moving. It also gives the amount of movement some randomness, but results in closer agents being more likely to be more attracted to one another [1].

The negative case is very similar, however instead of moving towards j, i is moved towards the boundary on the opposite side of j. Thus, if i is less than j on feature f, L is 1. On the other hand, if i is greater than j on f, then L is -1. In the rare case that $i = j$, then L is the opposite sign of i and j, and if $Z_{if,t} = Z_{jf,t} = 0$, then L is randomly picked to be -1 or 1. Now that L is described, we can see that $L - Z_{if,t}$ is simply the distance between L and i, which moves i towards L. We divide

this value by two in order to give greater strength to attraction, which makes the model more conservative (as it reduces polarization). Unlike positive interactions, the greater the distance is the stronger the repulsion, and thus we multiply by the distance, rather than its inverse. Finally, we multiply by a uniform random number, just like in the positive influence case. Again, this results in i necessarily either becoming further from j or not moving. It also gives the amount of movement some randomness, but results in agents with a greater distance between them being more likely to repel one another [1]. This is both sensible and similar to the positive case.

Once i has been updated on all N dynamic feature dimensions (with the final party dimension remaining static), the update is considered completed [1].

3.3.3 Measures of Polarization

We need to measure political polarization, both as a metric once the model terminates, as a stopping condition, and as a trigger to cause an exogenous shock [1], which I will discuss further in the next subsection.

As mentioned above, we measure political polarization as extremism, a measurement of dispersion, and partisan polarization, a measurement of distinctness. Extremism is the expected standard deviation of a randomly chosen feature. Thus, we can calculate extremism by calculating the standard deviation of all agents on each issue, and then averaging over the N issue standard deviations. Partisan polarization is the expected difference between two random agents of opposite parties on any given issue. Thus, we simply calculate the average difference between all opposite party pair agents on every issue [1].

3.3.4 Exogenous Shock

The timing of the exogenous shock is governed by a single external parameter, σ which is set before the model begins running. Once the level of polarization is measured by extremism, the shock is initiated [1]. In theory, we also could have mapped σ to partisan polarization, or any other measure of political polarization.

Once the exogenous shock has been initiated, a new opinion dimension is added (or activated depending on implementation) to every single agent. This new opinion dimension is set to be 1 for all agents; however, the new dimension is

dynamic and thus can update as the model continues to run [1]. Additionally, a second parameter γ effects the importance of this new singular dimension. The distance between any two agents i, j is the weighted average between the euclidean distance on the original N opinion dimensions and the euclidean distance on the new exogenous shock dimension [1]. The weight of this averaging is given by γ as such:

$$D_{ij}^{issues} = D_{ij}^{N} * (1 - \gamma) + D_{ij}^{exo} * \gamma, \tag{3.5}$$

where D_{ij}^{N} is the euclidean distance of i and j on the original N opinion dimensions and D_{ij}^{exo} is the corresponding euclidean distance on the new dimension. I also note here that a simple way of implementing this is to set γ to 0 before the measure of polarization reaches σ. When using this sort of implementation, the coder must be careful to also not update the values of the shock dimension – which must remain 1 – until the polarization reaches σ.

4. INITIAL WORK AND A SEPARATE MODEL

Before working on the model described in the rest of the paper, and Polarization and Tipping points as a whole, I tried a different analysis of political polarization. During this period, I performed research combining natural language processing – an interest of mine – with network science. This was originally going to be an enhancement of our other model, but it never quite fit into the paper. Below, I will describe this model and research, the partial results I obtained, and suggestions for future research in this topic.

4.1 Word2Vec, Modularity, and Political Polarization

In order to use natural language processing to analyze political polarization, I downloaded data on congressional bills from congress.gov [33] and then used python packages including Keras [34] and NetworkX [35] to analyze this data. I had three overall goals. First, to find patters in congressional bills. Second, to examine those patterns across the Democrats and Republicans – the two main American Political Parties at the time of this thesis being written. And finally, third, to see if the party difference changed over time.

The data I had from congress.gov [33] provided a "billToopTerm" and "billSubjects" for each bill. The subjects consist of a list of topics that the bill covers, and the top term is a overall category to which the bill belongs, and is also always one of the bill subjects. My first idea for analysis was to use word2vec [36][37], implemented by Keras [34] on the subjects. I had to make a few specifications to the way I ran word2vec to account for the fact that the bill subjects are not actually natural language. For example, I had to make the context window as large as twice the maximum length of any list of subjects. In normal human language, words that appear close to one another are usually more likely to be relevant to their neighbors; however, the bill subjects were simply an alphabetized list. Thus, I had to make sure all of the subjects in a bill were considered equal in the eyes of Keras. This gave me a word embedding of length 512 of each and every subject that appeared

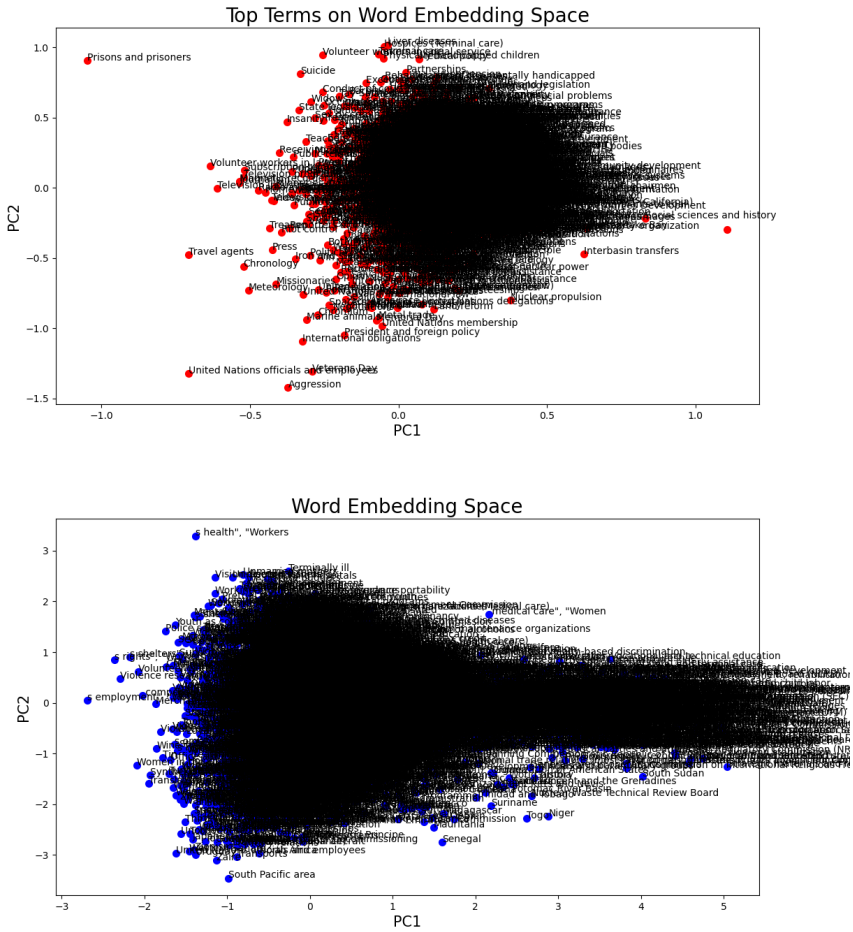

Figure 4.1: PCA of TopTerm and individual Subjects on Subject Word2Vec Word Embedding Space

in the data from congress.gov. I then went through every top term, and averaged over the vectors of subjects that's that appeared in bills with said top term. This gave me a vector value for each top term. The PCA of these embedding spaces can be seen in Figure 4.1.

I found the word embeddings to be sometimes informative on their own, but they were not always extremely meaningful. As a good example of what I mean by that vague statement, look no further than Fig 4.1. In the first two principle components of the Word Embedding Space, many African countries appear together, such as Togo, Niger, Senegal, and South Sudan; however, in between these four

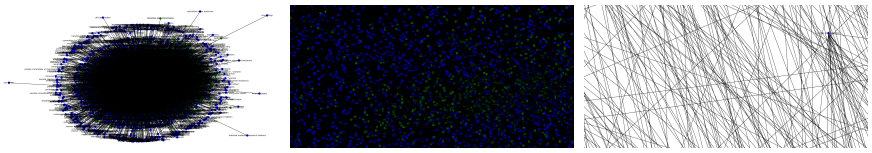

Figure 4.2: The Extremely Dense Naive Bipartite Graph of Bill Subjects and Top Terms at Three Levels of Increasing Magnification

Africans nations lies "Nuclear Waste Technical Review Board," which is in fact not a sovereign nation located in the continent of Africa – if my memory serves me correctly.

From here I switched to using modularity to analyze the top terms. For this rendition of my analysis I created a bipartite network of every bill subject and every bill top term. I naively connected every subject and top term where the subject and top term appeared in the same bill at least once. I then used NetworkX [35] to run Clauset-Newman-Moore greedy modularity [38] on this network. Unfortunately, this network was far too dense with far too many links per node resulting in a single giant cluster being returned with paltry modularity of 0 ± 10^{-16}. The extreme density of this Network can be seen in Fig 4.2

This clearly was not helpful, so I tried a number of methods to make the network more sparse, including requiring more shared bills for a top term and subject to be connected; however what seemed to work best was to create a graph of just the bill subjects, and then to connect the nodes if their similarity, as given by Keras's Word2Vec implementation [34][36][37], was over some threshold. This produced a number of interesting results, and networks with decent modularities in the range $(0.2, 0.4)$. While I did not to enough analysis in this area to provide a highly empirical review of this experiment, I will share here my favorite result from this method: One of the clusters generated was 51 nodes large. In what I have dubbed the "51 states" a cluster contained all 50 states, as well as – humorously and grimly – hurricanes, which was connected to Louisiana and Mississippi. This can be seen, with hurricanes cropped off, in Fig. 4.3.

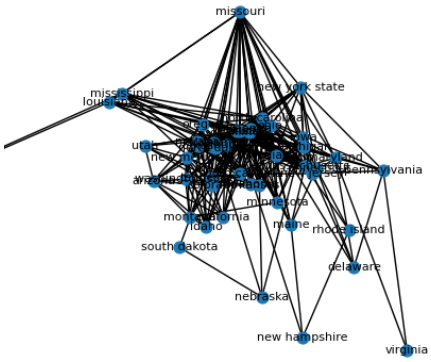

Figure 4.3: The "51" states

4.2 Suggestions for Future work

At this point, my researched switched to focusing on the model described in the rest of my thesis, but I still have a number of suggestions for myself or for some other researcher who picks up the mantle of this research.

First, I would conclude my investigation on modularity and Word2Vec similarity by finding an empirical, quantitative, and evidence based method of choosing a the similarity threshold above which nodes are connected. I then would use this threshold to produce a final network, and then perform some analyses on said network. I might also use this methodology to form a network of bill top terms, in addition to bill subjects.

Next, I would associate each bill with a sponsor, and thus a political party. From here, it would be simple to analyze the vector of the average subject of the average bill by party by year. A number of comparisons could be made from here, such as the difference between these two averages by year. This set of yearly differences would be a measure of polarization, specifically one of distinctness as defined by Bramson et al. [2]. I wanted to make this analysis myself, but ran into difficulty due to this data, which was included in THOMAS –the predecessor of congress.gov – but is no longer being easily queriable from congress.gov [33].

5. FUTURE WORK AND CONCLUSION

The model presented by my coauthors and I in "Polarization and Tipping points" [1] is a robust model which allows for a number of analyses. In "Polarization and Tipping points" we narrowly focus on the tipping point under a specific low agent, fully connected network meant to represent a legislative body. Our model works very well for this cause, and provides a number of interesting conclusions which are discussed in our paper [1]; however, our model is also more broad than this narrow focus and allows for significant modification for other analyses.

In my opinion, as well as that of my coauthors, one of the most interesting and pressing analyses that our model could be applied to is that of a larger, more general population, rather than that of a legislative body. As mentioned above this could be done by greatly increasing the number of nodes while also increasing the sparsity of the network to reduce connectivity. Immediately after making this change, the idea of p and q, the probability of an agent to be connected to any same party or opposite party node respectively in the form of a random graph [31][32], could be added back to the model. A more complex addition that my coauthors and I discussed for this type of model is a system allowing links to dynamically appear and disappear between agents. On the same lines of thinking, other parameters, such as dogmatism, α, or party identity, β could be allowed to dynamically change and or differ from agent to agent.

A second, unexplored analysis provided by our model is the use of static features. Anecdotally, identity seems to be playing an increasingly large role in our politics. These identities include but are not limited to gender, sexual orientation, ethnicity, nationality, and religion. Given that these identities generally do not change over the course of ones life, they could be treated as static features. For identities that may (rarely) change over the course of ones lifespan, semi-static features as described above, could be used. Semi-static features also could improve general population models by allowing party membership to (rarely) change over the course of a model's run. Using static and semi-static features other than party

could answer questions such as: does involving identity with politics increase or help mitigate political polarization? These analyses could then potentially provide information that would help reduce political polarization by encouraging or discouraging the use of identity in politics in accordance with the result of the research.

Finally, and potentially most importantly, is to tune model parameters to fit what we see in real world politics. While these models are inherently informative and explanatory, they lack a certain level of power without being related to the real world. This can be done by tuning parameters such as α, β, σ, γ, s, etc to real world data. An example of this for β, or party identity would be to analyze survey data or roll call voting data as done by Lu, Gao, and Szymanski in "The evolution of polarization in the legislative branch of government" [6]. This data could derive a measure of how much people care about issues compared to raw party-name-difference – exactly the definition of party identity. This real value of β would then allow for the prediction of future trends in political polarization. Predictions from past data, when compared to the current political climate, could allow researchers to tweak and improve the model. Predictions from current data could allow researchers to predict future trends and potentially influence how we deal with political polarization from a policy standpoint.

No matter what changes are or are not made to the model, the general backbone of the model described here and in "Polarization and Tipping points" [1] will hopefully allow for a rich and varied set of analyses on political polarization to be performed.

LITERATURE CITED

[1] Michael W. Macy, Manqing Ma, Daniel R. Tabin, Jianxi Gao, and Boleslaw K. Szymanski, "Polarization and tipping points," *Proceedings of the National Academy of Science,* **118**(50):e2102144118, Dec. 14, 2021.

[2] A. Bramson, P. Grim, D. J. Singer, W. J. Berger, G. Sack, S. Fisher, C. Flocken, and B. Holman, "Understanding polarization: Meanings, measures, and model evaluation," *Philosophy of Science*, **84**, no. 1, p. 115–159, 2017.

[3] N. Mccarty, *Chapter Nine. The Policy Effects of Political Polarization*, pp. 223–255. Princeton University Press, 2011.

[4] R. Sørensen, "Political competition, party polarization, and government performance," *Public Choice*, **161**, pp. 427–450, 12 2014.

[5] A. Gelman, "Economic divisions and political polarization in red and blue america," *Pathways*, vol. Summer, p. 3–6, 2011.

[6] X. Lu, J. Gao, and B. Szymanski, "The evolution of polarization in the legislative branch of government," *Journal of The Royal Society Interface*, **16**, p. 20190010, 07 2019.

[7] J. Moody and P. MUCHA, "Portrait of political party polarization," *Network Science*, **1,** 04 2013.

[8] M. Macy, D. DellaPosta, and Y. Shi, "Why do liberals drink lattes?," *American Journal of Sociology*, **120**, p. 1473, 03 2015.

[9] J. Xie, S. Sreenivasan, G. Korniss, W. Zhang, C. Lim, and B. Szymanski, "Social consensus through the influence of committed minorities," *Physical Review E*, **84,** p. 011130, 07 2011.

[10] D. Centola, J. Becker, D. Brackbill, and A. Baronchelli, "Experimental evidence for tipping points in social convention," *Science*, **360**, pp. 1116–1119, 06 2018.

[11] R. Axelrod, "The dissemination of culture," *Journal of Conflict Resolution*, **41,** no. 2, p. 203–226, 1997.

[12] K. Klemm, V. M. Eguíluz, R. Toral, and M. S. Miguel, "Global culture: A noise-induced transition in finite systems," *Physical Review E*, **67,** no. 4, 2003

[13] K. Klemm, V. M. Eguíluz, R. Toral, and M. S. Miguel, "Nonequilibrium transitions in complex networks: A model of social interaction," *Physical Review E*, **67**, no. 2, 2003.

[14] K. Klemm, E. V. M, R. Toral, and M. S. Miguel, "Role of dimensionality in axelrods model for the dissemination of culture," *Physica A: Statistical Mechanics and its Applications*, **327**, no. 1-2, p. 1–5, 2003.

[15] K. Klemm, E. V. M., R. Toral, and M. S. Miguel, "Globalization, polarization and cultural drift," *Journal of Economic Dynamics and Control*, **29**, no. 1-2, p. 321–334, 2005.

[16] D. Centola, J. C. González-Avella, V. M. Eguíluz, and M. S. Miguel, "Homophily, cultural drift, and the co-evolution of cultural groups," *Journal of Conflict Resolution*, **51**, no. 6, p. 905–929, 2007.

[17] A. Flache and M. Macy, "Local convergence and global diversity: The robustness of cultural homophily," 03 2007.

[18] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis and simulation," *Journal of Artificial Societies and Social Simulation*, vol. 5, 07 2002.

[19] R. Hegselmann and U. Krause, "Opinion dynamics driven by various ways of averaging," *Computational Economics*, vol. 25, pp. 381–405, 06 2005.

[20] R. Hegselmann and U. Krause, "Truth and cognitive division of labour: First steps towards a computer aided social epistemology," *Journal of Artificial Societies and Social Simulation*, vol. 9, 06 2006.

[21] G. Deffuant, S. Huet, J. Bousset, J. Henriot, G. Amon, and G. Weisbuch, "Agent-based simulation of organic farming conversion in allier département," *Complexity and Ecosystem Management: The Theory and Practice of Multi-agent Systems*, pp. 158–187, 01 2002.

[22] G. Deffuant, "Comparing extremism propagation patterns in continuous opinion models," *Journal of Artificial Societies and Social Simulation*, vol. 9, 06 2006.

[23] F. HEIDER, "Attitudes and cognitive organization," *The Journal of psychology*, vol. 21, pp. 107–12, 01 1946.

[24] Z. Wang and W. Thorngate, "Sentiment and Social Mitosis: Implications of Heider's Balance Theory," *Journal of Artificial Societies and Social Simulation*, vol. 6, no. 3, pp. 1–2, 2003.

[25] N. Hummon and P. Doreian, "Some dynamics of social balance processes: Bringing heider back into balance theory," *Social Networks*, vol. 25, pp. 17–49, 01 2003.

[26] P. Gawronski, P. Gronek, and K. Kulakowski, "The heider balance and social distance," *Acta Physica Polonica B*, vol. 36, 03 2005.

[27] M. Macy, J. Kitts, A. Flache, and S. Benard, "Polarization in dynamic networks: A hopfield model of emergent structure," pp. 162–173, 01 2003.

[28] J. Kitts, "Social influence and the emergence of norms amid ties of amity and enmity," *Simulation Modelling Practice and Theory*, vol. 14, pp. 407–422, 05 2006.

[29] S. A. Rice, *Quantitative methods in politics*. 1928.

[30] D. Abrams, H. Yaple, and R. Wiener, "Dynamics of social group competition: Modeling the decline of religious affiliation," *Physical review letters*, vol. 107, p. 088701, 08 2011.

[31] P. Erdos and A. Renyi, "On random graphs," *Publ. Math. Debrecen*, vol. 6, pp. 290–297, 1959.

[32] E. N. Gilbert, "Random Graphs," *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141 – 1144, 1959.

[33] K. S. Badeaux, "Congress.gov," *The Charleston Advisor*, vol. 16, pp. 17–20, Apr. 2015.

[34] F. Chollet *et al.*, "Keras." Git, 2015.

[35] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using networkx," in *Proceedings of the 7th Python in Science Conference* (G. Varoquaux, T. Vaught, and J. Millman, eds.), (Pasadena, CA USA), pp. 11 – 15, 2008.

[36] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013.

[37] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," 2013.

[38] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E*, vol. 70, p. 066111, Dec 2004.