Stochastic Blockmodels for Modelling Political Engagement

by

Cameron Raymond

A thesis submitted to the
School of Computing
in conformity with the requirements for
the course CISC 500 - Undergraduate Thesis

Queen's University
Kingston, Ontario, Canada

March 2020

Supervisor: Dr. Robin Dawes

Abstract

This is my abstract.

Acknowledgments

Blah blah blah.

Contents

Introduction

1.1 Background

The advent of social media has enabled political parties to engage with the broader populous in new and unforeseen ways. The ability to bypass the traditional mediating forces of mass media allows for an unfiltered promotion of policy, ideology and party stances. This is specifically interesting in Canada's political system which has historically been defined by large brokerage parties. In order to win a diverse range of electoral districts across Canada, these "big tent" parties try to appeal to various political persuasions. Political adverts, and policy have traditionally been the conduits through which brokerage parties attempt to accommodate different ideologies, but social media allows for a direct, granular approach to political messaging which is completely novel. Social networks, formed through new media like Twitter, are inherently relational and thus lend themselves well to being represented as graphs. Therefore, as political strategy becomes increasingly digital, the use of graph theory can potentially illustrate how large brokerage parties organize and along what axes Canadians engage with political parties.

1.2 Motivation

The way information is distributed and received has changed significantly over the past decade. Cogburn and Espinoza-Vasquez argue that Barrack Obama's 2008 presidential campaign was a watershed moment in social media campaigning – and in the subsequent decade, from Macron to Brexit to the Five Star Movement, social media has played an increasing role in how politics is conducted [4]. Between 2013 and 2018, the share of Canadian federal media expenditure spent on digital advertising rose from 27% to 65%, a 140% increase, making the study of new media critical from a social science perspective [1]. Additionally, rises in political polarization, populism and a decline in trust in political institutions in the 21^{st} century has been a topic of popular debate. Ezra Klein argues in his 2019 book, *Why We're Polarized*, that this due to a shift in preferences for parties over policies [10]. If this is the case, then this preference to engage along party lines rather than choosing to engage with specific issues should pattern engagement. An empirical analysis of how users behave and engage with political parties online should privilege the relational aspect of social media. Social network analysis helps avoid the pitfalls of survey data, famously described by Allen Barton as "a sociological meat grinder, tearing the individual from [their] social context" [7].

Graph theory's use in social network analysis, also called network science, has already been applied to explore problems in marketing, sociology and epidemiology – but there is a gap in analysis of political engagement online. Therefore, the contribution of this thesis is a novel, robust mathematical process for analyzing different axes of political engagement online in a purely relational manner. A secondary outcome of making relationships a "first-class citizen" in this work will be an organic analysis of what issues produce the most engagement. This deviates from traditional survey data that ask test subjects which issues concern them – and instead uses observations of past behaviour to model variations in local connectivity to answer the question: "what do people actually care about?" All of this will be done in the context of the 2019 Canadian federal election and the tweets of Canada's five major, english speaking party leaders: Andrew Scheer, Elizabeth May, Jagmeet Singh, Justin Trudeau, and Maxime Bernier.

1.3 Research Question

The primary question concerning this project is: in the lead up to the 2019 Canadian federal election, did politically active users on Twitter engage with political elites along the axis of issues (also referred to as topics or policies) or parties? If Ezra Klein is correct, then *who* produces the message will pattern engagement more than *what* the message is; this will act as the initial null hypothesis, with the alternate hypothesis being that who produces the message is equally to or less important than what the message is.

The secondary question to be explored is: during this period, what topics produced the highest level of engagement? And, what topics were important to the broader populous versus a party leader's base?

Social Media and Politics

2.1 Political Communication in the Digital Era

Technological innovation often precedes political disruption; Martin Luther's 95 theses could only be spread at scale by virtue of Gutenberg inventing his printing press first [8]. This example had obvious political ramifications and is a testament to the impact of the means of communication on the political sphere. McNair defines political communication as:

1. All forms of communication undertaken by politicians and other political actors for the purpose of achieving specific objectives.

2. Communication addressed to these actors by non-politicians such as voters and newspaper columnists.

3. Communication about these actors and their activities, as contained in news reports, editorials, and other forms of media discussion of politics. [11]

This is a broad, outward focused definition that includes most public, political discourse – verbal or otherwise. However, social media has changed all three aspects of this definition in novel ways. Most notably, the rate at which politicians can communicate information to achieve "specific objectives" has increased as a result of rapid nature of social media, and the granularity at which actors can target the these messages has gotten smaller due to the massive swaths of user data available [14]. Therefore, the value of using computational methods to study political communication in the 21^{st} century is a natural fit.

2.1.1 Technology's Implications for Democratization

While fully exploring the impact of social media on democracy is out of the scope of this project, it is an important justification for why research in this area needs to be conducted. The traditional model of the media being a mediating force through their reporting, commentary and analysis is no longer valid. Previously, political actors needed to use the media to broadcast their messages to their desired audiences [11]. Social networking sites allow these same actors to reach audiences in the millions without having to gain access to the media first; in this effect, the second and third elements of McNair's definition are also being transformed. Yascha Mounk argues that this has given voice to political outsiders who would be shut out from mainstream platforms. Thus, social media may not be inherently democratic or undemocratic, as it has contributed to democratic backsliding and overturning authoritarian governments, but it can certainly have a destabilizing affect [12]. Therefore, a better understanding of new media's ramifications is critical, and empirical modeling can aide in this understanding.

2.2 Canadian Brokerage Politics

While it is clear that technology is changing how information is received, and thus must also be changing how politics is conducted, it may not be clear the role of Canadian politics in this research. However, Canada's political system is the perfect environment to test the importance of political messaging, because relative to most liberal democracies, the system is dominated by party politicians. As Carty put it:

> No obvious simple geographic reality, no common linguistic or religious homogeneity, no common revolutionary experience or unique historical moment animated [Canada] or gave it life. Canada was created when a coalition of party politicians deemed it to be in their interest to do so, and it has been continuously grown, reshaped and defended by its politicians. [2]

Thus, it is not surprising that Canada's electoral system encourages electoral pragmatism – and developed large, "big tent" parties that are among the most organizationally weak and decentralized of established democracies [2]. This system defines political parties as brokers of the often conflicting, weakly integrated electorate – as opposed to mobilizers of distinct communities, articulating claims rooted in their pre-existing interests. In this way, parties act as the principal instruments of national accommodation, rather than democratic division [2].

Rationale

The dominance of parties in Canadian politics, their amorphous ideological stances, and the many intersectional geographic, linguistic and religious cleavages have given birth to what's been coined the brokerage party system. [2]. The need to capture pluralities in a diverse range of electoral districts gives Canada a unique set of dynamics - making it an interesting case to explore.

Modelling Unstructured Text Data

3.1 Background

Blah blah blah

3.2 Latent Dirichlet Allocation

Blah blah blah

Graph Theory and Computational Social Science

Graph theory is the study of mathematical structures, called graphs, which are used to model pairwise relations between entities. Graphs consist of a finite set of vertices, V, and a set of ordered pairs of vertices, E, called edges. A graph can be defined by the tuple, $G = (V, E)$. The graphs built in this project have added constraints and is defined as below:

- *Vertices*: Let $V_1 = \{v_1, v_2, ..., v_n\}$ be the set of party leaders; $V_2 = \{v_1, v_2, ..., v_m\}$ be the set of tweets by the party leaders described in section 1.2; and let $V_3 = \{v_1, v_2, ..., v_k\}$ be the set of users who retweet tweets. Let the total set of vertices, V, be equivalent to $V_1 \cup V_2 \cup V_3$.

- *Edges*: Let E be the set of edges. Allow the edge $(v_1, v_2) \in E$ if and only if $v_1 \in V_1, v_2 \in V_2$ or $v_1 \in V_3, v_2 \in V_2$. By this definition, we will only allow edges from a party leader vertex to a tweet vertex, or from a"retweeter" vertex to a tweet vertex.

4.1 Spectral Graph Theory

Spectral graph theory studies the structures of graphs via the eigenvectors of their adjacency matrix, Laplacian matrix, or some other variant of the two. The set of eigenvalues for a graph if size n, $\{\lambda_1, ..., \lambda_n\}$, is called the spectrum of a graph [17]. As Hammond et al. note, graph spectra are closely related to major graph invariants [3]. Graph spectra have been used in image segmentation and object recognition tasks, as well as in studying the stability of molecules; Elghawalby and Hancock demonstrated how the euclidian distance between graphs' spectra track the edit distances between graphs [6, 3]. As such, spectral graph theory is useful in comparing the underlying structure of two graphs – allowing for nuanced distance and similarity metrics.

4.1.1 Network Laplacian Spectral Descriptor

While various distance metrics are tracked by graph spectra, most are not size invariant or scale adaptive. Size invariant similarity metrics capture that two graphs that are structurally similar as close together, regardless of magnitude. Scale adaptive similarity metrics would be able to capture both local and global features of a graph. Tsitsulin et al. developed the Network Laplacian Spectral Descriptor (NetLSD), a state of the art similarity measure, that creates extracts a compact heat trace signature from a graph's normalized Laplacian spectrum using the heat kernel [17]. This, in effect, models how heat diffuses throughout a network over; with local features being captured in the immediate time-steps after the nodes are "heated" (and only affecting adjacent nodes) and global features being captured as heat becomes further diffuse. Figure 4.1 shows how the two graphs can be similar at a global level and local level, but differ at an intermediate scale [17].

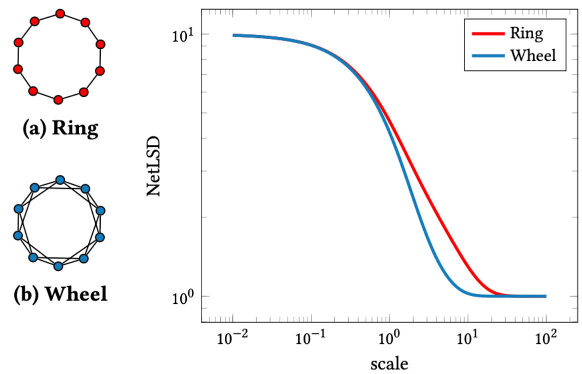

Figure 4.1: NetLSD Heat Trace Signatures for two Similar Graphs

The heat trace (h_t) for a graph at time t is calculated by taking the eigenvalues of the graph's normalized Laplacian matrix, and summating their exponentiation multiplied by $-t$.

$$h_t = \sum_{j}^{n} e^{-t\lambda_j} \tag{4.1}$$

Given that a graph with n nodes will have n eigenvalues and thus have a higher heat trace value at any given time – the heat trace can be normalized against an empty graph, which has all zero eigenvalues [17]. The *heat trace signature* then is a vector of different heat traces at different times denoted by $h(G)$:

$$h(G) = \{h_t\}_{t \geq 0} \tag{4.2}$$

Two graphs can be compared via the L_2 distance among trace signatures.

4.2 Random Graphs

Real world interactions are not deterministic. That is – who your friends are, whether you contract a disease, or whether you choose to retweet a tweet are the result of random processes, patterned by your connections with others. Therefore, modelling how social networks form requires investigating the underlying of the processes that generate these random ties. As Robins' discusses in his article, *a tutorial on methods for the modeling and analysis of social network data*, graphs where edges are generated in some stochastic process can help illustrate how relationships drive behaviour [15].

4.2.1 Stochastic Blockmodels

Initially, random graph models were primarily developed to "model variations in local connectivity of each network actor, the degree of local clustering among actors, and the general distribution of connectivity." [15] However, intrinsic characteristics of actors in a network often patterns the relationships found; inter vs intra-group dynamics drive Shakespeare's Romeo and Juliet, and how the characters behaved was largely determined by which family they were a part of [5]. Blocks therefore group all nodes of the same characteristics. The proportion of all possibles edges within a group – and the proportion of all possible edges that span to all other blocks – formulate the probabilities of forming edges in the generative blockmodel. Figure 4.2 visualizes a friendship network with each node (representing a person) coloured according to their sex, and edges between nodes denoting friendship between them.

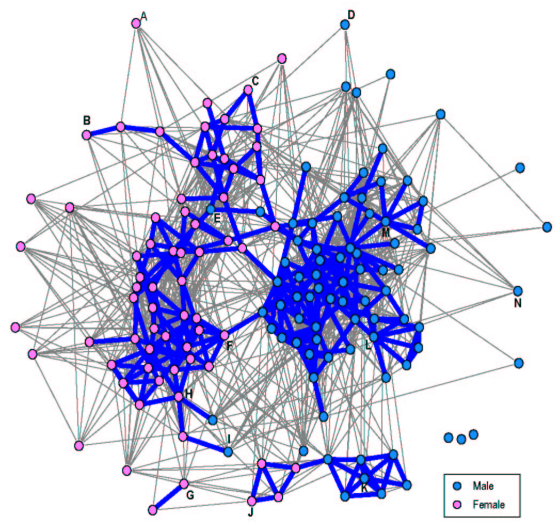

Figure 4.2: Blockmodel Example With Nodes Coloured by Characteristic

Of all the possible edges for a network of that size, the ones observed are overwhelmingly between members of the same sex, with fewer edges spanning that characteristic. These are attributes that shape behaviour that blockmodels attempt to capture.

Stochastic blockmodelling, first proposed by Holland et al. in 1983, is a subset of blockmodels that takes classes of nodes to be latent [9]. The goal, therefore, of stochastic blockmodels are to determine the latent classes necessary to define the blockmodel.

For the purposes of this thesis, stochastic blockmodels will be adapted and extended to capture how users on Twitter show preferences for party leaders and/or policies when deciding whether to engage with tweets. The model will have to realistically model the spectrum of users; from those who only engage with one party leader, to those who retweet multiple party leaders, to those whose engagement driven by specific issues.

4.3 Measures of Centrality

Centrality is a measure of prominence for vertices within a graph. For the purpose of this thesis, it will be used to measure the relative importance of different topics tweeted about in the lead up to Canada's 2019 federal election.

4.3.1 Background

There are various different ways of measuring vertex centrality that have successfully been applied to problems in marketing, economics and epidemiology; Stephenson and Zelen explored the utility of centrality measures in studying the social dynamics of Gelada baboons. [16]. Common centrality measures include measures of degree and betweenness. This thesis will focus on the notion that central vertices are close to other central vertices, which is one of the founding intuitions behind Google's "page-rank" algorithm and eigenvector centrality.

4.3.2 Eigenvector Centrality

As Newman lays out in his 2016, Mathematics of Networks: "the eigenvector centrality [...] accords each vertex a centrality that depends both on the number and the quality of its connections: having a large number of connections still counts for something, but a vertex with a smaller number of high-quality contacts may outrank one with a larger number of mediocre contacts." [13]

If we define the eigencentrality of vertex x as $C_E(x)$, where $C_E(x)$ is proportional to the average eigenvector centrality of x's neighbours, multiplied by some constant

λ:

$$C_E(x) = \frac{1}{\lambda}\sum_{j=1}^{n} A_{xj} C_E(x) \tag{4.3}$$

By defining the vector of centralities as $C_E(X) = (C_E(x_1), C_E(x_2), ...)$ this equation can be rewritten as $\lambda C_E(X) = \lambda A$, and it is evident that CE(V) is an eigenvector of the adjacency matrix with eigenvalue, λ [13]. By Perron-Frobenius theorem, picking the largest eigenvalue of A will result in all elements of $C_E(X)$ being non-negative [13].

Thesis Contribution

Here we go This is the big one

5.1 Topic Modelling

Model topics?

5.1.1 Hyper-Parameter Tuning

5.1.2 Results

5.2 Topic Centrality

5.2.1 Total Network Topic Centrality

Define what it is, show the full graph again.

5.2.2 Party Leader Topic Centrality

Now, consider a fact stating that relation `r` is total, i.e.,

5.2.3 Results

...upper bound on Embee's performance:

$$\text{upper bound is} \begin{cases} O(bN^2) & \text{if } scope \leq 16 \\ O(bF) & \text{if } scope > 16 \end{cases}$$

5.3 Stochastic Blockmodels

Describe generic algorithm without specifying if its the party/topic model

5.3.1 Artificial Neural Network Adaptation

Stochastic Party Leader Blockmodel

The stochastic party leader model generates user behaviour only taking into account the previous party leaders each user had engaged with prior. For each user, when deciding which tweet they are to retweet, their retweet history is converted into a probability distribution (ex. $history = [JT = 0, AS = 1, JS = 3, EM = 2, MB = 0]$ generates $probs = [JT = 0.09, AS = 0.19, JS = 0.36, EM = 0.27, MB = 0.09]$). In this sense, it models a world in which politically engaged Twitter users only engage along the axis of party leaders, with a complete disregard for the topics tweeted about.

Stochastic Topic Blockmodel

Conversely, the stochastic topic blockmodel models a world in which politically engaged Twitter users only engage along the axis of topics. Here, the topic history of a user is converted into a probability distribution, and with a probability of ϵ, that user will retweet the topic with the highest activation – regardless of which party leader tweeted it.

Stochastic Hybrid Blockmodel

The final model developed is a hybrid of the stochastic party leader block model, and the stochastic topic blockmodel. Here, two history vectors for each user are captured – the n dimensional party leader history vector, and the k dimensional topic history vector. After each respective vector is converted into a probability distribution, the weight of a retweet of topic i by party leader j is determined by the function:

$$weight(partyleader = i, topic = j) = \alpha P(i) + (1 - \alpha)P(j) \tag{5.1}$$

Where $P(i)$ is index i of that user's *party leader* probability distribution, $P(j)$ is index j of that user's *topic* probability distribution, and α is some constant that determines the relative weighting of the two. As α approaches 1, the hybrid model becomes equivalent to the stochastic party leader blockmodel – and as α approaches 0 the model approaches the stochastic topic blockmodel. This model then generates different "worlds" in which users' political engagement falls on the spectrum from only caring about *party leaders* to only caring about *topics*.

5.3.2 NetLSD for Describing Political Engagement

The final objective of comparing the relative importance of topics and party leaders in driving political engagement requires comparing the structure of target graph shown in 4, and various hybrid models generated with different values of α. Using the Network Laplacian Spectral Descriptor described in section 4.1.1 by Tsitsulin et al. the optimal alpha value can be determined in a scale-adaptive, size-invariant, and permutation-invariant manner [17].

5.3.3 Results

Summary and Conclusions

6.1 Summary

6.2 Other Work

As dictated by the Queen's School of Computing, 25% of the evaluation for CISC 500 is left to be arranged between the student and their supervisor. Given the nature of this research, lab work or developing software appeared out of scope. I believe that one of the benefits of applied research is its ability to take abstract concepts and frameworks and show their utility in solving complicated problems. This took the form of performing a lecture on the project's research and the application of graphs in Dr. Robin Dawes' CISC 235: Data Structures course.

6.2.1 International Network for Social Network Analysis

The findings and methodologies proposed in this thesis will be presented at the International Network for Social Network Analysis' annual Sunbelt conference in Paris, France.

6.3 Future Work

* better ANNs * different contexts (UK, US, Chile, France) * more variables? * more efficient way of fitting the model

6.4 Conclusion

Bibliography

[1] *Annual report on Government of Canada advertising activities.* Public Works and Government Services Canada., 2018.

[2] R. Kenneth Carty and William Cross. Political parties and the practice of brokerage politics. *The Oxford Handbook of Canadian Politics*, pages 191–207, 2010.

[3] Fan RK Chung and Fan Chung Graham. *Spectral graph theory.* Number 92. American Mathematical Soc., 1997.

[4] Derrick L. Cogburn and Fatima K. Espinoza-Vasquez. From networked nominee to networked nation: Examining the impact of web 2.0 and social media on political participation and civic engagement in the 2008 obama campaign. *Journal of political marketing*, 10(1-2):189–213, 2011.

[5] Patrick Doreian, Vladimir Batagelj, and Anuska Ferligoj. *Generalized blockmodeling*, volume 25. Cambridge university press, 2005.

[6] Hewayda Elghawalby and Edwin R Hancock. Measuring graph similarity using spectral geometry. In *International Conference Image Analysis and Recognition*, pages 517–526. Springer, 2008.

[7] Linton Freeman. The development of social network analysis. *A Study in the Sociology of Science*, 1:687, 2004.

[8] Nathan Gardels and Nicolas Berggruen. *Renovating democracy: Governing in the age of globalization and digital capitalism*, volume 1. Univ of California Press, 2019.

[9] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[10] Steven Levitsky and Daniel Ziblatt. *How democracies die*. Broadway Books, 2018.

[11] Brian McNair. *An introduction to political communication*. Taylor & Francis, 2017.

[12] Yascha Mounk. *The people vs. democracy: Why our freedom is in danger and how to save it*. Harvard University Press, 2018.

[13] Mark EJ Newman. The mathematics of networks. *The new palgrave encyclopedia of economics*, 2(2008):1–12, 2008.

[14] David W. Nickerson and Todd Rogers. Political campaigns and big data. *Journal of Economic Perspectives*, 28(2):51–74, 2014.

[15] Garry Robins. A tutorial on methods for the modeling and analysis of social network data. *Journal of Mathematical Psychology*, 57(6):261–274, 2013.

[16] Karen Stephenson and Marvin Zelen. Rethinking centrality: Methods and examples. *Social networks*, 11(1):1–37, 1989.

[17] Anton Tsitsulin, Davide Mottin, Panagiotis Karras, Alexander Bronstein, and Emmanuel Müller. Netlsd: hearing the shape of a graph. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2347–2356, 2018.