Multiview Representation Learning for Political Science Research

Etienne Gagnon, Political Science Department

McGill University, Montreal

February, 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

Master of Arts, Political Science

Abstract

What is the best way to utilize social media data for political science research? Social media data is heterogenous in nature, meaning that it offers different types of information that are hard to analyze simulatenously. In this thesis, I propose multi-view representation learning, a machine learning framework that learns functions to jointly optimize different sets of vectors, as a technique to analyze heterogenous data. Multi-view learning has interesting potential applications to political science research. Applied research in Political Science typically focuses on one aspect of data. Multi-view learning makes it possible to combine information obtained from the different aspects of data to analyze an outcome. I apply multi-view learning to tweets produced by Canadian Members of Parliament to detect informal social links within the Liberal Party of Canada. The resulting representations correlate better with real-life parliamentary networks than other representation methods currently in use in the literature.

Abrégé

Comment maximiser l'exploitation des données recueillies sur les réseaux sociaux en Science Politique? Les données recueillies sur les réseaux sociaux sont de nature hétérogène, ce qui signifie qu'elles contiennent différents types d'information difficile à exploiter de façon simultanée. Dans ce mémoire, je propose l'apprentissage multi-vue, une perspective en apprentissage automatique qui optimize conjointement différents ensembles de vecteurs, afin d'améliorer l'analyse de données recueillies sur les réseaux sociaux en Science Politique. L'apprentissage multi-vue promet d'intéressantes applications pour la science politique. La recherche en Science Politique tend à se concentrer sur un seul type de données. Appliquée au réseaux sociaux, l'apprentissage multi-vue permet l'exploitation combinée de données textuelles et de données réseaux. J'applique l'apprentissage multi-vue au tweets produits par les députées à la chambre des communes afin de détecter des liens sociaux informels au sein du Partie Libéral du Canada. Les représentations créées avec cette technique sont plus corrélées à un réseau parlementaire hors-ligne que d'autres méthodes présentement en utilisation dans la littérature.

Acknowledgements

Thanks to Professor Aaron Erlich for 2 years of mentorship and the supervision of this MA thesis. Thanks to Annie Chen for being a great labmate and her support through the completion of this MA. Thanks to Professor Elizabeth Gidengil for acting as examiner. Thanks to Professor Dietlind Stolle for comments on an early draft of this thesis. Thanks to Aengus Bridgeman for being an outstanding upperclassmen to our department and a constant source of support for younger students. Thanks to my parents for their financial support throughout my studies. Thanks to Daniel St Germaine for being a positive presence in the office and for proofreading the final version. Thanks to Sean Dempsey for additional proofreading. Thanks to Sean Nosek for his help in navigating the required procedure to have this thesis accepted in time. Thanks to Tara Alward for her help in meeting the departmental deadlines to deposit this thesis. The current research was partially funded by the McGill Center for Cultural and Social Data Science The current research was carried out independently by Etienne Gagnon. Thanks to Alex Wang for his help in securing funding and suggestions of algorithms to use.

Table of Contents

List of Figures

List of Tables

In spite of an increasing number of studies on social media in recent years, political scientists have yet to explore methodologies that can best account for the unique characteristics of social media data (Zhuravskaya et al., 2019). Notably, social media produces very large amounts of heterogenous data. Users output both text, videos, images, as well as social interactions with other users. There is a need to introduce methods that are able to jointly leverage different types of data to the discipline.

In this paper, I assess the potential of multi-view social media user representations for political science research. User representations designate low-dimensional vectors that represent the behavior of users on social media. Multi-view representations combine representations from different sources into a single representation. Multi-view representations allow us to synthetize the information from the different aspects of the data to obtain an unified representation of the user. Multi-view representations are currently used in a variety of applications, including predicting users' brand preferences (Pennacchiotti and Popescu, 2011), substance use (Ding et al., 2017) and various demographic characteristics of users (Benton et al., 2016; Kosinski et al., 2013).

I lay out a framework to apply multi-view representations to social media data in political science. I propose an application that uses tweets made by politicians to infer informal social links within political parties. By creating user representations of politicians, I attempt to detect intra-party relationships within the Canadian parliament.

1 Proposed Approach:

Social media content is heterogenous in nature, which makes it difficult to analyze. A single user is defined in terms of the textual content they produce, the images they post, the videos they upload, or the people they most often interact with. The traditional approach to user representations on social media is to use domain knowledge to manually define features relevant to some task (Pan and Ding, 2019). Examples of features are the presence of a certain set of keywords in users' messages, or whether they follow certain accounts.

In this thesis, I forgo manually defined features in favor of automated user representations. I create automated user representations by learning latent variables present in the data. Latent variables are variables that are not directly observed in the dataset but rather inferred through a mathematical model. Automatically learned representations present significant advantages over traditionally defined features. Complex data produces rich, non-linear interactions that are difficult for even a domain expert to detect or express in terms of simple variables. This is true of both textual and network data, where linear parametric methods cannot effectively capture many important relationships (Bengio et al., 2014). Latent variables are learned directly from the data, to most efficiently retain the information important to the task at hand, while discarding irrelevant information. This approach is the core of what is called *representation learning* in machine learning.

The key to jointly leveraging heterogenous types of data is to use typical representation learning techniques to create separate vector representations of each type of data. Data represented in vector form can then be related through a mathematical function that maximizes the ressemblance between the vectors. In this paper, I construct user representations from low-dimensional vectors of their textual content and their interaction graphs. I use two single-view representations to create a unified representation through a multi-view learning technique called generalized canonical correlational analysis (GCCA). I apply GCCA to create representations of the MPs in the Liberal Party of Canada (LPC) on the basis of their tweets. I cluster MPs using the k-means clustering algorithm to identify groups of MPs that share similar behavior. The representations are a new tool that makes it possible to observe intra-party politics in the Canadian context. I first detail how separate vector representations of textual content and social interaction data are made. I then demonstrate how to create a unified representation through GCCA.

1.1 Learning Single-View Representations

Multi-view learning first requires the production of different sets of representations, or "single-view" representations for the different categories of data that I want to leverage.

This means that I use a separate training process for the textual and network data. A multi-view framework then associates both types of representations. Though I present a general framework to make multi-view representations, other researchers should adapt the techniques to make single-view representations to their own application and data.

The Skip-Gram Algorithm

As a preliminary step, I present the skip-gram algorithm, which contributes to both types of single-view representations for this application. The skip-gram algorithm is a method that creates vector representations of observations that also contain information about the context that they appear in. Mikolov et al. (2013) first presented the algorithm to create vector representations of word meaning by encoding the context that the word appears in. The approach is based on the idea of "distributional semantics" (Harris, 1954), which states that words that appear in similar contexts have a similar meaning. This algorithm is extendable to other settings, such as graph representation, where the context of an observation yields important insights (Perozzi et al., 2014).

Take a sequence of words W composed of n words.

$$W^n = (w_0, w_1, ..., w_{n-1}, w_n)$$

Where each individual word w is part of a vocabulary V. The researcher pre-defines a context window of size k. Words in the k positions around a given word are considered as that word's context C. The goal of the skip-gram algorithm is to find a set of parameters θ that maximizes the probability $p(C|w;\theta)$, which is the probability that the context words are predicted as being in the context of w. Formally:

$$\arg\max_{\theta} \Pi_{w \in W} \Pi_{c \in C} p(c|w;\theta)$$

In accordance with the neural network literature, a softmax function fits the parameters θ.

$$p(c|w;\theta) = \frac{e^{v_c * v_w}}{\sum_{c' \in C} e^{v_c * v_w}}$$

Where v_w is a vector representation of the current word and w_c are vector representations of the context words. Extracting a value v_w is therefore the real aim of the algorithm. The denominator normalizes this quantity against the sum of all possible contexts. The θ parameters are the vector representations in the softmax. A single layered neural network typically fits the parameters. A single layered neural network is a 2-stage classification model. In the first stage, the neural network takes in an input vector α composed of j features over X observations and creates a set of m derived features Z.

$$Z_m = \sigma(\alpha_{0m} + \alpha_m X)$$

Where σ is a non-linear transformation[1] applied to the linear relationship inside of the parenthesis. α_{0m} represents an intercept value and Z_m represents a derived feature in Z. In the second stage, a linear relationship to the outcome is estimated using the derived features as the input. In a classification task with K possible outcomes this gives:

$$V_k = \beta_{0k} + \beta_k Z$$

and the softmax is applied to the V values to yield a probability value. How does the neural network converge towards an optimal set of weights? The neural network is assigned a "loss function", which is a criteria that it tries to minimize. The appropriate criteria typically depends on the type of outcome that the neural network is trying to predict. A commonly used loss function is the cross-entropy loss, expressed with the following formula. w is an observation in the set of all observations W, C is a vector that represents possible context observations to that observation, y is a binary indicator that takes the value 1 if c is actually found in the context window of w and 0 otherwise, p is the

[1]Traditionally a sigmoid function, although in recent years the rectified linear unit function (ReLU) has been the most popular choice.

probability that the neural network assigns to c being classified as being in the context of w:

$$-\sum_{w\in W}\sum_{c\in C} y_{c,w}\log(p_{c,w})$$

The neural network weights are fitted using gradient descent. It is also possible to use a more complex architecture to learn more complex relationships. However, empirical results find that training simpler models on a larger corpus yields better representations than training complex architecures over a smaller corpus (Bengio et al., 2003; Mikolov et al., 2013).

This means that the algorithm learns similar vector representations for words that share similar context words. When the size of the vocabulary V is large, the denominator of the softmax is too computationally expensive to compute. In these situations it is typical to use a hierarchical or negative-sampled variant of the softmax. I refer the reader to Goodfellow et al. (2016) for a thorough treatment of this topic.

The Textual View

The first type of information I leverage is textual data. The goal is to learn a low-dimensional representation of a single user's content. The text data has 3 levels: The word level, the tweet level and the User level. In other words, what do we know about the words used by a user, what do those words tell us about the tweet that contains them, and what do the tweets tell us about the user?

Traditional text analysis methods treat words as discrete units of analysis (ie: the word is present or not in the document). However, in recent years, low-dimensional dense vector representations of words are driving much of the progress in various NLP task. Many techniques exist to create these low-dimensional representations, such as the skip-gram algorithm from the previous section. Current models learn word representations from billions of words large corpuses. A large corpus makes it possible for the skip-gram algorithm to learn a wide variety of associations between words, which gives very detailed representations of a word's meaning (Mikolov et al., 2013). Dense vector representations perform well in many learning tasks since they carry a lot of pre-learned information about the words. It is easier to complete learning tasks from that starting point compared to sparse one-hot vector[2] representations, where a model needs to learn all the information about the vocabulary, alongside the specific relationship of interest.

For this paper's application, there are roughly 10 million words in the corpus, which is not large enough to train reliable word representations. When the corpus is not large enough it is common to use word-vectors pre-trained on a large corpus of related data. I use GloVe[3] pre-trained vectors from the Stanford NLP group trained on a corpus of billions of tweets (Pennington et al., 2014). Applied research shows that pre-trained generic representations do not perform significantly better or worse than representations trained on a domain specific corpus for political science applications (Spirling and Rodriguez, 2019).

Creating higher level document representations is an active area of research in Natural Language Processing. Current research uses techniques that rely on aggregating the words in the document (Harris, 1954; Salton and Buckley, 1988), deriving document representations from topic models (Ng and Jordan, 2002), learning unsupervised document representations from a large corpus (Le and Mikolov, 2014) and learning document representations from a supervised task (Wieting et al., 2016). The preferred technique depends on the downstream task for which the representation is used, as well as the available data in the corpus. In this situation, I create representations of the author of the tweet from the text. A supervised task that predicts the author from the content is therefore a natural choice. Supervised task-specific document representations show a strong in-task performance compared to more general unsupervised embedding methods, at the expense of

[2]A one-hot encoding is a vector representation of categorical variables where the category being represented takes a value of 1 and all other categories a value of 0. Encoding the word "like" in a 1 000 words large vocabulary gives a vector of length 1 000 with 0s in positions representing all other words and 1 in the position associated to "like".

[3]GloVe is an algorithm for training word representations that is broadly analoguous to the skip-gram algorithm from word2vec.

their generalizability to other tasks (Wieting et al., 2016). Different political science applications of multi-view representations may require to adopt a different strategy to best represent documents.

The process that generates a text content representation of a user is the following:

1. As a pre-processing step, I machine translate French tweets to English with the google translate API. All tweets are lemmatized to allow for an easier retrieval of their embedding in the GloVe pre-trained corpus.

2. I query individual words in each tweet to a lookup table that returns a 50 dimensional GloVe pre-trained word embedding upon matches. I reject words that are not present in the GloVe vocabulary. GloVe provides word vectors for 1.2 million English words. This means that rejected words will mostly consist of very infrequent words in the English language and proper nouns. The fact that GloVe's vectors are trained on a dataset of over 2 billion tweets means that vectors are available for most forms of informal language. This includes a large set of abbreviations that are common on Twitter. Emoji and hashtag representations are not available in the standard GloVe corpus and are therefore rejected. On average, 10% of the tokens per tweet are rejected at this step.

3. Tweets are now a sequence of word representations. A neural network[4] takes the sequence as an input and predicts the author of the tweet. I extract the final hidden layer of the neural network and treat it as a tweet representation. The intuition behind this step is that the final hidden layer of the neural network represents a set of latent variables that relate the tweet's content to its author. Similar MPs' tweets therefore appear in similar regions of the vector space.

4. I average all tweet representations from a single MP along each dimension, yielding an overall user representation of the content produced. All users now have a repre-

[4] I use a Bi-directionnal Long-Short Term Memory (Bi-LSTM) network in this paper, though it is possible to apply any arbitrary architecture.

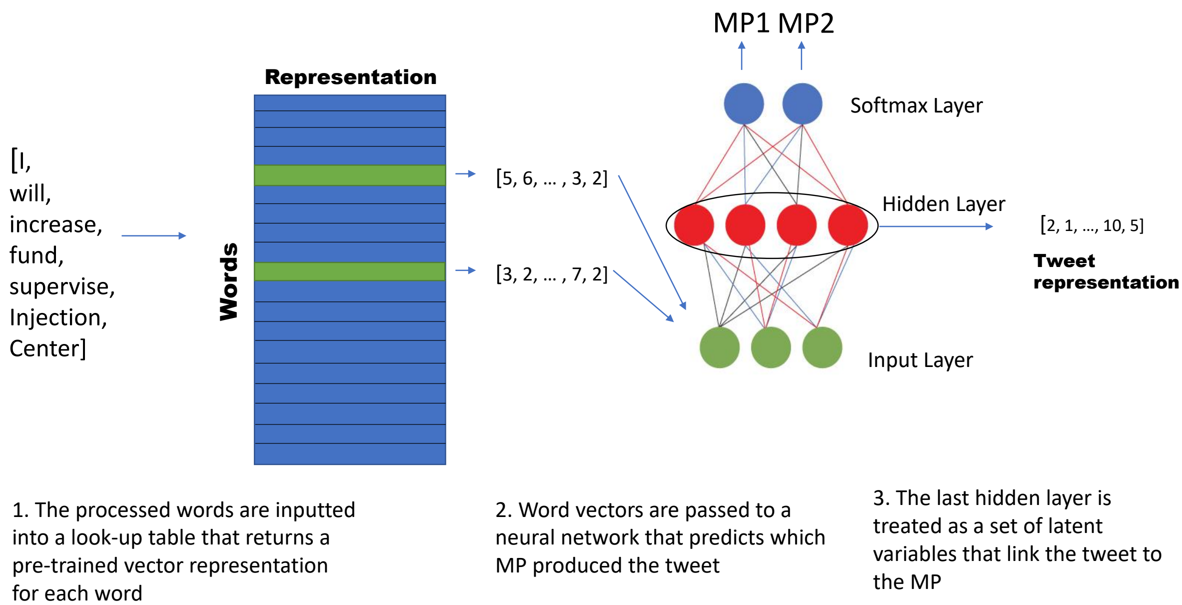


Figure 1: The pipeline that creates a tweet representation. The tweet representations of each MP's tweets are averaged along each dimension to yield a MP representation

sentation of equal dimensionality in the same vector space, which makes it possible to easily compare them. This type of averaging is common practice when making user level representations from social media data (Benton et al., 2016; Ding et al., 2017).

The Social View

I create a vector representation for the social interactions between the users through a process broadly analogous to word vector creation. The current application is to the Twitter interactions of politicians. There are two types of Twitter actions that I consider an interaction between users. Interactions are when a user retweets another user,[5] or when a user mentions another user in their tweet.[6]

[5]I do not make a distinction between retweets and ”quotes”. The latter being a retweet that includes an additional comment by the retweeter.

[6]When users use the @ operator to mention another user's handle.

All interactions are discrete actions that are unilaterally taken by a user. The best way to express them is as a directed graph $G = (V, E)$, where vertices V represent a user and edges E represent a type of interaction from user $i \in V$ to user $j \in V$ for any given user i and j. A multigraph representation, with different types of edges for the different types of possible interactions may appear to be the most natural here. The problem is that typical methods for graph representation do not generalize to multigraph structures. I represent the different types of interactions in 2 separate graphs, with edges in graph A being the mentions and edges in graph B being the retweets.

To learn a vector representation of each graph, I apply a slight modification of the Deepwalk algorithm (Perozzi et al., 2014). Deepwalk is an algorithm that works as follows:

1. Initiate a simple graph $G = (V, E)$.

2. Start a large number of short random walks of a set length from each vertex.

3. Collect the sequence of vertices visited by a walk starting from a given vertex. This sequence is the context of that vertex. This process is illustrated in Figure 2.

4. Use the skip-gram algorithm to find a set of k derived features that maximize the probability that a vertex has a similar representation to vertices that often appear in its random-walk neighborhood. Taking the example walk from Figure 2:

$$p(1, 2, 3, 4|BlueNode)$$

   is one of the quantities that skip-gram tries to maximize, alongside the result of all other random walks. The quantity is maximized through the process described in section 1.1.

I modify this algorithm by using biased walks instead of random walks. The biased walks pick an edge based on the probability w assigned to it. The probability of going from user

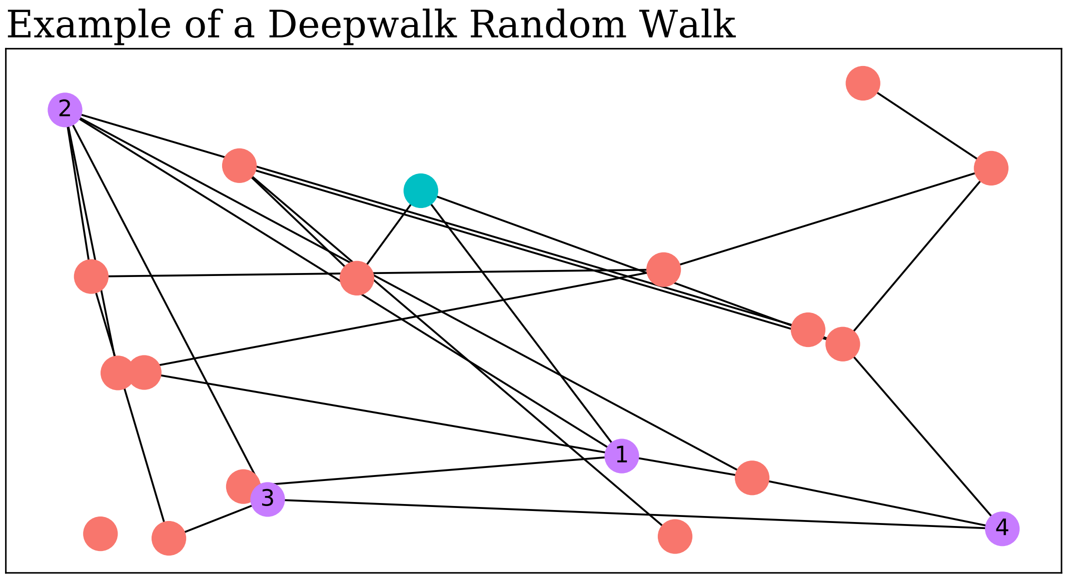


Figure 2: Example walk from Deepwalk. The walk originates at the blue node and visits the purple nodes in the indicated sequence. In the full algorithm a large number of similar walks of length 4 would be initiated from each node. The nodes visited by random walks (purple nodes) originating from a node is its context in the skip-gram algorithm.

i to user j is the ratio of the number of interactions from user i to user j, divided by the total number of interactions with other users initiated by that user. In other words:

$$w_{ij} = \frac{k_{ij}}{\sum_{u}^{V} k_{iv}}$$

One can imagine the graph as a language model, with vertices representing words. Words visited by a random walk that starts from a word $i \in V$ are the context words to that word (Perozzi et al., 2014). From this point of view, the vertex representations of Deepwalk are analogous to the word representations from the text representation.

A potential concern with this method is how the inclusion of highly popular accounts may affect the representations. As an example, Justin Trudeau is the most retweeted MP for all Liberal MPs in the dataset, due to the fact that he is the Prime Minister. A high number of retweets of Justin Trudeau may not necessarily be indicative of a close relationship. There are two possible issues related to this. The first is the questions of whether highly retweeted MPs will have an inordinate impact on the representations. The second is whether backbenchers interact with the party leadership on a different basis than with other backbenchers. Regarding the first of these points, including highly popular accounts actually does not have an important impact on single observation's representation. According to information theory, the information I carried by an event x is proportional to the inverse of the logarithm of its probability of happening.

$$I(x) = -log(P(x))$$

Intuitively, since this event is likely to happen, there is already an expectation that it will happen and it therefore does not contribute much information when it occurs. A related concept to information is Shannon's entropy $H(x)$, which is the expectation of the information provided by a distribution.

$$H(x) = -\sum_{x} P(x) \log P(x)$$

The probability of Justin Trudeau having been retweeted by another MP can be expressed as the conditional probability $pr(JT|MP)$. If we take the extreme case where Justin Trudeau is retweeted by every single MP, then $pr(JT|MP)$ will be 1 for every single MP, meaning that the contribution of the presence of Justin Trudeau to the entropy of the distribution of any MP's retweets will be equal to:

$$P(1) \log P(1) = 0$$

Justin Trudeau being retweeted by another MP actually provides little information for the neural network to update its weights.[7] This is an idealized scenario but it demonstrates that including Justin Trudeau does not have an inordinate impact on the representations.

[7] A corollary of this is that MPs who are retweeted by half of the MPs are the ones with the highest impact on the representations.

Another way to understand the impact of highly retweeted MPs is to consider that representations generated through the skip-gram algorithm are made to resemble observations found in their context as much as possible. If we take two vector representations V_1 and V_2, the difference in their embedding $D(V_1, V_2)$ is a function of the symmetric difference in the set of observations found in their context $C1$ and $C2$. If Justin Trudeau is found in the context of all observations, then he will not appear in any of the symmetric differences between two contexts. He will therefore not contribute to the difference between any of the vector representations.

The question of whether backbenchers interact with the party leadership in a way that may alter the results significantly is an intersting question. I explore this question in Appendix 1.

1.2 Learning a Multi-view Representation

A multi-view framework is a method that relates different sets of representations. In this case, the idea is to find some common dimension that relates social activity and content produced. This calls for a "co-regularization style" algorithm (Zhao et al., 2017). Co-regularization algorithms penalize representations of an observation for being dissimilar to other representations of the same observation. I follow other recent studies on social media user representations that use general canonical correlation analysis (GCCA) (Benton et al., 2016).

GCCA is a technique that learns a single representation from multiple-views. GCCA finds linear-transformations of the different single-view representations to maximize their linear correlation in transformed space. GCCA is an extension of CCA, or canonical correlation analysis. In CCA, two linear projections w_x and w_y are found for vectors X and Y, which correspond to different views of the data. The linear projections attempt to maximize this linear correlation coefficient:

$$\frac{w_x^T C_{xy} w_y}{\sqrt{(w_x^T C_{xx} w_x)(w_y^T C_{yy} w_y)}}$$

Where C_{xx}, C_{yy} and C_{xy} are co-variance matrices that correspond respectively to $\frac{1}{n}XX^T$, $\frac{1}{n}YY^T$, and $\frac{1}{n}XY^T$. Maximizing the previous formula corresponds to the following optimization problem. The way to solve this optimization problem is discussed in more details in appendix 2.

$$\max w_x, w_y \;\; w_x^T C_{xy} w_y$$

$$s.t\ w_x^T C_{xx} w_x = 1, w_y^T C_{yy} w_y = 1$$

Expressed differently, maximizing the correlations between the projected vectors is equivalent to minizing the following function:[8]

$$g_{w_x,w_y}(X,Y) = ||w_x X - w_y Y||^2$$

Subject to the same constraint as earlier. Observed as the last formula, CCA imposes a regularization constraint on both sets of projected representations, in such a way that they ressemble one another as much as possible. GCCA is a more general version of CCA that can account for more than 2 sets of vectors. GCCA finds transformed embedding G and linear transformations U_i, in such a way that:

$$\underset{G,U}{\arg\min} = \sum_i \|G - X_i U_i\|^2 s.t\ G^T G = I$$

Where X_i is the data from the i th view. This creates a single representation that summarizes and puts in relation information from heterogenous sources.

[8] A demonstration of this can be found in Shawe-Taylor et al. (2004).

2 Application: Detecting Intra-Party Groups

I apply multi-view social media user representations to the detection of intra-party groups in the Canadian parliament. The acquisition of political power through electoral competition necessitates the cooperation of various actors within a party, who typically share a set of common interests but also present important divergence in their policy emphasis (Greene and Haber, 2017). Many studies in political science apply text-processing techniques that detect latent ideology to make inferences with regards to intra-party politics. Social media content produced by politicians is an interesting tool to observe intra-party politics, since it makes it possible to observe politicians behaving in a low-cost setting. In contrast to approaches that focus on detecting politician's ideology, I propose multi-view user representations to observe intra-party dynamics that are not based on the ideological distance between politicians, but on other factors such as personal affinities, clientelistic links, or shared regional interests.

Intra-party groups are a feature of intra-party politics in most party systems (Adams et al., 2011; Harmel et al., 1995; Kölln and Polk, 2017). Depending on the context, intra-party groups are more or less institutionalized. They can range from simple groupings of politicians that share affinities to institutional party factions, defined as "any intra-party combination, clique, or grouping whose members share a sense of common identity and common purpose and are organized to act collectively – as a distinct bloc within the party – to achieve their goals." (Zariski, 1960). Competition between intra-party groups affects the elaboration of the party platform and policy agenda (Ceron, 2012), coalition formation (Gianetti and Benoit, 2009) as well as portfolio allocation (Greene and Haber, 2017).

In spite of their important influence on party politics, intra-party groups are under-developed in terms of both conceptual approaches and hypotheses (Verge and Gómez, 2012). This is due to the fact that intra-party groups are often an informal, sometimes disorganized feature of a party's social network, making them hard to observe directly. Most studies have centered on a few cases: Italy's Christian-Democrats, Socialists and

Post-Socialists, Spain's Socialist Party, and Japan's Liberal-Democratic Party , with some work also directed at the United Kingdom and the United States (Boucek, 2009). A shared trait of these parties is that their groups are well institutionalized factions. This public and institutional status is due to there being a gain to be had from publicizing the existence of factionalism inside of the party. An example of this is the Japanese Liberal-Democratic Party (LDP). The Japanese electoral system prior to the 1994 electoral reform consisted of a single non-transferable vote (SNTV) in multi-member districts (MMD). The LDP, which typically obtains the large majority of the vote share in Japanese elections, had to present multiple candidates in each district to maximise its potential of winning seats. The LDP candidates engaged in fierce intra-party competitions for the seats inside of a district. For LDP candidates, publicizing their membership in an institutionalized faction of the party allowed to present ideological separation from other LDP candidates (Cox and Rosenbluth, 1993). This was often key in winning seats, since winning votes from the supporters of same-party candidates is easier than obtaining the support of other parties' supporters (Thayer, 1969).

However, in most party systems, intra-party politics are difficult to observe . Therefore, there is a need for techniques that effectively measure intra-party politics, which will in turn allow for empirical studies of intra-party politics and its effect in other contexts. Recent scholarship suggests that social media can play an important role in making this phenomenon visible.

2.1 Observing Intra-Party Politics on Social Media.

Social media is a new means of observing intra-party politics. Traditionally, studies in intra-party politics use actions carried out in a legislative body as a source of data. One example is a rich literature that leverages cross-country comparisons of roll-call votes to infer the effect of political institutions on party cohesion (Carey, 2007; Godbout and Høyland, 2017; Hix and Noury, 2016; Indriason and Kristinsson, 2015; Rice, 1925). The effect of a party member's ideology on his roll-call votes record is another often studied phenomenon (Ceron, 2015; Jenkins, 2006). Another literature analyzes the text of parliamentary speeches or debate (Ceron, 2015; Proksch and Slapin, 2010; Schwarz et al., 2017), or the text of party platforms (Catalinac, 2016; Ceron, 2016). Social media offers interesting advantages over traditional types of content. One advantage of social media is it can be organized into a network structure, which provides additional information when compared to simple text (Barberá et al., 2019). Second, social media data makes it possible to obtain individual assessments of politicians at a very low cost (Bond and Messing, 2015). Third, when compared to content produced in parliament, social media data is less prone to the confounding effect of party discipline, where the party leadership prevents individual legislators from showing their true preferences (Carrubba et al., 2008; Depauw and Martin, 2009; Ecker, 2017). This is because observing politician's behavior in settings where the cost of not acting cohesively with the rest of the party is low is more likely to be indicative of their true preferences (Ceron, 2017; Gianetti and Benoit, 2009). The cost incurred by not acting cohesively with the rest of the party is minimal on social media, where politicians are not as heavily scrutinized as in other settings, such as media addresses or parliamentary speeches (Ceron, 2017).

Studies that leverage social media up to now have mostly attempted to scale politicians on an ideological axis (Ceron, 2017), which makes it possible to infer the effect of ideological differences between politicians of a same party on intra-party politics. Ideology is typically measured through either social network analysis or text content analysis. In the social network analysis approach, studies analyze the follower-following network of social media users. An adjency-matrix of user relationships is created and a latent variable representing ideology is extracted through dimensionality reduction techniques (Barberá, 2015; Bond and Messing, 2015; King et al., 2016). In the text analysis approach, algorithms that aim to extract ideology positions from text, such as the wordfish or wordshoals algorithms (Lauderdale and Herzog, 2016; Slapin and Proksch, 2008), are applied to the tweets produced by politicians to infer their ideology (Boireau, 2014). Ceron (2017) notably looks at the effect of ideological distance found on social media posts on endorsements between candidates of the same party. These studies are inspired by a larger literature that uses text-analysis methods on legislative or electoral content produced by politicians, to assess their ideology (Catalinac, 2018; Ceron, 2015; Proksch and Slapin, 2010; Schwarz et al., 2017).

Ideological scaling of politicians is an important influence for the intra-politics literature, due to the fact that it makes it possible to obtain direct measurement of politician behavior. However, qualitative accounts of intra-party groups show that, while different ideological tendencies exist within parties, politicians do not necessarily form intra-party groups on the basis of ideological proximity (Boucek, 2009; Rose, 1964). In some contexts, politicians from the same region will form regional groups that are better able to bring benefits to local constituents (Croissant and Chambers, 2010; Slider, 2010). In other party systems, such as Japan and Italy, clientelistic links exist between a group leader and subordinate politicians. The decision to join a group is then mainly influenced by the ability of a group leader to provide ressources to less influential politicians (Bettcher, 2005; Curtis, 1999). Legislative bodies form small communities where politicians often interact on a personal basis. Personal affinities can often play a role in how politicians associate (Jeffrey, 2017). Finally, politicians may associate with other politicians that prioritize the same issues, even when said issues are not tied to a specific ideology (Heinkelmann-Wild et al., 2019; Hilderman and Thomas, 2013).

Methods that observe intra-party politics only on the basis of politicians' ideology estimates are not effective for observing other dimensions on which politicians may associate. The current paper's approach offers two important innovations to previous work. First, the literature on estimating ideology positions shows that meaningful information can be obtained from both politician's social networks and their produced content. Multi-view representations combine two independently successful approaches into unified representations, allowing to jointly leverage the data. Second, the previous literature looks at politicians using uni-dimensional ideology measures. The multiview approach does not limit the representation of politicians to a set number of dimensions, allowing to possibly observe more complex relationships between political actors.

3 Case: The Liberal Party of Canada

I test multi-view representation on the Liberal party of Canada during the 42nd Canadian parliament, which was active from December 3rd, 2015 up to the general election of October 21st, 2019. Canada is an interesting case in the study of intra-party politics. A salient characteristic of the Canadian party system is the high level of party discipline exhibited by elected representatives. While early iterations of the Canadian parliament showed internal dissent in political parties, contrarian members were gradually replaced by more loyal leadership supporters, leading to party members showing near perfect cohesion on parliamentary votes (Godbout and Høyland, 2017). Various theories attempt to explain this transformation, including the gradual professionalization of MPs (Kam, 2009), party leaders restraining the ability of MPs to introduce legislation (Cox, 2005) or partisan sorting and the ability of the government to control the legislative agenda (Godbout and Høyland, 2017).

Strong party discipline makes intra-party politics very difficult to observe in the Canadian context. However, intra-party politics are not perfectly consensual. While overt dissent is rare in the Canadian parliament, intra-party disagreements do happen. They are typically handled behind closed doors (Malcolmson et al., 2016). To inform expectations with regards to the kinds of groupings that will be found by user-representation proximity in the Liberal Party of Canada, I offer a short account of its intra-party politics.

The Liberal Party of Canada

The LPC is the oldest and longest-serving political party in Canada. It dominated Canadian federal politics for much of its history. While there is a tradition of unity and loyalty to the party's leadership, the late 1990s and 2000s were marked by internal competition between the supporters of different party figures. Accounts of caucus meetings of the LPC in the 2000s highlight the lengthy discussions necessary to manage the rift between supporters of former Prime Minister Jean-Chrétien and Prime Minister Paul Martin, as well as a proliferation of "sub-caucuses" focused on specific issues during that period (Hilderman and Thomas, 2013; Jeffrey, 2017). After years of electoral misfortune through the late 2000s, the party re-unified under the leadership of Justin Trudeau and seized power once again in 2015.

The state of intra-party politics in the LPC following the election of 2015 is uncertain. No recent academic work offers a comprehensive account of what current dividing lines exist between the MPs. There exists evidence to suggest that the caucus is not completely unified. MPs have spoken out against a perceived centralization of power within Justin Trudeau's leadership group (Berthiaume, 2016). Are old dissenssions from the pre-Trudeau era still in place within the LPC? By studying the LPC's multi-view representation, I attempt to gain a better understanding of its internal dynamics.

3.1 Data

The data I use for this project consists of the entirety of the tweets, retweets and replies of the 176 sitting MPs from the Liberal Party of Canada posted from January 1st, 2018 to the 43rd Canadian federal election, held on October 21st, 2019. I gathered tweets through the Twitter API. Figure 3 shows the distribution of the count of tweets produced by politicians in the dataset over the time period.There is a total of $270,822$ tweets contained in the dataset. The figure shows that there is a time trend in the number of tweets, with MPs' twitter usage increasing as a function of time. This is likely due to the increasing importance of using social media to reach out to voters, as well as the beginning of a pre-electoral period from 2019. The important decrease in the number of tweets around January 2019 is likely due to politicians tweeting less during the holiday period.

A difficulty when working with Canadian data is that politicians, especially from the Province of Quebec, tend to use both French and English when appealing to consituents.

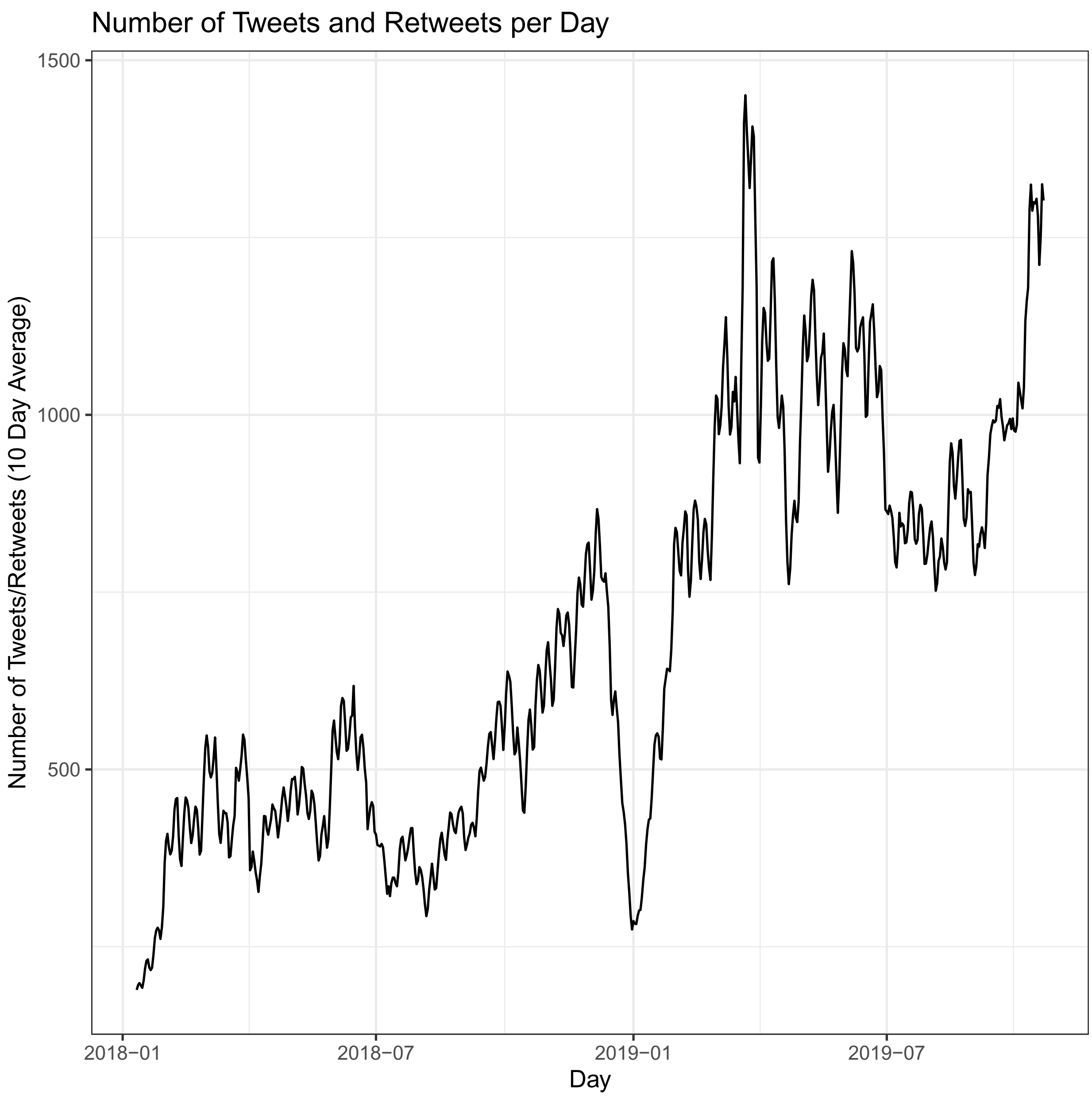


Figure 3: Tweets per day in the dataset.

	Min	5%	25%	50%	Mean	75%	Max	Std. Dev
Tweets	3	62	323	736	808	1184	3199	588
Retweets	3	63	251	620	715	1073	3725	749.5
Mentions	2	4	51	107	129.92	182	480	107.67

Table 1: Distribution of Tweets/Retweets/Mentions per MP.

The corpus contains $51,716$ French tweets. To allow the processing of both English and French tweets on similar grounds, French tweets are machine translated using the Google translate API. A more general problem when working with social media is that the level of usage of the platform is not constant across all users. There is a concern that some MPs are not active enough to obtain a reliable MP representation. Fortunately, all LPC MPs exhibit some level of social media usage. Table 1 shows descriptive statistics on the distribution of tweets and retweets by individual member of parliament. Even MPs at the 5th percentile of Twitter usage provide hundreds of posts to analyze. Low social media usage is therefore not an issue in the current project. The number of tweets and interactions recorded provide enough data points to create accurate representations of almost every MP.

3.2 Clustering:

For the specific application of detecting intra-party groups, I aggregate the MP representations into a number of clusters. The groups here can be considered as a way to heuristically aggregate politicians who have similar representations. This clustering of the data is done through the K-means algorithm. K-means is a method that partitions n vectors of identical dimensions into k clusters to minimize a pre-determined distance measure. In this case I use the sum of within cluster squares. Formally this is expressed as :

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{X \in S_i} \|x - \mu_i\|^2$$

One issue with this algorithm is the necessity of using a predetermined number of clusters, meaning that I need to give a prior expectation of the number of groups that will be detected. Some heuristic methods that rely on visualizing principal components of the data in 2 dimensions or finding critical values of k where the reduction in within-cluster variance stops decreasing when adding more clusters can also be leveraged Romesburg (2004).

An important issue with k-means clustering is that, as the number of dimensions increases, the significance of euclidean distance as a measure of distance diminishes. This is due to the phenomenon known as the "curse of dimensionality" where adding dimensions to the data makes it sparser. In the limit, euclidean distance becomes meaningless as a measure of distance. Various techniques can be used to counteract this phenomenon in multi-view representation learning. One possibility is to limit the dimensionality of the single-view representations, thus diminishing the effect of the curse of dimensionality. The second possibility is to use dimensionality reduction techniques, such as principal component analysis (PCA), to select certain dimensions in the data that are especially informative. In the current application I use PCA in 2 dimensions. PCA is a procedure that uses an orthogonal transformation of the data to produce a set of variables that are not linearly correlated. The retained variables are the principal components of the data that explain the most variance. This makes it possible for k-means clustering to function efficiently at clustering the data and also makes it possible to visualize the data in 2 dimensional plots. Visualizing the data helps with the interpretability of the analysis.

4 Results

The following section presents the intra-party description of the Liberal Party of Canada that results from applying the proposed method. I discuss the projection of the produced representation of the LPC. I then discuss the clusters found in the LPC. I analyze cluster membership with respect to a set of socio-demographic factors that may be related to

MPs' social proximity. I finally use topic modelling to highlight differences in the content produced by members from the different groups. The full list of MPs and their cluster is shown in appendix 3.

4.1 Representation of the LPC

I applied the methods laid out in section 1 to the tweets produced by MPs of the LPC. I obtained a set of 64 dimensional multi-view embeddings. Using PCA in 2 dimensions I show a visualization of the produced embeddings in figure 4. MPs in various sections of the representation are labelled on the plot for perspective.

A few things stand out about the produced representation. First, the projected representations do not form clearly separated clusters. This result is consistent with accounts of the LPC's intra-party dynamics, which underscore the party's unity behind Justin Trudeau's leadership.

Some well documented relationship between MPs are well represented by the multi-view representation. Justin Trudeau and Mélanie Joly, who are childhood friends and still entertain close links, have very close representations on the projection. Chrystia Freeland, rumored to be a possible heir to Justin Trudeau (Mitrovica, 2019), and current Deputy Prime Minister of Canada, is also located close to the Prime Minister on the projection. On the other hand, some relationships are not as well depicted. I do not observe proximity between Justin Trudeau and Seamus O'Regan, in spite of the two men having a close relationship.

4.2 Choosing a K-Number of Groups.

As mentioned in section 3.2. I need to set a number of k-clusters to input into the K-means algorithm. I do not have any theoretical expectation with regards to the appropriate number of k-groups that needs to be found. I therefore apply heuristic methods to find an appropriate number of groups in the party. The first method I apply is the "elbow

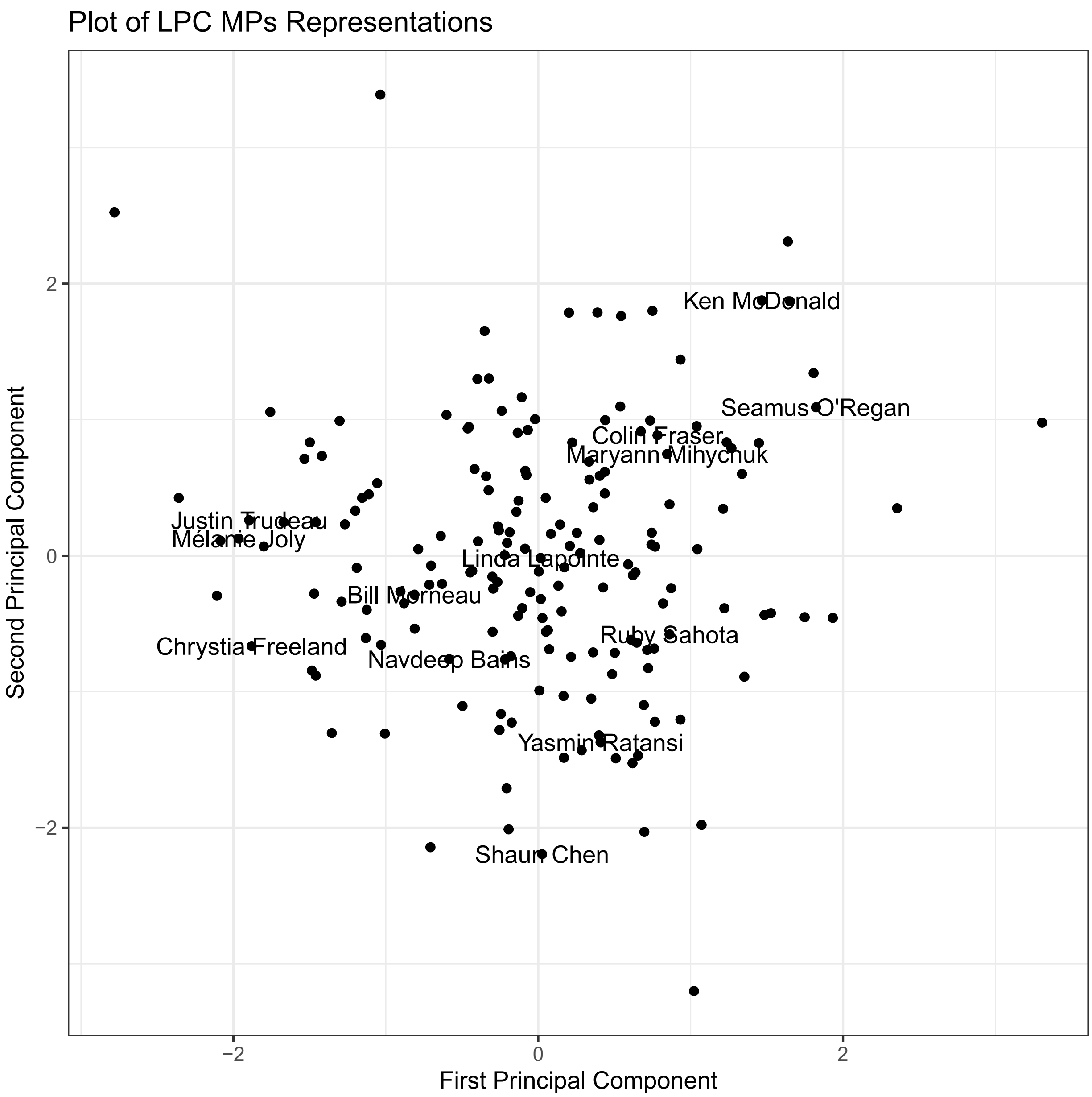


Figure 4: LPC MPs projected in 2 dimensional space.

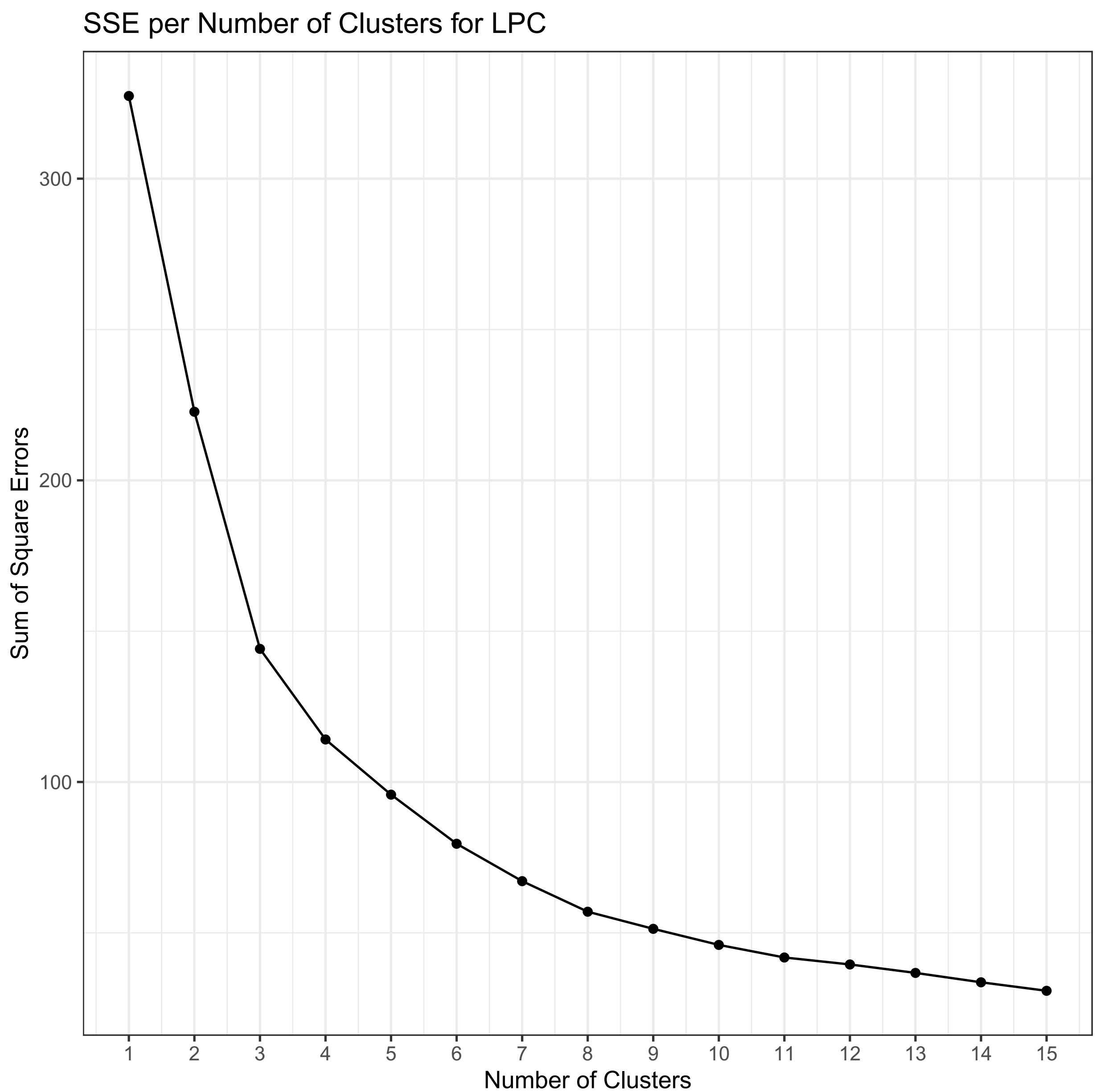


Figure 5: Reduction in the SSE per K-Cluster.

method" (Thorndike, 1953). This method relies on clustering at different numbers of k and calculating the sum of squared errors (SSE) between each point and its cluster center. The k value retained is a value after which the SSE decreases less sharply when adding more clusters.

Figure 5 shows the elbow plot for the PLC. There does not appear to be a single number where the decrease in SSE suddenly becomes very small, or in other words, a very clear "elbow." The reduction in SSE when incrementing K breaks sharply when k is equal to 3, and stalls when k is has a value of about 7. To favor a parsimonious model of the LPC, I propose 3, the smaller value, as the number of k-groups to analyze.

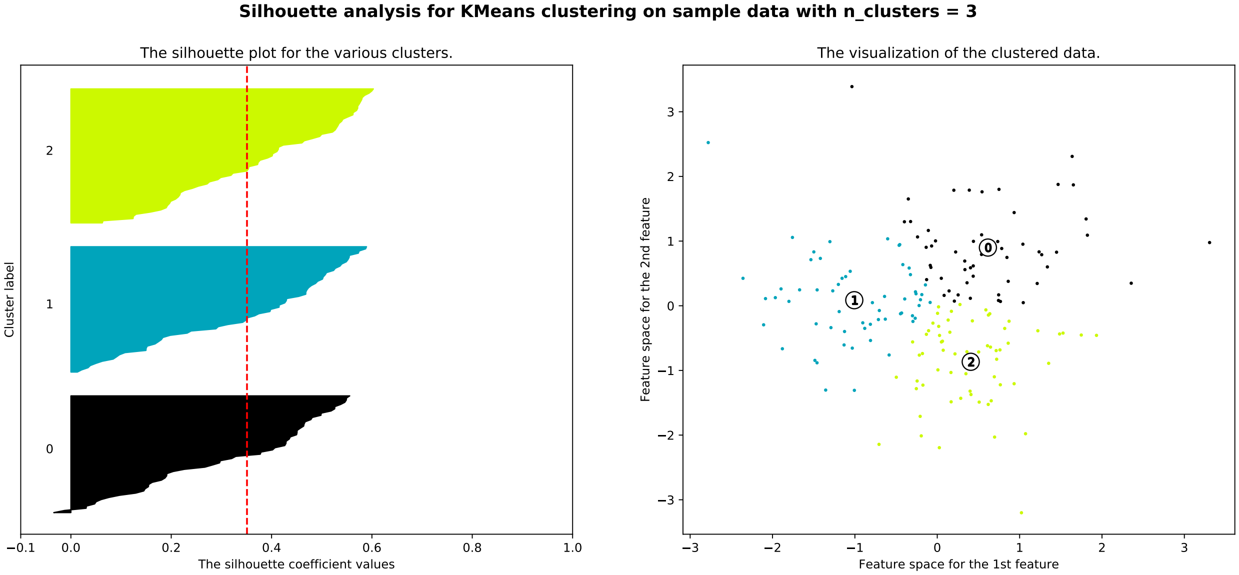
I use a second validation method to insure that 3 is a valid number of clusters for this study: silhouette analysis. Silhouette analysis is a method of validating the quality of a clustering. The silhouette value for a given datapoint i in cluster C_k is a function of its similarity to points in the same cluster and dissimilarity to points in the neighboring cluster, defined as the cluster which i is not a part of whose points have the least mean distance to i. A negative silhouette value therefore indicates that i resembles a different cluster more closely than its own cluster. In such a case the clustering is of poor quality. On the contrary, higher values indicate that the clustering is of good quality.

The two highest mean silhouette scores were obtained when clustering with 3 and 7 groups. Figure 6 shows the result of silhouette analysis. When K is equal to 3, the thickness of the silhouette plots are balanced, indicating roughly equal sized groups. Furthermore there are few observations with negative values compared to the silhouette analysis when K is equal to 7. Finally, when K is equal to 3, no cluster is entirely found under the overall mean silhouette value, such as group 6 when k is equal to 7. Silhouette analysis therefore confirms that 3 is a valid number of clusters to use in the analysis.

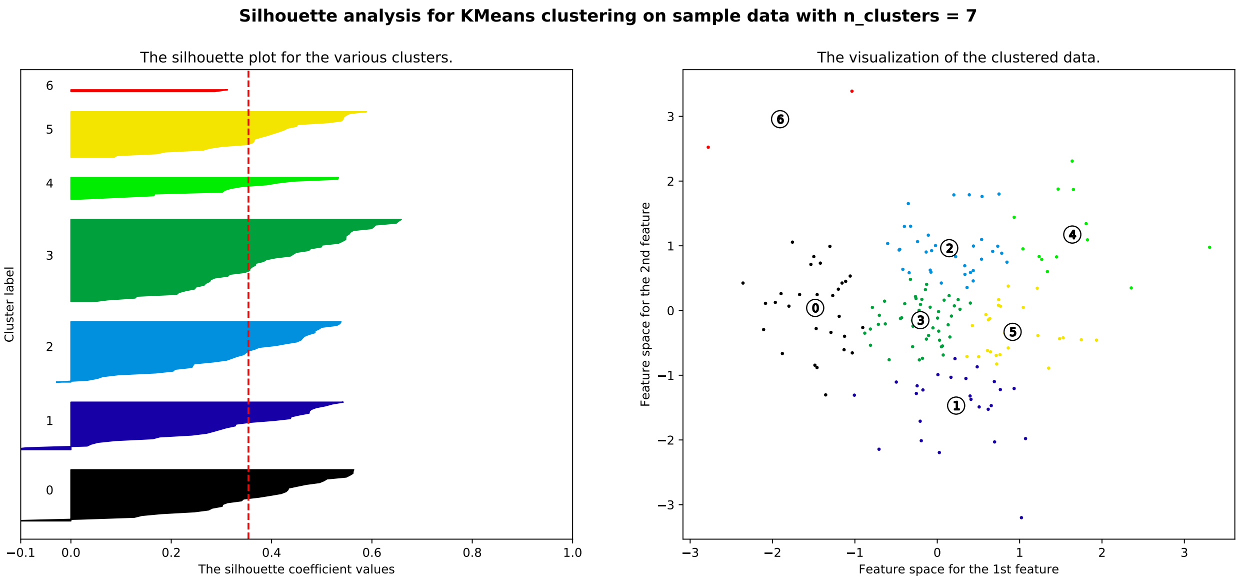
The above evidence suggests that 3 groups is a suitable number of clusters for the LPC. Figure 7 shows the resulting clustering in 2 dimensions. The following section analyzes the composition of the groups.

4.3 Determinants of Same Group Membership

The politician representations were created by analyzing the textual content they produced and their social interactions. In the current section I assess whether MPs who belong to the same detected groups share demographic characteristics. I use the detected clusters as an outcome and analyze them using standard parametric techniques. I assess various theoretically informed expectations with regards to factors that may explain why



(a) Plot for three clusters



(b) Plot for seven clusters

Figure 6: The silhouette method for 3-7 clusters.

parliamentarians tend to associate with one another . Note that the analysis here is only observational and that I do not make claims regarding causality when examining these factors.

To find possibly important predictors of why MPs are part of the same group, I apply multinomial logistic regression, a regression technique that makes it possible to analyze

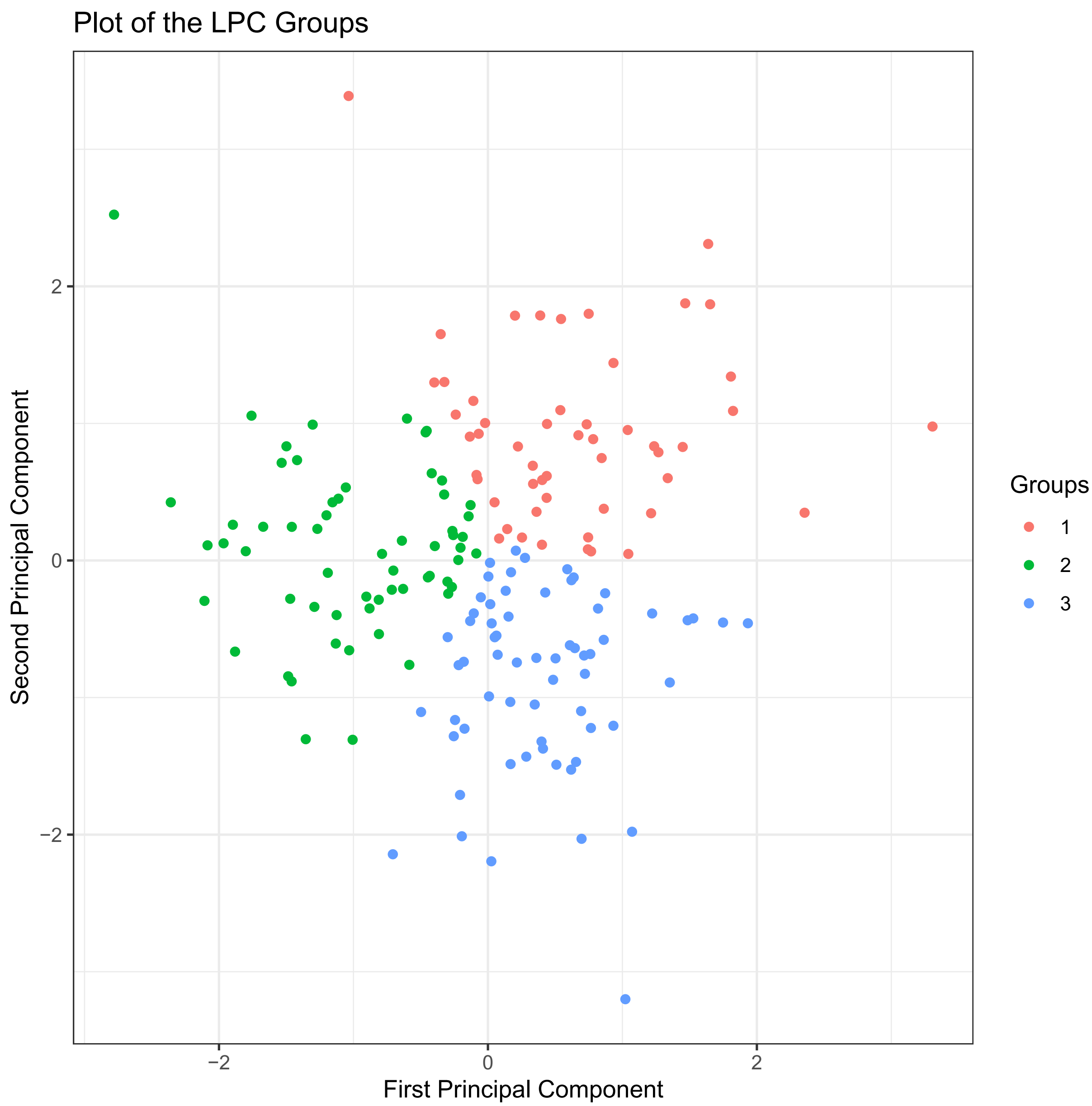


Figure 7: LPC Groups Visualized.

categorical outcome variables with multiple classes. I regress this variable on a set of variables that the intra-party politics literature has found to be important to explaining why politicians tend to associate.

1. Woman MP: A dummy variable coded 1 if the MP is a woman. Some scholars argue that increased representation of women is having an effect on intra-party politics, notably in terms of how women legislators can affect policy making within the party (Erzeel and Celis, 2016), or the internal party structure (Stockemer and Sundström, 2019). This variable tests for whether women MPs are more likely to be found in one of the groups.

2. Region: A variable that indicates which region the MP is from. Part of the literature on intra-party politics proposes that MPs tend to form groups to bring benefits to local constituents (Croissant and Chambers, 2010; Slider, 2010). To test for this effect I coded each MP as being part of one of Canada's traditional region: Quebec, Ontario, The Atlantic, The Prairies and British-Columbia. [9]

3. Age: Considering that the data comes from social media, I control for age to insure that observed representation proximity is not simply a factor of generational differences in social media usage. It also assesses a theoretical argument made in the literature, which is that young parliamentarians tend to share some policy interests, especially with regards to topics that are of concern for younger people, such as college education (Stockemer and Sundström, 2019).

4. Minority: A dummy variable coded 1 if the MPs is a visible minority. Previous research shows evidence that MPs from minority groups tend to propose more legislation related to minority rights (Mügge et al., 2019; Saalfeld, 2011). I test for whether minority MPs tend to associate more on the basis of a shared minority status.

[9] I excluded the 3 MPs from the territories from this analysis due to there being too few of them to draw meaningful conclusions.

5. Minister: A dummy variable coded 1 if the MP is or has been a minister in Justin Trudeau's cabinet. This variable tests for the presence of a leadership group in the party.

6. Ideology Score: The Wordfish measure of latent ideology applied the MP's tweets. The intra-party politics literature treats ideological proximity as an important factor in determining whether same party MPs tend to associate. This variable both tests this assumption and ensures that the other factors I test for are not simply explained by MP ideology. Here latent ideology was estimated by Wordfish. The Wordfish algorithm applied to a group of documents gives them a score of -1 to 1, ranging from ideologically-left to ideologically-right, based on differences in the words found in each document.

7. Rural MP: A dummy variable coded 1 if the MP is from a rural circonscription. I define Rural circonscriptions as circonscriptions with under average population density. Prior research shows that rural politicians can associate on the basis of shared rural interests, especially with regards to agricultural policy (Bryan, 2019). With this covariate I assess whether some form of "rural power" is found in the LPC.

It is difficult to succintly interpret the result of multinomial regression coefficients, since they represent log-odds ratios for each category with regards to a reference category. I use simulated probabilities for each of the variables to observe which socio-demographic factors are related to membership in a group.

Regionalism Matters...

MPs' regions correlate importantly with the group to which they belong, even controlling for other factors. Figure 8 shows simulated first differences from the previous model, when a MP's region changes from the Atlantic to any other major region.

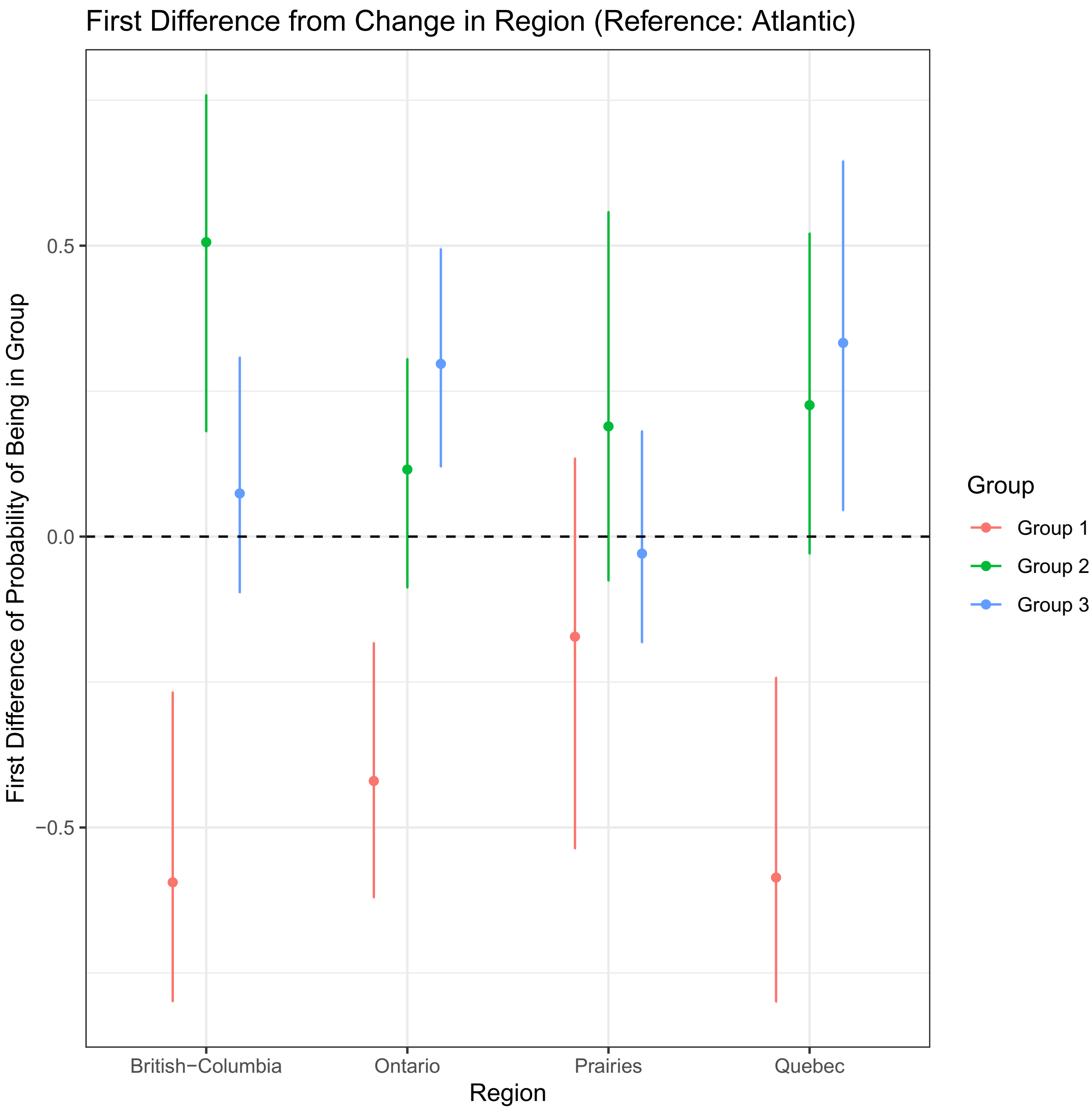


Figure 8: First difference of MP's region changing from Atlantic to another region.

Being in a non-atlantic region decreases the probability of belonging to Group 1 substantially, indicating that MPs from the Atlantic region are much more likely to be found in group 1. The MP's region changing from the Atlantic to Ontario increases the probability of belonging to Group 3. The MP's region changing to Quebec predicts an increase in the probability of belonging to either Group 2 or 3. MPs from British-Columbia see an important increase in their probability of belonging to Group 2. Finally, there is no statistically significant effect in the first difference in the probability of belonging to any of the groups when the MPs region changes from the Atlantic to the Prairies. The results indicate that regional cliques such as those found in different national contexts, as observed in Croissant and Chambers (2010) and Slider (2010), may be present within the LPC, especially with regards to MPs from Atlantic Canada.

Regionalism can also be observed in Figure 9, which shows the projection of the LPC's representation, with MPs color-coded by region. Broadly speaking, MPs from Ontario dominate the bottom part of this plot, MPs from Quebec and BC the left-side of the plot, and MPs from the Atlantic the top of the plot. Figure 9 gives a clear visualization of how regionalism presents itself in the LPC MP's online interactions.

... But so does Ideology

MP ideology is also an important predictor of which group an MP will appear in. I analyze this relationship using the simulated probability of belonging to each group at a given ideological scaling value. Figure 10 shows simulated probabilities of belonging to the same group as a function of ideological distance. A MP's latent ideology score does not appear to be predictive of whether the MP will belong to Group 1. However, a simulated increase in the ideology score is highly predictive of a MP appearing in Group 2. MPs with the lowest latent ideology scores are simulated to appear in Group 2 at an under 15% probability while MPs with the highest scores are predicted to have a near 50% probability of appearing in it. The inverse relationship is seen in Group 3, where MPs'

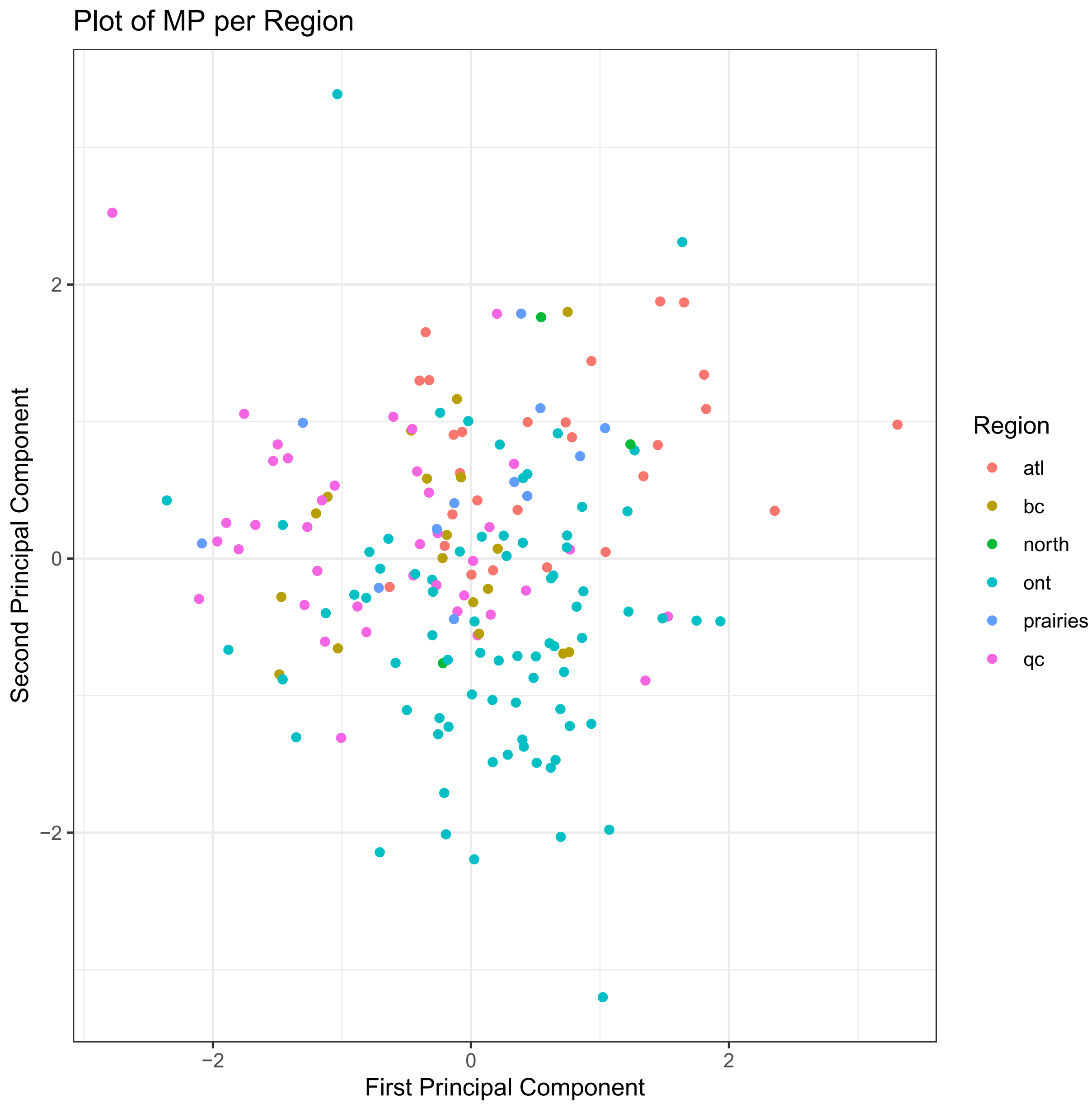


Figure 9: MPs per region in 2d space.

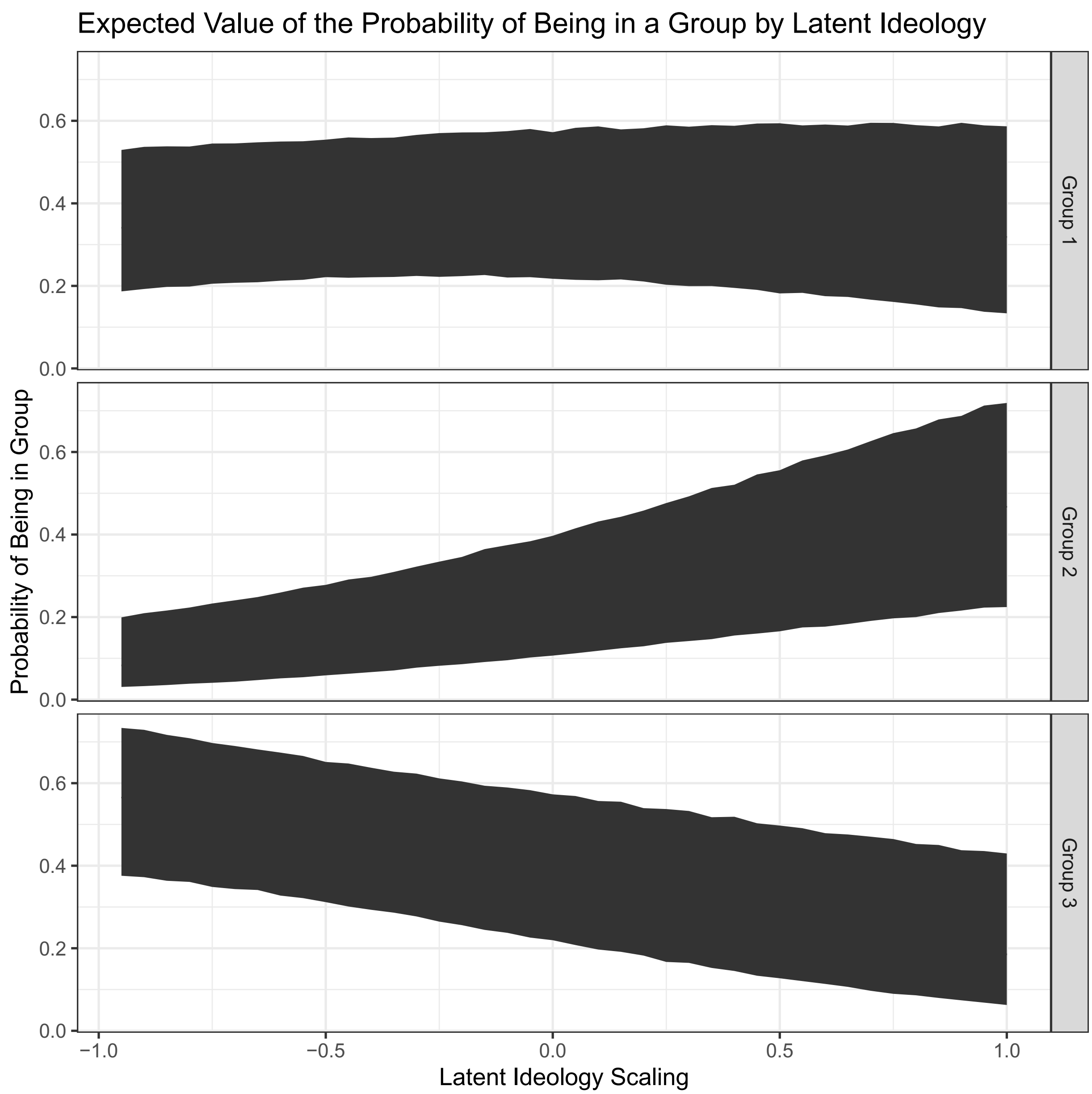


Figure 10: As latent ideology score increases MPs become more likely to be part of Group 2.

simulated probability of appearing in it decreases drastically as their latent ideological scaling score increases.

Minority Dynamics

I also find that a MP being a visible minority is associated with the group to which the MP belongs. Figure 15 shows the simulated first difference in the probability of belonging to a group when the MP is a minority. Being a minority MP increases the simulated probability of being in Group 3 by over 20%. Even controlling for other factors, minority MPs have a tendency to share similar interactions online, which leads them to be found together in Group 3. I show the 2d projection of the LPC with minority MPs color coded in figure 12. Minority MPs are mostly found in the center and bottom part of the plot. Figure 12 brings evidence that minority MPs within the LPC are more socially proximate.

Rural and Urban MPs

I also observe some level of association between a MP being from a rural region and their group membership. Figure 13 show the first difference in the probability of belonging to each group when the MP's region type changes from urban to rural. No first difference is statistically significant at the 95% level of confidence. However, there is still a sizeable 16% increase in the predicted probability of belonging to group 1 if a MP is from a rural region, a near statistically significant change. The probability of belong to group 3 decreases by a similar amount. Group 2 seems unrelated to the MP's circonscription type.

MP Gender

I observe whether MPs of the same gender tend to be more represented in the same group. Figure 14 shows the first difference in the probability of belonging to a group when the MP's gender goes from male to female. Overall, MP gender does not seem related to membership in any of the groups. No first difference is statistically significant and none of the effect size is important.

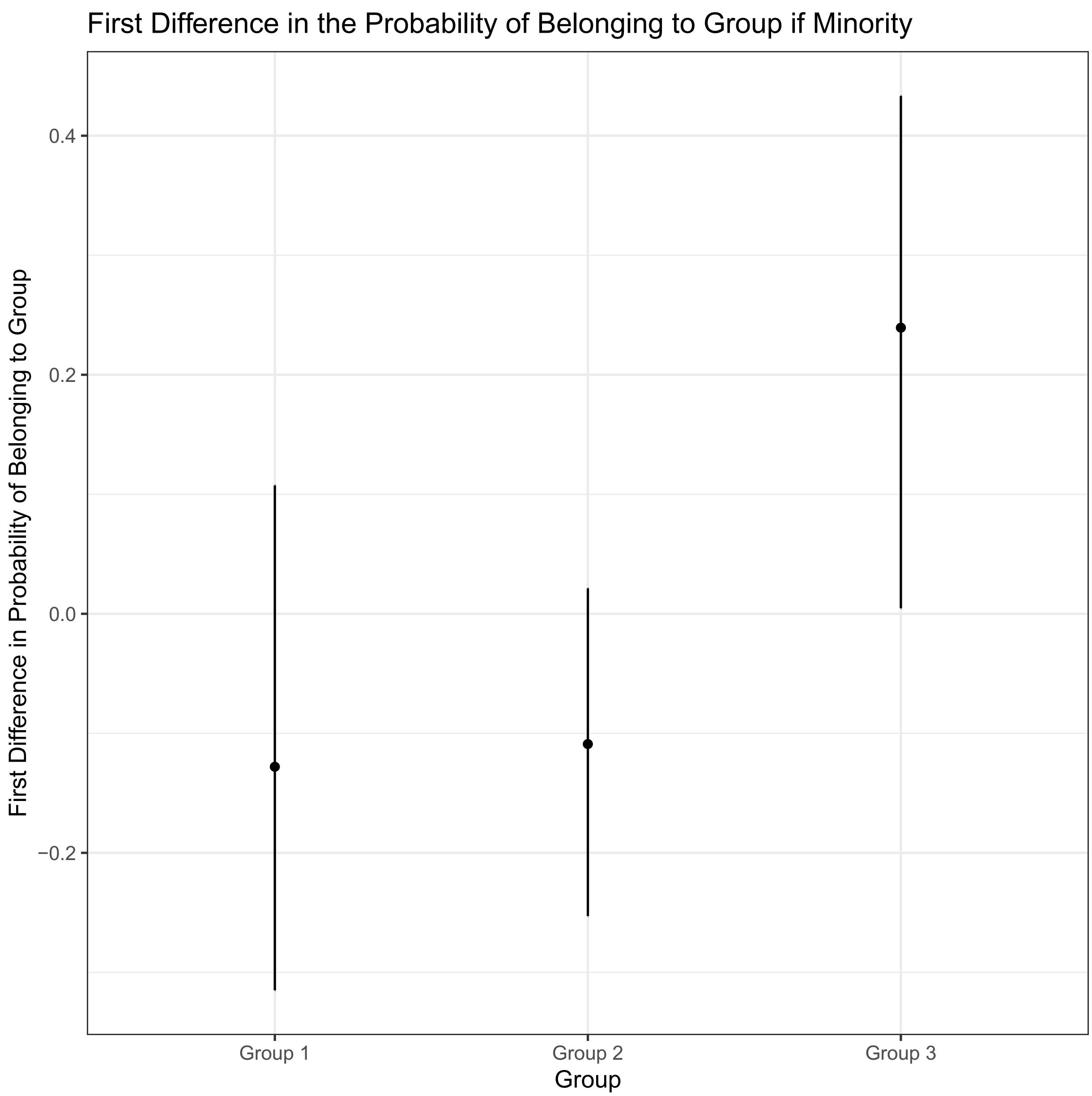


Figure 11: First difference of the probability of appearing in each group based on whether the MP is a visible minority.

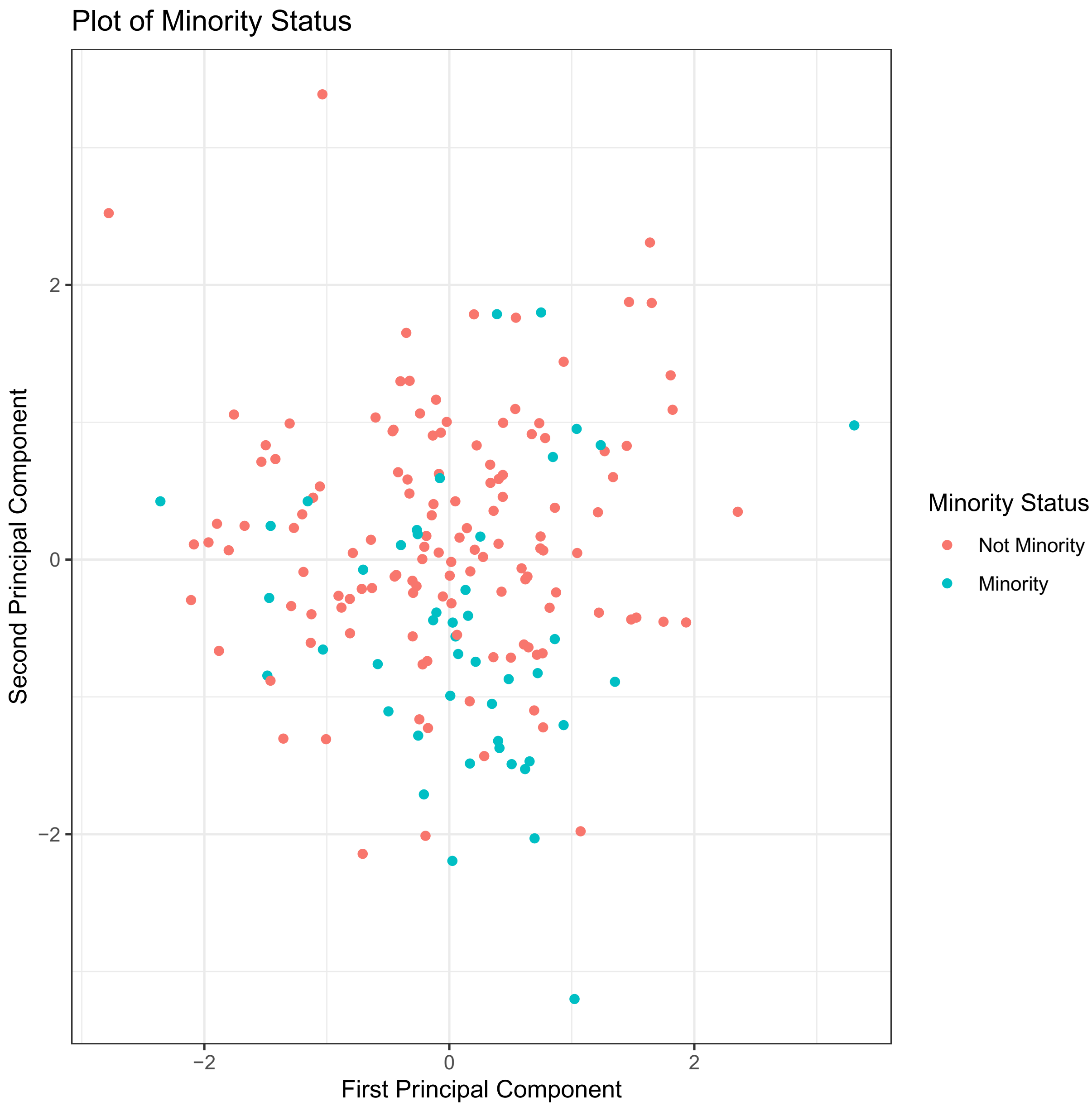


Figure 12: MP color coded by minority status.

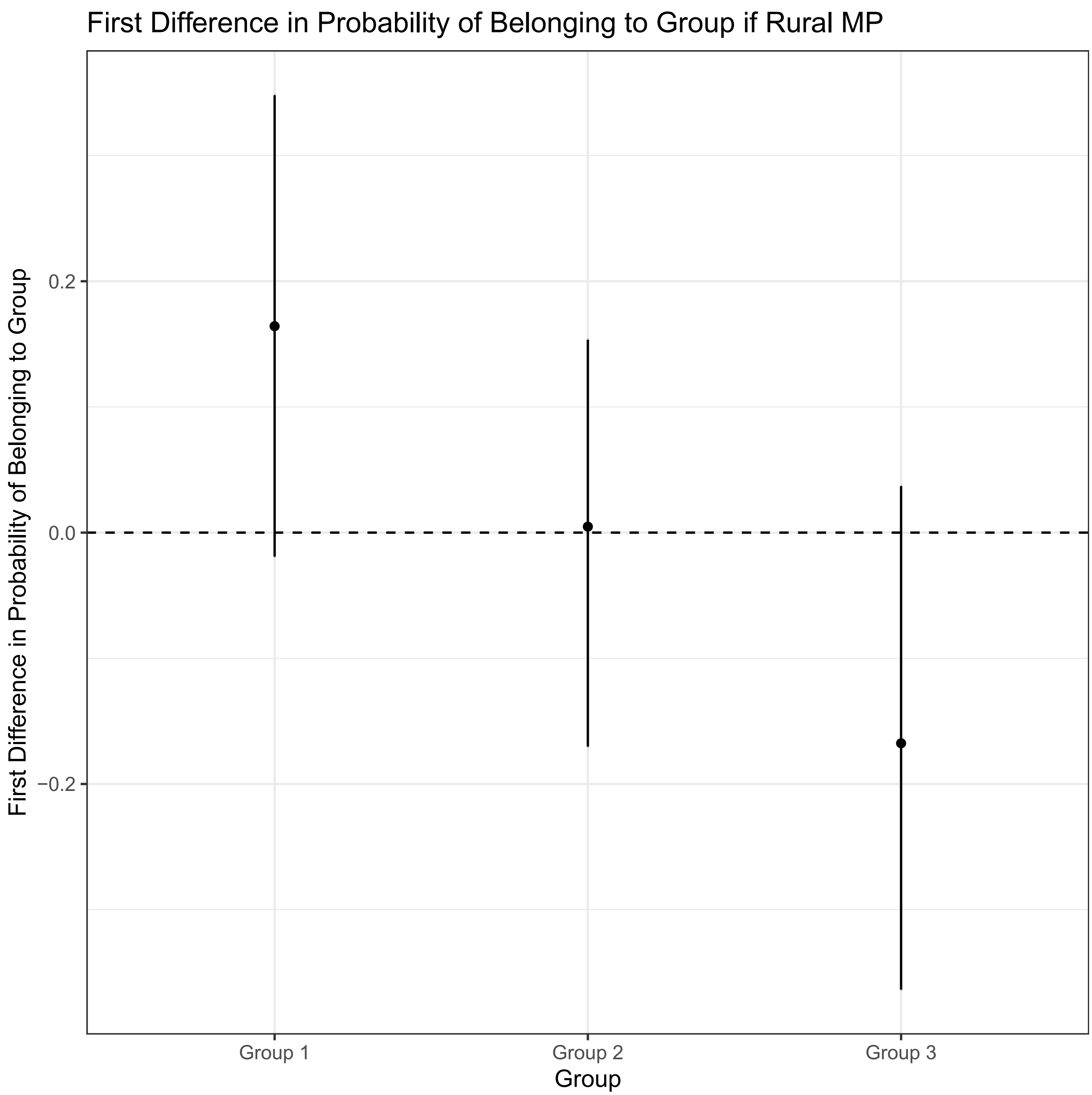


Figure 13: First Difference when MP circonscription type goes from urban to rural.

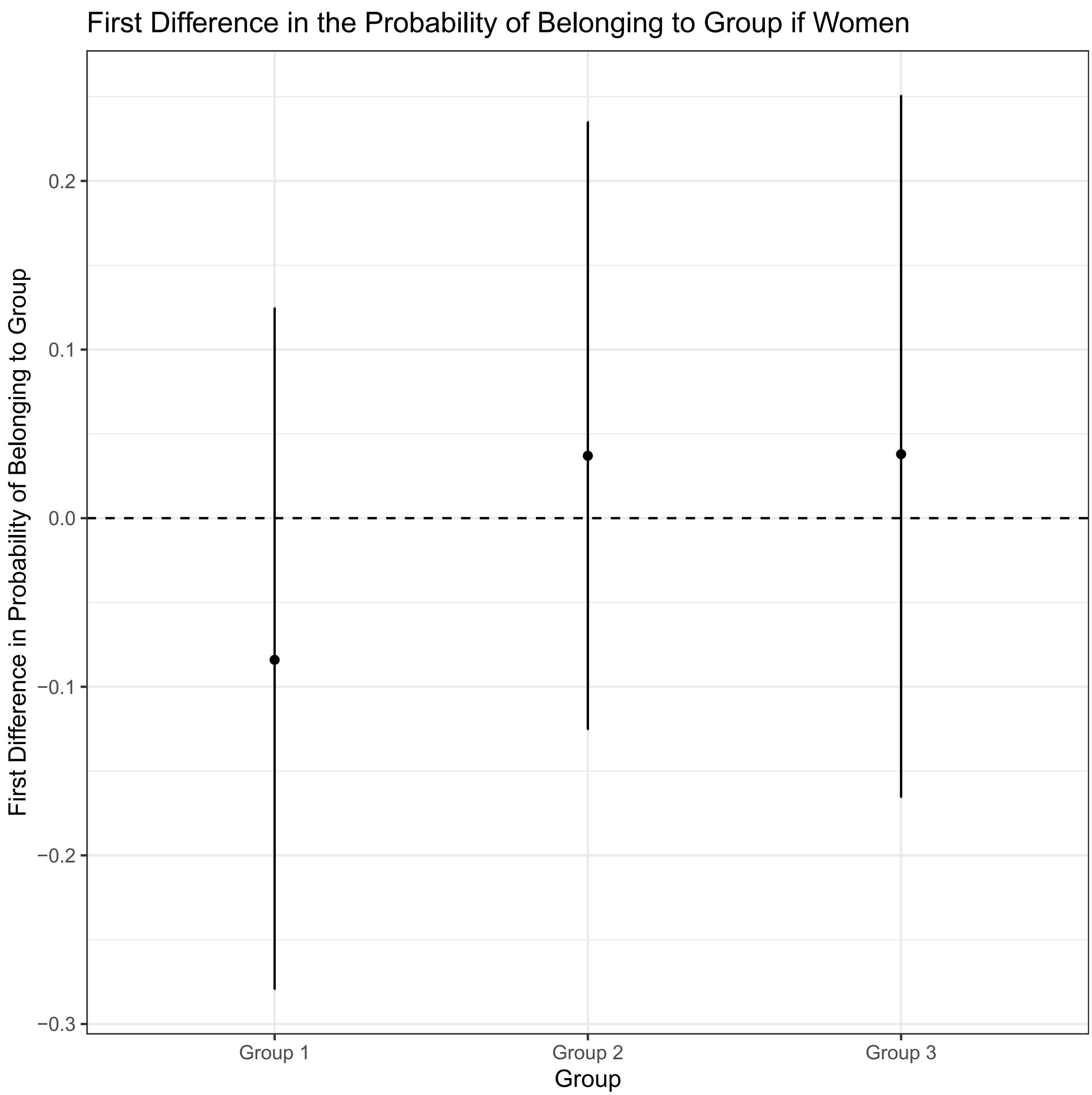


Figure 14: First Difference when MP gender goes from Male to Female.

Minister

Is being a minister in the Trudeau cabinet associated with membership in one of the groups? In other words, does one of the groups represent a leadership cluster in the LPC? Figure 15 shows the first difference when MP status changes from backbencher to minister. A MP being a minister does not predict a statistically significant change in the probability of belonging to any of the groups. The increase in the probability of belonging to group 2 is nevertheless quite important and nearly statistically significant. Group 2 is the group that contains Justin Trudeau. There therefore appears to be a tendency for ministers to be closer to Justin Trudeau's group.

Overall I find that three covariates are especially predictive of which group MPs will belong to. The MP's region is highly predictive of which group he will belong to. MP region establishes an especially strong relationship between being a MP from Atlantic Canada and belonging to Group 1. The MPs' estimated ideology is another strong predictor. A higher ideology score predicts MPs to be in Group 2 and a lower ideology score predicts them to be in Group 3. Minister Status and whether a MP is from a urban or rural region are also appear to influence group membership, though statistical significance is hard to establish. Finally, being a visible minority predicts MPs to be more likely to be in Group 3. How to interpret the principal components extracted from the data? The extracted dimensions are difficult to succintly summarize. The first principal component runs from right to left with negative values expressing closer proximity to the party leadership and MPs from Quebec while positive values express proximity to Atlantic provinces. The first component therefore summarizes regional separation between the MPs. The second principal component runs top to bottom, with higher values indicating MPs from a more rural, caucasian background while negative values associated to MPs from urban communities and ethnic minorities.

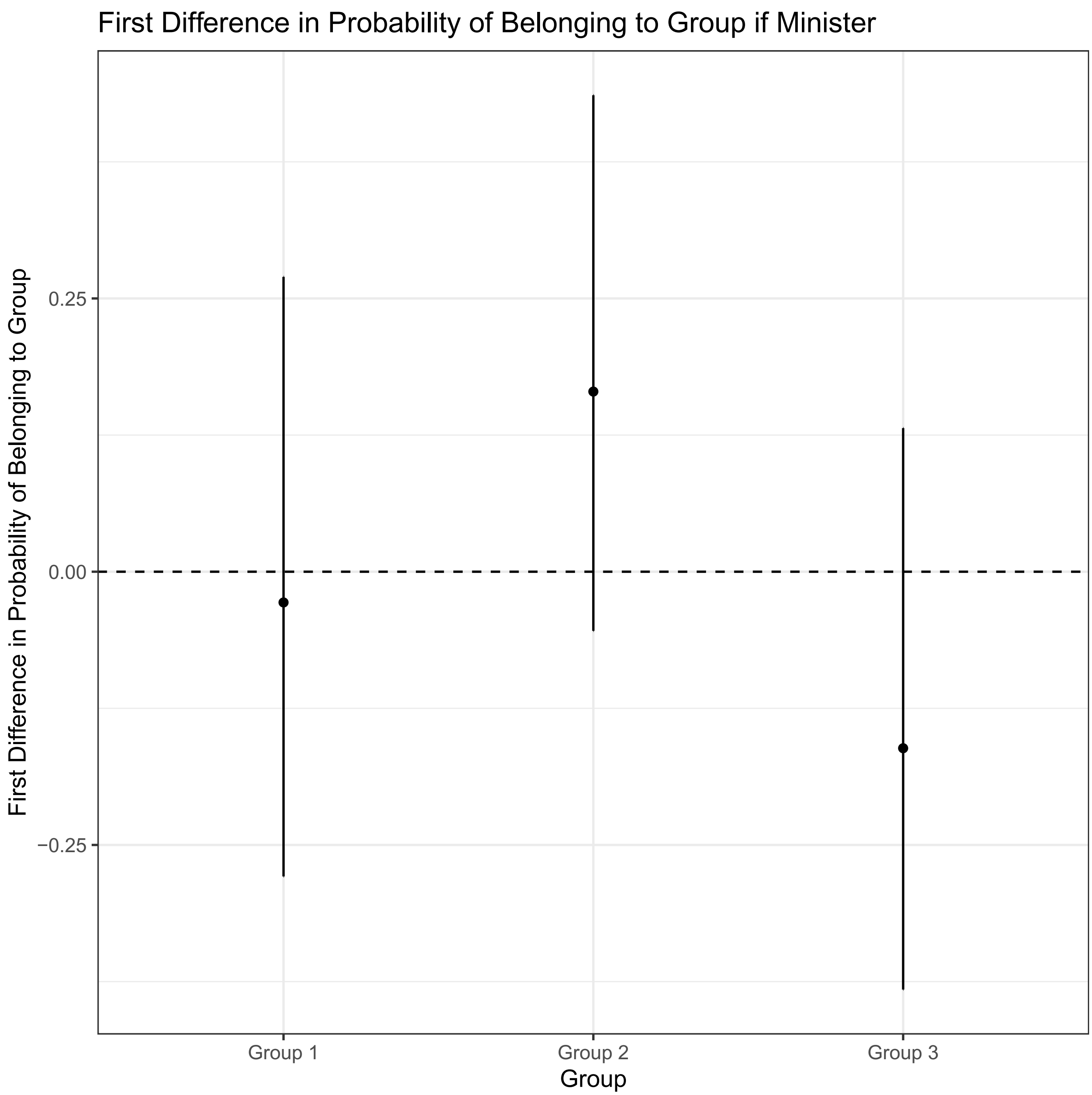


Figure 15: First Difference when MP status changes to Minister.

4.4 Topics

Are there differences between the groups in terms of the topics that they discuss on social media? I assess this using unsupervised topic modelling techniques. Unsupervised topic modelling aims at extracting a set of relevant topics from a body of documents. I separate the tweets of every group and apply Latent Dirichlet Allocation (LDA), to the tweets of every group. LDA is a technique that works best on larger documents, since it assumes that each documents contains a mixture of topics. Tweets from each group are concatenated on the basis of the day they were posted at. This type of tweet aggregation gives a larger document which improves the output of the LDA model (Hong and Davidson, 2010). LDA is used to obtain a set of 110 topics [10] discussed by MPs from each group. Twitter data concatenated by day often contains many topics related to time-specific events [11] that have little political interest. I follow the methodology from (Barberá et al., 2019), and qualitatively assess which topics are relevant to the current analysis. Results are shown in table 2.

There is some distinction in the topics found in the different group's tweets. In Group 1, topics are mainly focused on the economy (eg: International trade, budget, loans), the Canadian health system (eg: drugs policy, health) as well as issues that are distinct to Atlantic Canada, such as fishing and fisheries or topics related to Acadians. This is not surprising considering that MPs from Atlantic Canada form a sizeable portion of this group's membership.

Group 2, in comparison, exhibits many topics related to minority issues. This includes the rights of minorities, first nations, as well as topics related to homophobia and islamophobia. Interestingly enough, in spite of group 2's tweets containing topics related to minority issues the most, more minority MPs are actually found in Group 3, not Group 2.

Finally, Group 3 exhibits a variety of topics, some related to the economy (eg: energy policy, steel industry), issues pertaining to minority groups and other topics, such as the

[10] LDA was tested for varying numbers of topics and 110 was selected as number that produced the most coherent topics.

[11] eg: St-Patrick's Day, Christmas, etc.

Group 1		Group 2		Group 3	
Topic ID	Topic	Topic ID	Topic	Topic ID	Topic
1	International trade	1	Marijuana legalization	4	Energy policy
9	Budget	7	Minority issues	7	Abortion
10	Parental support	11	House ownership	8	Minority issues
11	Student Loans	14	Transgender issues	14	Foreign policy
24	Health	17	First nations	24	Steel industry
25	Women issues	19	Prescription drugs policy	29	Minority issues
26	Marijuana legalization	20	Economic issues	35	New Zealand mosque shooting
30	Infrastructures	25	Health issues	36	Firearms policy
44	Veteran celebration	34	Racism	51	House ownership
46	Acadian issues	36	Islamophobia	63	Tourism and immigration
47	Health	43	New Zealand mosque shooting	75	Mental illness
50	Firearms policy	45	Homophobia	94	Prescription drugs policy
55	Social issues	47	Firearms policy	100	Environment
58	Veteran celebration	52	Minority issues	101	Women in science
60	Fishing and fisheries	56	Elderly issues	108	Holocaust remembrance
71	Canadian industries	66	Infrastructure		
72	Prescription drugs policy	77	Holocaust remembrance		
75	Prescription drugs policy	80	House ownership		
76	Acadian issues	96	Health issues		
83	Fishing and fisheries	101	Abortion		
88	Iqualuit flood				
97	Minority issues				
102	Wildlife conservation				
103	Fishing and fisheries				
105	Fishing and fisheries				

Table 2: Main topics found in each group's tweets.

environment or the role of women in science. Overall, Group 3's content has a less clear orientation than the content produced by Group 1 and 2.

The current section analyzed the representations of the LPC. I find that embedding proximity is largely predicted by the region in which the MP is elected and their ideology, although other factors such whether the MP is part of a minority group and whether the MP is a minister in the Trudeau government may also be associated with group membership. Factors such as whether the MP's gender or whether the MP is from a rural circonscription, however, do not predict same group membership. I also find differences in the content produced by the different groups.

5 Validation

Do multi-view representations accurately represent proximity between Canadian MPs? For MPs' online behavior to be a valid measure of the MPs social network, the groups detected in section 4 should relate to parliamentarians' offline behavior. As mentioned in section 2, it is very difficult to find observable manifestations of intra-party politics, since intra-party networks are typically not formalized, or even organized. I propose the network of lobbyist-MP interactions in the parliament as a real-life phenomenon that is indicative of proximity between MPs, and to which I can relate multi-view representations. I first build a network representation of the liberal MPs, where MPs are related if they often interact with the same lobbyists. I will then observe whether the groups I detected in section 4 correlate to this network. The assumption behind this test is that both MP-Lobbyist interactions and social media representations are influenced by the social structure of the party. If a correlation is detected between these two networks, it shows that online interactions are indicative of offline relationships between politicians. I compare the performance of the groups created through multi-view representation with other representation methods currently in use in the literature, wordfish scores and guided ideology estimates.

5.1 Lobbying: A valid Proxy

Why might lobbyist-MP interactions be informative with regards to intra-party politics? An important literature assesses the motivations of lobbyists when arranging meetings with legislators. A developing consensus is that lobbyists tend to target ”ally” MPs, who are favorable to their policy objectives (de Figueiredo and Richter, 2014). Lobbyists are “budget-driven.“ They have limited ressources and access to MPs is costly. For this reason, the most effective use of ressources for lobbyists is to access “ally“ MPs, who they can easily meet with, and provide them with expertise regarding an issue. Lobbyists act as informational ressource providers, that give MPs policy expertise necessary to convince other same party MPs of a policy orientation (Hall and Deardorff, 2006; Hall and Miller, 2008). On the contrary, accessing MPs not already aligned with the lobbyist's interest is costly, and it is difficult to persuade them (Schnakenberg, 2017). Access to ally legislators is further fuelled by "revolving door lobbyists," which refers to ex-staff members of politicians who leverage their personal connections to gain access to their previous employer (Blanes i Vidal et al., 2012; LaPira and Thomas, 2014; Tyllström, 2019). MPs often contacted by the same lobbyists are therefore likely to share characteristics and policy preferences with one another. This is true of the Canadian parliament, where previous research finds evidence of network homophily in the network of MP and lobbyists interactions. Network homophily designates a network where individuals who share characteristics are more likely to be connected (Hopkins, 2019). Another recent study finds that lobbyists in the Canadian parliament often leverage personal relationships with politicians to access them (Boucher and Cooper, 2019). Recent research on intra-party politics identifies the behavior of lobbyists as a promising way of observing intra-party politics in contexts where little data is available on intra-party dynamics (Yadav, 2017). Typical research in this approach leverages social network analysis techniques to correlate shared ties with lobbyists and MP behavior. A notable finding is that MPs who are contacted by the same lobbyists are more likely to co-sponsor bills in parliament (Fischer, 2019).

Lobbyist-MP interactions is a complex phenomenon. Nevertheless, there is considerable evidence that lobbyist and MP interactions are indicative of intra-party relationships. I make the claim that the lobbyist-MP network is a valid way to observe proximity between politicians offline, and that it can be used to validate whether multi-view representations created from online content correlate to MPs' offline behavior. The argument is summarized in Figure 16.

5.2 Validation Method

I leverage data on lobbyist MP encounter registered in the Canadian "Registry of Lobbyists," the record of meetings between paid lobbyists and certain designated public office

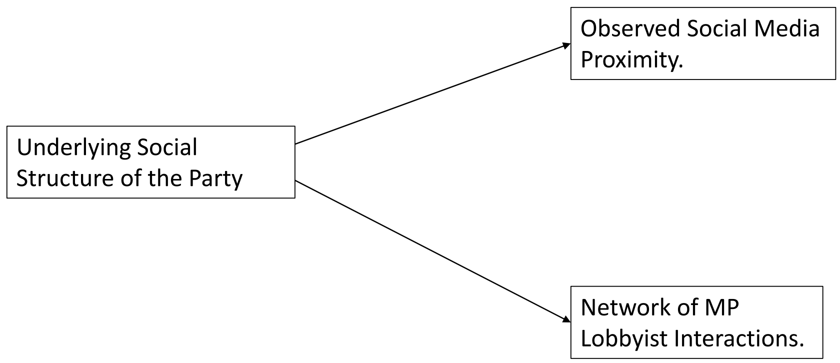


Figure 16: The argument of Section 5. In so far as there is no causation between observed social media proximity and the network of MP and lobbyist interactions, I can attribute correlation between the two phenomena to the influence of the underlying social structure of the Party. Detecting a positive relationship then indicates that observed social media proximity is indicative of that social structure.

holders, a category which includes Canadian MPs. I gathered all encounters between lobbyists and LPC MPs from the beginning of 2018 to the 2019 election. This time period matches the time period during which the tweets were collected. I found approximately $10,000$ lobbyist MP interactions. Most lobbyists only have a single visit to a MP in the time period. To better capture the idea of ally MP that is prominent in the literature on MP-Lobbyist interactions, I only observe Lobbyist-MP pairs that have met twice or more, leaving about $1,000$ MP-lobbyist pairs. The data is organized into a bi-graph, where nodes of type "lobbyist" and nodes of type "MP" are linked if there are registered encounters between them. A projection of the bi-graph where MP nodes have an edge if they are both linked to the same lobbyists in the bi-graph is created. This projection is the network of interest, since it relates MPs based on whether they have same lobbyist interactions.

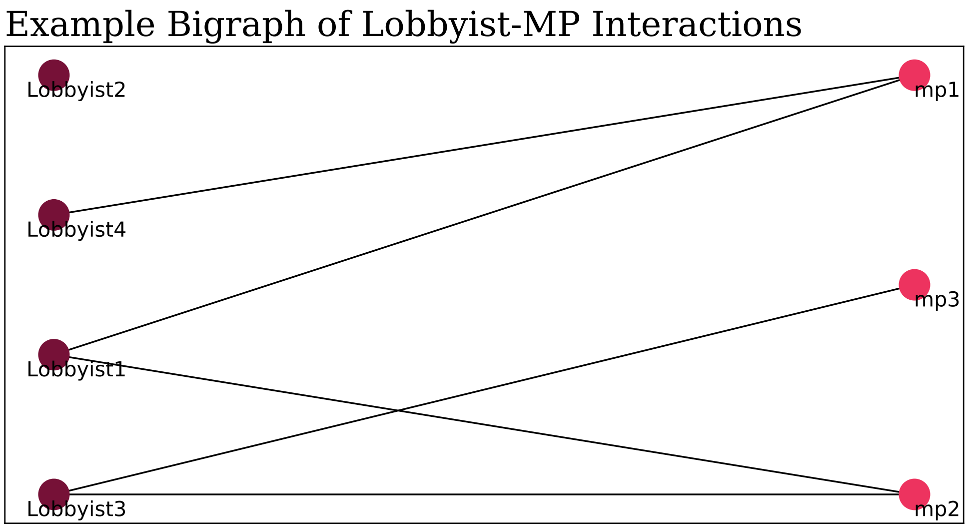


Figure 17: Example bigraph of MP-Lobbyist interactions. MP1 and MP2 are connected through Lobbyist1. MP2 and MP3 are connected through Lobbyist3.

An example bigraph of MP-Lobbyist interactions is shown in Figure 17 and the resulting projected graph of MP connections is shown in Figure 18. For the current analysis the graph of interest is the graph of Figure 18.

I give the group attributed by multi-view representations as a node attribute to each MP nodes. I then calculate the attribute assortativity coefficient of the network, a statistic that indicates to what extent nodes with a given attribute in a network tend to share links with one another (Newman, 2003). The attribute assortativity coefficient ranges from -1 to 1. A negative value indicates that nodes that share the same attribute have fewer links while a positive value indicates that nodes that share an attribute tend to share links. An example network with edges colored based on whether they are between same attribute nodes is shown in Figure 19. In other words, do MPs from the same group detected in section 4 tend to associate with the same lobbyists? I use Monte Carlo case resampling to assess whether the detected coefficient is statistically significant at conventional levels.

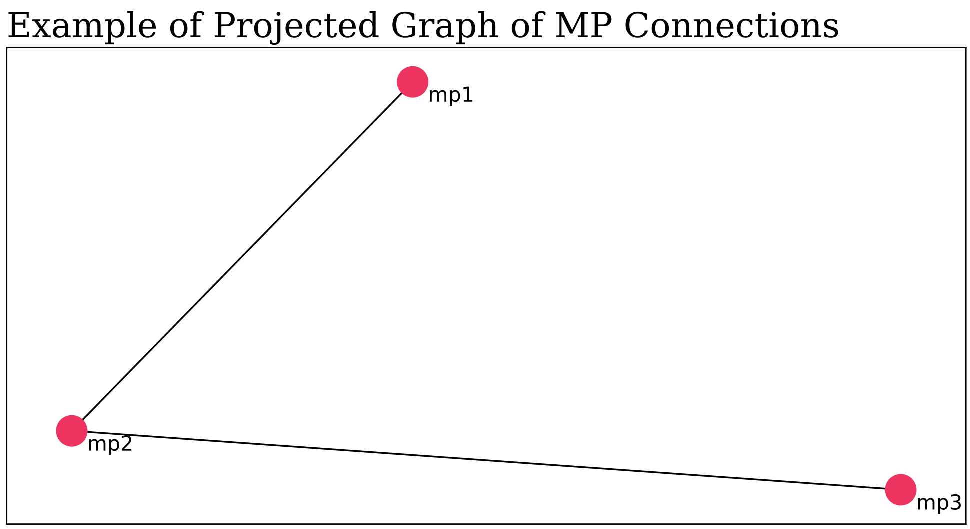


Figure 18: Projection Graph of MP connections based on figure 17. MPs who were connected to the same lobbyist are connected in the projection. This is the graph of interest for the validation test.

5.3 Comparison to Baseline Methods

Is there an added value to using the multi-view framework to represent politicians, when compared to typical representation methods currently used in the literature? I compare the performance of multi-view representations on the previous validation task to other methods of representing politicians. In section 2.1, I discussed an important literature on scaling the ideology of politicians. I compare the performance of multi-view representations to algorithms for the ideological scaling of politicians currently in use in the literature. This makes it possible to assess whether multi-view representation learning is a technique that offers advantages over currently employed methodologies in Political Science.

In a recent piece on observing intra-party politics on social-media, Ceron (2017) uses the wordfish algorithm to assign an ideology score to Italian MPs from a party, on the basis of the content of their social media posts. He finds that distance between these scalings is predictive of cohesive behavior between members of the same party, such as endorse-

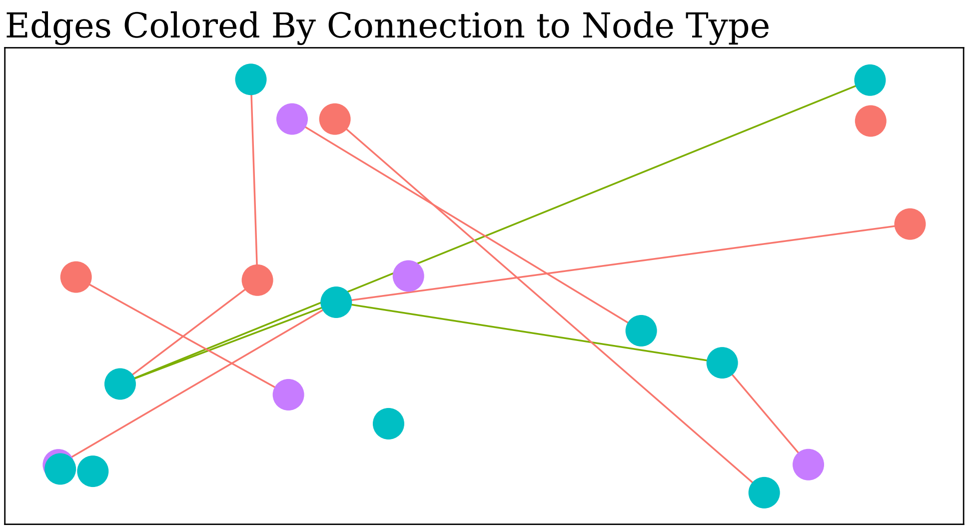


Figure 19: Example network with nodes colored by attribute. Green edges are shared by same attribute nodes and red edges are shared by nodes with a different attribute. The attribute assortativity coefficient is positive when the number of green edges is larger than you would probabilistically expect it to be with no association between attribute and edges.

ment of candidates in leadership races. I treat this method as a baseline to beat. I obtain wordfish scores based on the content produced by politicians in my dataset and compare their predictivity in the validation task to the multi-view representations. To obtain a discrete attribute that can be assigned to nodes and easily compared to the groups obtained through multi-view embeddings, I separate MPs in three groups based on whether their wordfish score is in the lower tertile, middle tertile or upper tertile of the distribution.

I also compare the performance of multi-view representations to "Guided Ideology Estimates." Guided ideology estimates are a measure of document ideology recently proposed in Rheault and Cochrane (2020), that relies on training a vector representation of a political document and calculating its cosine distance to a set of seed words representing left-wing and right-wing positions. I calculated guided ideological estimates for all the tweets and aggregated them per MP to obtain a guided ideology score. The seed words used to identify left-right positions is provided in Rheault and Cochrane (2020). Like for wordfish, I separate the MPs in three bins to obtain a discrete attribute.

5.4 Validation Performance

I observe the correlation between the group attributed to a node and the structure of the graph in Figure 20.

Attribute correlation coefficients with groups made through multi-view representations, wordfish and guided ideology are shown. Multi-view representation based groups have about a 0.1 attribute correlation with the structure of projected MP-lobbyist network. 95% bootstrapped confidence intervals are shown on the plot. The correlation coefficient on the multi-view representation grouping is easily statistically significant at conventional levels of significance, indicating a positive relationship between the detected groups and the MP-lobbyist networks. The baseline methods using groups made through wordfish and guided ideology estimates both have an attribute correlation coefficient of approximately 0.03 with the network. The estimates are statistically significant but multi-view representation based groups show a stronger relationship to the MP-lobbyist network's structure.

A correlation of 0.1 between representations of social media content posted by Members of Parliament and MP-Lobbyist interactions is a substantially important result that, I argue, can only be explained by the fact that both phenomena are related to the party's internal social structure. The current validation task gives strong evidence that multi-view representations built from social media content correlate to offline relationships between politicians. The validation task also shows they are more effective at doing so than alternative methods currently in use in the literature.

How does the presence of highly influential MPs in the sample affect the correlation between these networks? There is a possibility that the detected relationship is caused by lobbyists visiting especially influential MPs, such as cabinet members. I explore this question in Appendix 1.

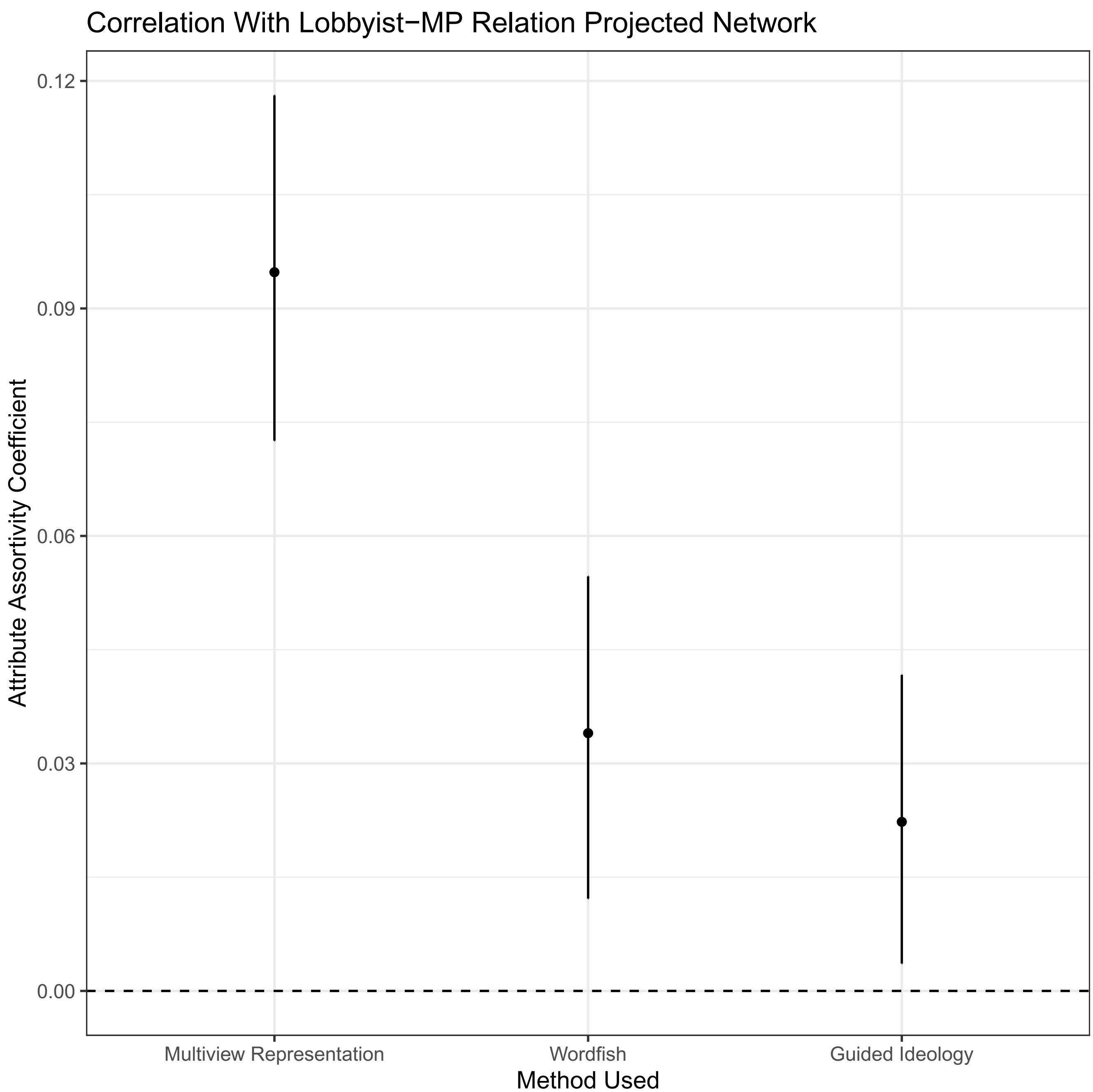


Figure 20: Performance of different methods on the validation task.

6 Discussion: Implications for Political Science

The current paper has many implications for the study of political science. In recent years, technical improvements in representation learning, mostly driven by progress in deep learning methods, have contributed to many important technological advances in computer science and engineering (Bengio et al., 2014). Political scientists are also starting to leverage representation learning to better represent social phenomena. Recently published articles in political methodology leverage Long-Short Term Neural Networks (LSTMs) to classify political text (Chang and Masterson, 2019), or use document-embeddings assign ideology scores to documents (Rheault and Cochrane, 2020). Multi-view representations can contribute to further improve the modelling of political content. I present the representation of social-media content as an especially important area where this metho can be applied, but applications are not limited to social media content. Many sources of data in political science are heterogenous. A concrete example is parliamentary speeches, where important information is available both in the content and in knowing the debate in which the speech is pronounced (Lauderdale and Herzog, 2016). The increasing importance of video (Qi et al., 2016) and audio data (Proksch et al., 2019) in applied political science research gives additional opportunities to leverage heterogenous data.

This article treated the multi-view representations as an outcome of interest and analyzed them using dimensionality reduction and various parametric statistical tests. Applications of representation learning often use low-dimensional vectors as inputs into a classifier that attempts to predict a certain feature of the data. Benton et al. (2016) and Ding et al. (2017) obtained impressive results when predicting social media user attributes using multi-view representation learning. Political scientists can use multi-view representations as inputs in a machine learning framework to improve predictions of political outcomes.

The current paper challenged assumptions regarding Canadian political science. Section 4 highlighted some intra-party dynamics of the Canadian LPC that have been ignored up to now in the literature on Canadian intra-party politics. Amongst important findings, minority MPs in the LPC are associated in their online behavior. While I do not currently make any claim regarding the importance of this phenomenon for Canadian politics, I believe that Canadian political scientists should pay attention to minority MPs as a parliamentary force. While regionalism is a well-studied phenomenon in Canadian political science, its effect on intra-party politics is not well developed, both theoretically and in terms of empirical studies. I find evidence that MPs interact on a regional basis within the Canadian LPC, especially with regards to MPs from the Atlantic region. This is another phenomenon that Canadian political scientists should consider in future studies. In general, strong party discipline in Canada gives the impression that its intra-party politics are top-down, directed by the leadership, and that Canadian parties are monolithic in nature. The current research challenges this assumption and invites more research on intra-party politics in Canada.

While the resulting representations accurately represent general relationships within the Canadian LPC, readers should avoid drawing conclusions about any individual MP on the basis of the results presented in this paper. The generated representations are an approximation based on twitter posts and interactions. The representations may not be perfectly accurate depictions of a given individual.

7 Conclusion

The current paper proposed multi-view representation learning as a method to improve the analysis of social media content. I used multi-view representation to jointly leverage social and textual data produced by Canadian MPs on social media to observe their social proximity. Overall, multi-view representation learning is found to be an effective tool for the study of political science. The multi-view representations allow us to explore the intra-party politics of the Liberal Party of Canada. I discover previously unexplored relationships, such as the fact that minority MPs are very proximate in the LPC. I find that multi-view representations created using online content correlate well to the offline network of lobbyist and MPs. Potential applications of this method are not limited to the study of intra-party politics. It could in fact be extended to any topic that benefits from the analysis of social media data, or more broadly heterogenous data.

I suggest that political scientists familiarize themselves with the multi-view learning framework. Multi-view representation learning is not limited to the GCCA algorithm employed in the current study. Many different techniques of relating heterogenous data exist and can bring important methodological contributions to the study of political science. In CCA inspired methods, Kernel-CCA (Akaho, 2007) or Deep CCA (Galen et al., 2013) extend the CCA framework to non-linear correlations, using respectively kernel smoothing methods and deep neural networks. Another family of multi-view learning algorithms relies on the principle of co-training . In co-training algorithms, a classifier trained on one view is used to predict the labels of the other view (Blum and Mitchell, 1998). The predicted labels are then used as part of an augmented training set for the other view. Co-training is useful in situations where there are few labelled samples and many unlabelled samples available. These different techniques can also be easily applied to political science research.

Bibliography

Adams, J., Ezrow, L., and Somer-Topcu, Z. (2011). Is Anybody Listening? Evidence That Voters Do Not Respond to European Parties' Policy Statements During Elections. *American Journal of Political Science*, 55(2):370–382.

Akaho, S. (2007). A kernel method for canonical correlation analysis. *arXiv:cs/0609071*. arXiv: cs/0609071.

Barberá, P. (2015). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis*, 23(1):76–91.

Barberá, P., Casas, A., Nagler, J., Egan, P. J., Bonneau, R., Jost, J. T., and Tucker, J. A. (2019). Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data. *American Political Science Review*, 113(4):883–901.

Bengio, Y., Courville, A., and Vincent, P. (2014). Representation Learning: A Review and New Perspectives. *arXiv:1206.5538 [cs]*. arXiv: 1206.5538.

Bengio, Y., Ducharme, R., Vincent, P., and Jauvin, C. (2003). A Neural Probabilistic Language Model. *Journal of Machine Learning Research*, pages 1137–1155.

Benton, A., Raman, A., and Dredze, M. (2016). Learning Multiview Embeddings of Twitter Users. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, volume 2, pages 14–19, Berlin, Germany. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.

Berthiaume, L. (2016). 'This isn't what Liberals are about': Some party members bristling over perceived efforts to quash dissent. *National Post*.

Bettcher, K. E. (2005). Factions of Interest in Japan and Italy: The Organizational and Motivational Dimensions of Factionalism. *Party Politics*, 11(3):339–358.

Blanes i Vidal, J., Draca, M., and Fons-Rosen, C. (2012). Revolving Door Lobbyists. *The American Economic Review*, 102(7):3731–3748.

Blum, A. and Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory.*

Boireau, M. (2014). Determining Political Stances from Twitter Timelines. In *Proceedings of the 2014 Conference on Electronic Governance and Open Society*, NY, USA. Association for Computing Machinery.

Bond, R. and Messing, S. (2015). Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook. *American Political Science Review*, 109(1):62–78.

Boucek, F. (2009). Rethinking Factionalism: Typologies, Intra-Party Dynamics and Three Faces of Factionalism. *Party Politics*, 15(4):455–485.

Boucher, M. and Cooper, C. A. (2019). Consultant Lobbyists and Public Officials: Selling Policy Expertise or Personal Connections in Canada? *Political Studies Review*, 17(4):340–359.

Bryan, F. M. (2019). *Politics In The Rural States : People, Parties, And Processes*. Routledge.

Carey, J. M. (2007). Competing Principals, Political Institutions, and Party Unity in Legislative Voting. *American Journal of Political Science*, 51(1):92–107.

Carrubba, C., Gabel, M., and Hug, S. (2008). Legislative Voting Behavior, Seen and Unseen: A Theory of Roll-Call Vote Selection. *Legislative Studies Quarterly*, 33(4):543–572.

Catalinac, A. (2016). From Pork to Policy: The Rise of Programmatic Campaigning in Japanese Elections. *The Journal of Politics*, 78(1):1–18.

Catalinac, A. (2018). Positioning under Alternative Electoral Systems: Evidence from Japanese Candidate Election Manifestos. *American Political Science Review*, 112(1):31–48.

Ceron, A. (2012). Bounded oligarchy: How and when factions constrain leaders in party position-taking. *Electoral Studies*, 31(4):689–701.

Ceron, A. (2015). Brave rebels stay home: Assessing the effect of intra-party ideological heterogeneity and party whip on roll-call votes. *Party Politics*, 21(2):246–258.

Ceron, A. (2016). Inter-factional conflicts and government formation: Do party leaders sort out ideological heterogeneity? *Party Politics*, 22(6):797–808.

Ceron, A. (2017). Intra-party politics in 140 characters. *Party Politics*, 23(1):7–17.

Chang, C. and Masterson, M. (2019). Using Word Order in Political Text Classification with Long Short-term Memory Models. *Political Analysis*, pages 1–17.

Cox, G. W. (2005). *The Efficient Secret: The Cabinet and the Development of Political Parties*. Cambridge University Press.

Cox, G. W. and Rosenbluth, F. (1993). The Electoral Fortunes of Legislative Factions in Japan. *American Political Science Review*, 87(3):577–589.

Croissant, A. and Chambers, P. (2010). Unravelling Intra-Party Democracy in Thailand. *Asian Journal of Political Science*, 18(2):195–223.

Curtis, G. L. (1999). *The Logic of Japanese Politics: Leaders, Institutions, and the Limits of Change*. Columbia University Press. Google-Books-ID: NHS5Eew3ce0C.

de Figueiredo, J. M. and Richter, B. K. (2014). Advancing the Empirical Research on Lobbying. *Annual Review of Political Science*, 17(1):163–185.

Depauw, S. and Martin, S. (2009). Legislative party discipline and cohesion in comparative perspective. In *Intra-Party Politics and Coalition Governments*, pages 103–120. Routledge, London.

Ding, T., Bickel, W. K., and Pan, S. (2017). Multi-View Unsupervised User Feature Embedding for Social Media-based Substance Use Prediction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2275–2284, Copenhagen, Denmark. Association for Computational Linguistics.

Ecker, A. (2017). Estimating Policy Positions Using Social Network Data: Cross-Validating Position Estimates of Political Parties and Individual Legislators in the Polish Parliament. *Social Science Computer Review*, 35(1):53–67.

Erzeel, S. and Celis, K. (2016). Political parties, ideology and the substantive representation of women. *Party Politics*, 22(5):576–586.

Fischer, M. (2019). How MPs' ties to interest groups matter for legislative co-sponsorship. *Social Networks*, 57:34–42.

Galen, A., Raman, A., Bilmes, J., and Livescu, K. (2013). Deep Canonical Correlation Analysis. In *International conference on machine learning.*, page 9.

Gianetti, D. and Benoit, K. (2009). *Intra-Party Politics and Coalition Governments*. Routledge, New York, 1st edition.

Godbout, J.-F. and Høyland, B. (2017). Unity in Diversity? The Development of Political Parties in the Parliament of Canada, 1867–2011. *British Journal of Political Science*, 47(3):545–569.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT press.

Greene, Z. and Haber, M. (2017). Maintaining Partisan ties: Preference divergence and partisan collaboration in Western Europe. *Party Politics*, 23(1):30–42.

Hall, R. L. and Deardorff, A. V. (2006). Lobbying as Legislative Subsidy. *The American Political Science Review*, 100(1):69–84.

Hall, R. L. and Miller, K. C. (2008). What Happens After the Alarm? Interest Group Subsidies to Legislative Overseers. *Journal of Politics*, 70(4):990–1005.

Harmel, R., Heo, U., Tan, A., and Janda, K. (1995). Performance, leadership, factions and party change: An empirical analysis. *West European Politics*, 18(1):1–33.

Harris, Z. S. (1954). Distributional Structure. *WORD*, 10(2-3):146–162.

Heinkelmann-Wild, T., Kriegmair, L., Rittberger, B., and Zangl, B. (2019). Divided they fail: the politics of wedge issues and Brexit. *Journal of European Public Policy*, 0(0):1–19.

Hilderman, J. and Thomas, P. (2013). Climbing the ladder of dissent: Backbench influence in the Canadian House of Commons. Victoria, BC.

Hix, S. and Noury, A. (2016). Government-Opposition or Left-Right? The Institutional Determinants of Voting in Legislatures. *Political Science Research and Methods*, 4(2):249–273.

Hong, L. and Davidson, B. D. (2010). Empirical study of topic modeling in Twitter. In *Proceedings of the First Workshop on Social Media Analytics*, Washington D.C.

Hopkins, V. (2019). It's coming from inside the House (of Commons): Agenda control, accountability, and interest group lobbying in majoritarian parliaments. *Governance*, page gove.12454.

Indriason, I. H. and Kristinsson, G. H. (2015). Primary consequences: The effects of candidate selection through party primaries in Iceland. *Party Politics*, 21(4):565–576.

Jeffrey, B. (2017). The Liberal Party of Canada: Rebuilding, Resurgence, and Return to Power. In *Canadian Parties in Transition*. University of Toronto Press, 3rd edition.

Jenkins, S. (2006). The Impact of Party and Ideology on Roll-Call Voting in State Legislatures. *Legislative Studies Quarterly*, 31(2):235–257.

Kam, C. (2009). *Party Discipline and Parliamentary Politics*. Cambridge University Press.

King, A. S., Orlando, F. J., and Sparks, D. B. (2016). Ideological Extremity and Success in Primary Elections: Drawing Inferences From the Twitter Network. *Social Science Computer Review*, 34(4):395–415.

Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805.

Kölln, A.-K. and Polk, J. (2017). Emancipated party members: Examining ideological incongruence within political parties. *Party Politics*, 23(1):18–29.

LaPira, T. M. and Thomas, H. F. (2014). Revolving door lobbyists and interest representation. *Interest Groups & Advocacy*, 3(1):4–29.

Lauderdale, B. E. and Herzog, A. (2016). Measuring Political Positions from Legislative Speech. *Political Analysis*, 24(3):374–394.

Le, Q. and Mikolov, T. (2014). Distributed Representations of Sentences and Documents. *International Conference on Machine Learning*, page 9.

Malcolmson, P., Myers, R., Baier, G., and Bateman, T. (2016). *The Canadian Regime: An Introduction to Parliamentary Government in Canada, Sixth Edition*. University of Toronto Press. Google-Books-ID: J8o6DAAAQBAJ.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv:1301.3781 [cs]*. arXiv: 1301.3781.

Mitrovica, A. (2019). Chrystia Freeland: Trudeau's heir apparent. *Al-Jazeera*.

Mügge, L. M., Pas, D. J. v. d., and Wardt, M. v. d. (2019). Representing their own? Ethnic minority women in the Dutch Parliament. *West European Politics*, 42(4):705–727.

Newman, M. E. J. (2003). Mixing patterns in networks. *Physical Review E*, 67(2):026126.

Ng, A. Y. and Jordan, M. I. (2002). On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes. *Advances in neural information processing systems*, pages 841–848.

Pan, S. and Ding, T. (2019). Social Media-based User Embedding: A Literature Review. *arXiv preprint arXiv:1907.00725*.

Pennacchiotti, M. and Popescu, A.-M. (2011). A Machine Learning Approach to Twitter User Classification. In *Fifth International AAAI Conference on Weblogs and Social Media*.

Pennington, J., Socher, R., and Manning, C. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*.

Perozzi, B., Al-Rfou, R., and Skiena, S. (2014). DeepWalk: Online Learning of Social Representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710. ACM.

Proksch, S.-O. and Slapin, J. B. (2010). Position Taking in European Parliament Speeches. *British Journal of Political Science*, 40(3):587–611.

Proksch, S.-O., Wratil, C., and Wackerle, J. (2019). Testing the Validity of Automatic Speech Recognition for Political Text Analysis. *Political Analysis*, 27(3):339–359.

Qi, L., Zhang, C., Tavanapong, W., Peterson, D. A. M., and Sukul, A. (2016). Automated Coding of Political Video Ads for Political Science Research.

Rheault, L. and Cochrane, C. (2020). Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora. *Political Analysis*, pages 1–22.

Rice, S. A. (1925). The Behavior of Legislative Groups: A Method of Measurement. *Political Science Quarterly*, 40(1):60–72.

Romesburg, C. (2004). *Cluster Analysis for Researchers*. Lulu.com. Google-Books-ID: ZuIPv7OKm10C.

Rose, R. (1964). Parties, Factions and Tendencies in Britain. *Political Studies*, 12(1):33–46.

Saalfeld, T. (2011). Parliamentary Questions as Instruments of Substantive Representation: Visible Minorities in the UK House of Commons, 2005–10. *The Journal of Legislative Studies*, 17(3):271–289.

Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513–523.

Schnakenberg, K. E. (2017). Informational Lobbying and Legislative Voting: INFORMATIONAL LOBBYING AND LEGISLATIVE VOTING. *American Journal of Political Science*, 61(1):129–145.

Schwarz, D., Traber, D., and Benoit, K. (2017). Estimating Intra-Party Preferences: Comparing Speeches to Votes. *Political Science Research and Methods*, 5(2):379–396.

Shawe-Taylor, D. o. C. S. R. H. J., Shawe-Taylor, J., and Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press. Google-Books-ID: 9i0vg12lti4C.

Slapin, J. B. and Proksch, S.-O. (2008). A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science*, 52(3):705–722.

Slider, D. (2010). How United is United Russia? Regional Sources of Intra-party Conflict. *Journal of Communist Studies and Transition Politics*, 26(2):257–275.

Spirling, A. and Rodriguez, P. L. (2019). Word Embeddings: What works, what doesn't, and how to tell the difference for applied research. *Working Paper*.

Stockemer, D. and Sundström, A. (2019). Do young female candidates face double barriers or an outgroup advantage? The case of the European Parliament. *European Journal of Political Research*, 58(1):373–384.

Thayer, N. B. (1969). *How the Conservatives Rule Japan*. Princeton university press edition.

Thorndike, R. L. (1953). Who belongs in the family. *Psychometrika*.

Tyllström, A. (2019). More Than a Revolving Door: Corporate lobbying and the socialization of institutional carriers. *Organization Studies*, page 0170840619848014.

Verge, T. and Gómez, R. (2012). Factionalism in multi-level contexts: When party organization becomes a device. *Party Politics*, 18(5):667–685.

Wieting, J., Bansal, M., Gimpel, K., and Livescu, K. (2016). Towards Universal Paraphrastic Sentence Embeddings. *arXiv:1511.08198 [cs]*. arXiv: 1511.08198.

Yadav, V. (2017). Studying legislative party politics in data scarce environments: A new empirical approach. *Party Politics*, 23(2):135–147.

Zariski, R. (1960). Party Factions and Comparative Politics: Some Preliminary Observations. *Midwest Journal of Political Science*, 4(1):27.

Zhao, J., Xie, X., Xin, X., and Sun, S. (2017). Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38(November):43–54.

Zhuravskaya, E., Petrova, M., and Enikolopov, R. (2019). Political Effects of the Internet and Social Media. SSRN Scholarly Paper ID 3439957, Social Science Research Network, Rochester, NY.

8 Appendix 1: Analysis without Ministers Included

There is interest in how the inclusion of ministers affects the representations in the party. Part of the concern is that well known MPs will be interacted with often by other MPs without that being indicative of any real social proximity. I re-ran the analysis from section 4, excluding all ministers in the Trudeau government. The new clusters are presented in Figure 21. The results are briefly discussed in this appendix. [12] The general relationships found in the main analysis are robust to the exclusion of ministers from the sample. Region, minority status and ideology remain correlated with the group in which a MP is located. In the minister-less representations, the MPs' region seems to be even more indicative of the group. This is evident on Figure 22. MPs from Quebec, Ontario and Atlantic Canada's representations are clearly separated from one another. Minority MPs are also clearly more proximate. I show this in Figure 23. Minority MPs are mostly located in the top part of the plot. This result is also consistent with the main analysis.

Is the relationship between MP representations and the MP-Lobbyist network detected in section 5 caused by lobbyists mainly visiting ministers? I reproduce section 5's validation task in Figure 24. When removing ministers from the set of MPs, the correlation between groups generated by multi-view representations and the network of MP-Lobbyist interactions is not affected significantly. The correlation coefficient is slightly higher than in section 5. The baseline comparison methods lose statistical significance when ministers are removed from the sample.

Overall, I find that the paper's results are robust to the removal of ministers. When excluding ministers, the separation by region of MPs becomes more pronounced than when ministers are included. This may indicate that regional dynamics play an important role in backbencher interactions. Proximity between minority MPs is also a phenomenon reproduced when excluding ministers. It is important to keep in mind that ministers

[12] The multinomial logit model is not reproduced in this appendix. The reason is that there is perfect separation based on region in the backbencher representations, which makes the multinomial model numerically unstable.

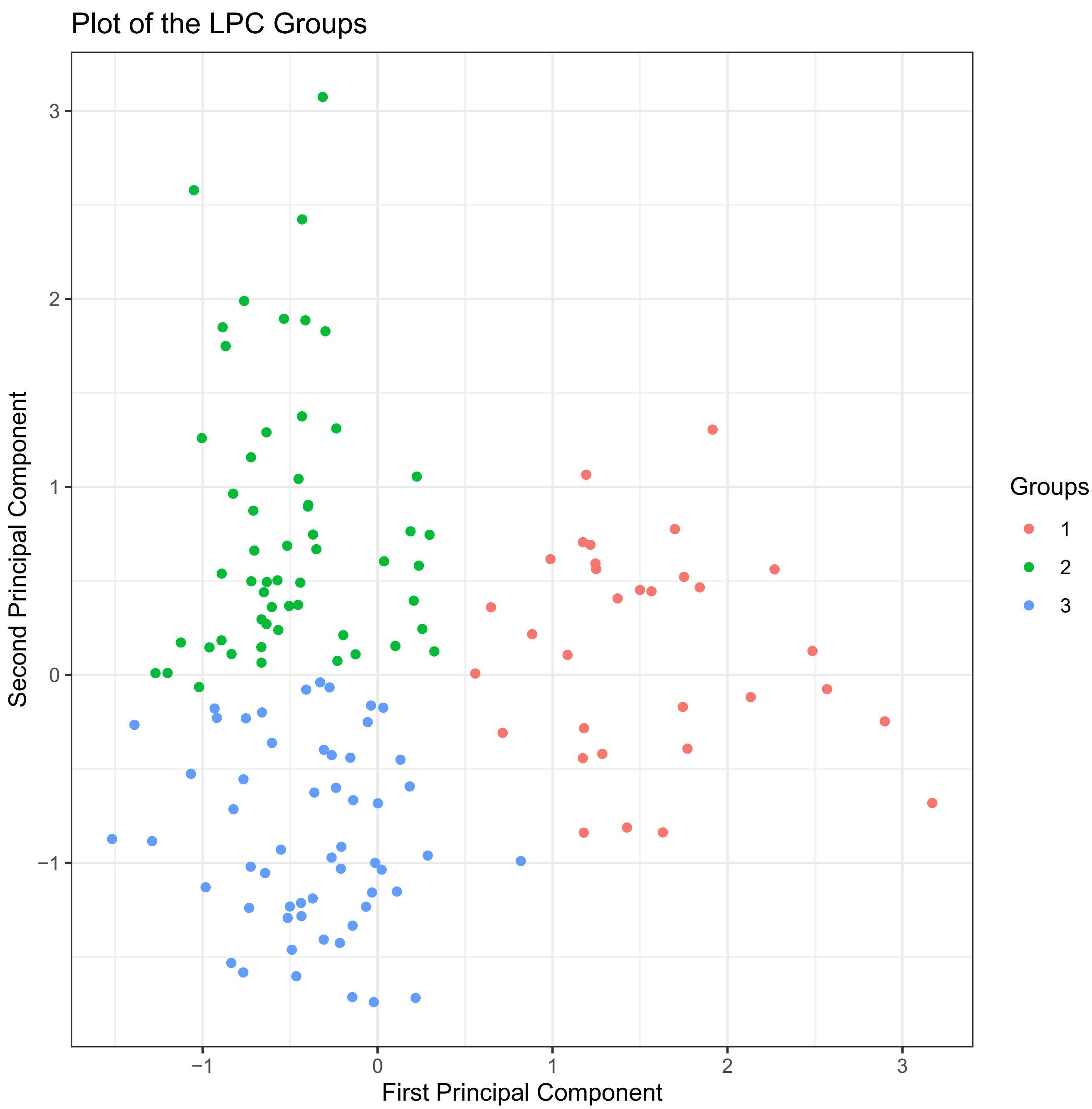


Figure 21: Groups based on backbencher representations only.

represent 19% of the original sample. Some degree of change from excluding such a large percentage of the observations is to be expected.

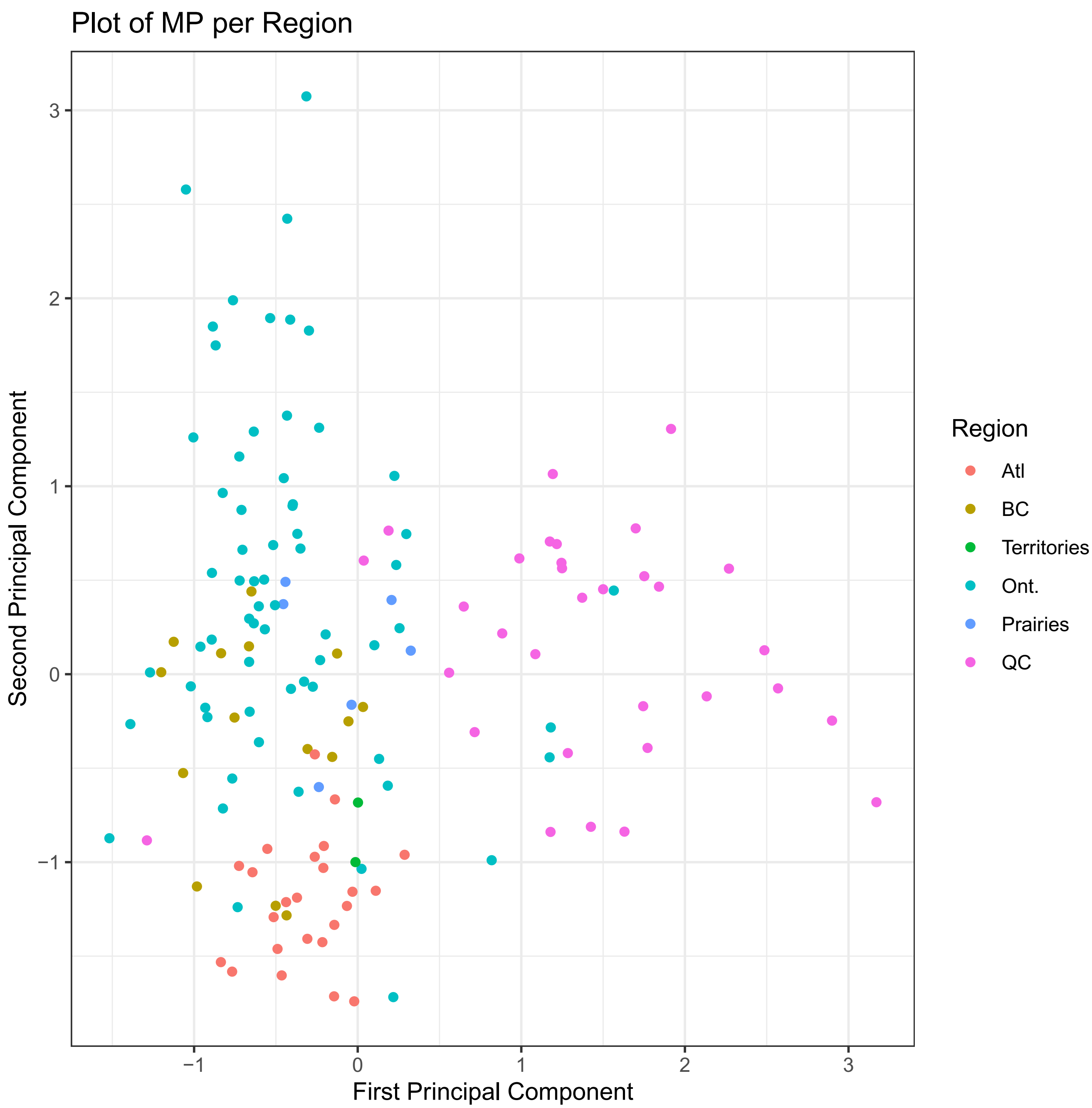


Figure 22: Representations of the backbenchers, color coded by region.

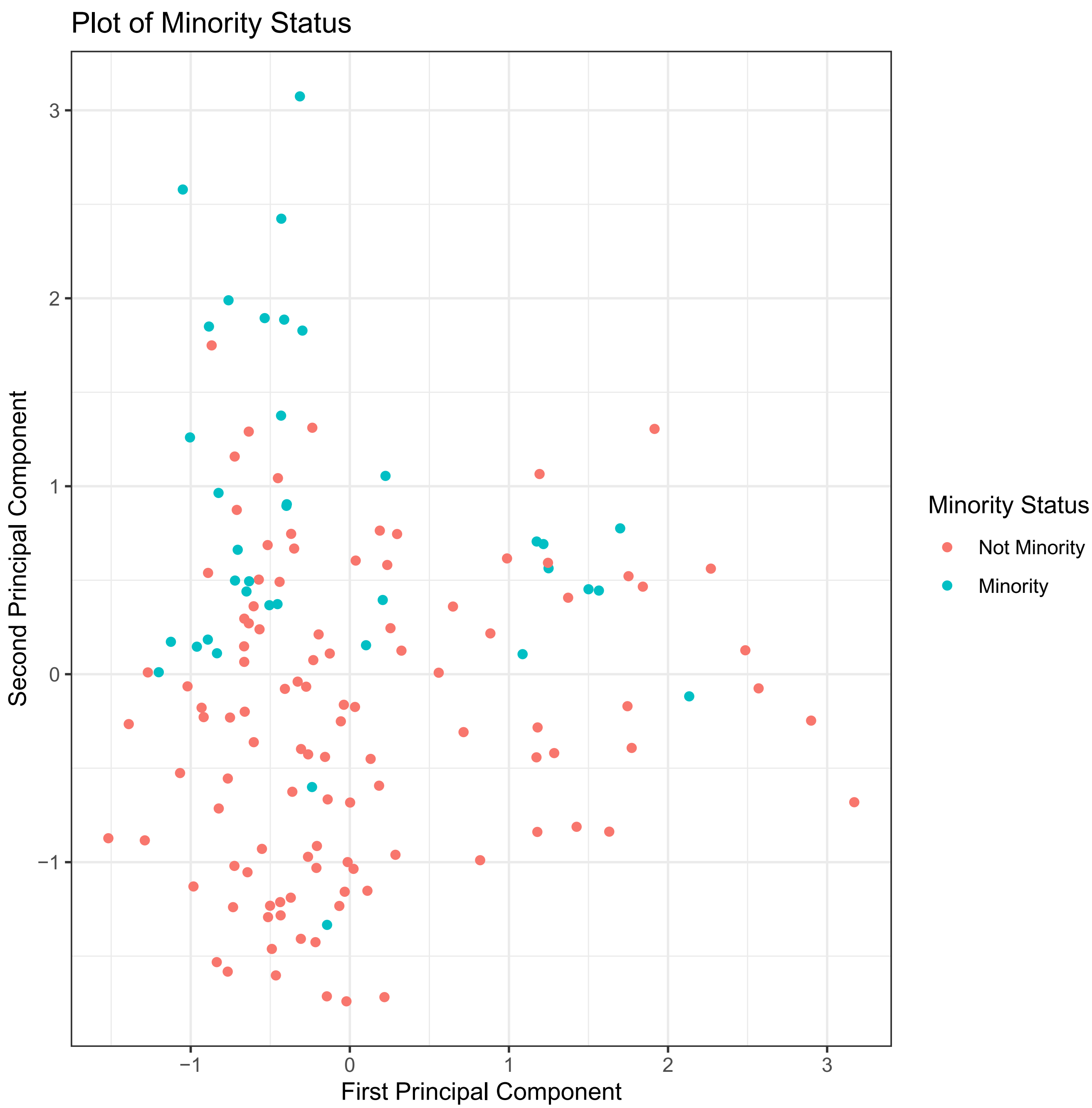


Figure 23: Backbenchers color-coded by minority status.

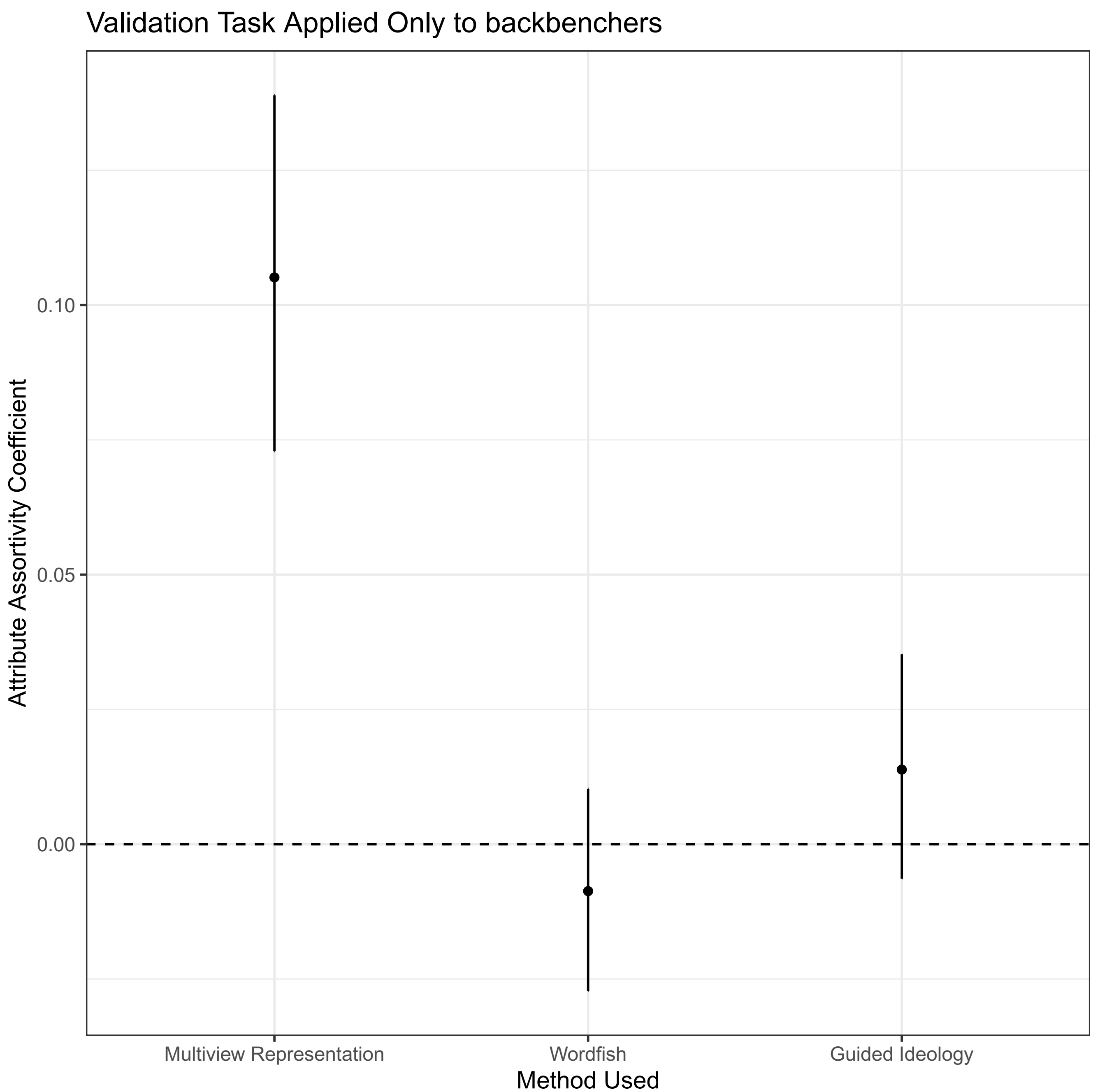


Figure 24: Validation task applied to backbenchers only.

9 Appendix 2: Steps to Solve the CCA Algorithm

Appendix 2 presents supplementary derivations with regards to how to find the transformations that optimize the correlation between 2 sets of vectors in CCA.

Recall that the solution can be found by solving the following optimization problem where C_{xx} is a covariance matrix for vector x, C_{yy} is a covariance matrix for vector y, w_x is a linear projection for vector x and w_y a linear projection for vector y. Finding the linear projections that maximize correlation between the two sets of vectors can be done through the following constrained optimization problem:

$$\max w_x, w_y \;\; w_x^T C_{xy} w_y$$

$$s.t\; w_x^T C_{xx} w_x = 1, w_y^T C_{yy} w_y = 1$$

By applying the Lagrangian multiplier technique to the optimization the following formula is obtained:

$$\max w_x^T C_{xy} w_y - \frac{\lambda_x}{2}(w_x^T C_{xx} w_x - 1) - \frac{\lambda_y}{2}(w_y^T C_{yy} w_y)$$

Taking derivatives of the previous formula with respect to w_x and w_y yields the following two equations:

$$C_{xy} w_y - \lambda_x C_{xx} w_x = 0$$

$$C_{yx} w_x - \lambda_y C_{yy} w_y = 0$$

Substracting w_x^T times the first equation from w_y^T times the second equation yields:

$$\lambda_x w_x^T C_{xx} w_x - \lambda_y w_y^T C_{yy} w_y = 0$$

Taking into account the constraints, this implies that $lambda_x$ and λ_y are the same. Denoting this value by λ, the following system of equations can be inferred from the previous equations:

$$C_{xy}w_y = \lambda C_{xx}w_x$$

$$C_{yx}w_x = \lambda C_{yy}w_y$$

The vectors w_x and w_y can therefore be found by solving a generalized eigenvalue problem.

10 Appendix 3: Full list of MPs and their group

Appendix 3 presents all MPs and their associated group.

Candidate	Group
Adam Vaughan	1
Ahmed Hussen	3
Alaina Lockhart	1
Alexandra Mendès	1
Ali Ehsassi	3
Amarjeet Sohi	2
Andrew Leslie	2
Andy Fillmore	3
Angelo Iacono	2
Anita Vandenbeld	3
Anju Dhillon	3
Anthony Housefather	3
Anthony Rota	1
Arif Virani	3
Bardish Chagger	2
Bernadette Jordan	1
Bill Blair	2
Bill Morneau	2
Bob Bratina	3
Borys Wrzesnewskyj	2
Brenda Shanahan	2
Bryan May	3
Carla Qualtrough	2
Carolyn Bennett	3

Catherine Mary Mckenna	3
Celina Caesar-chavannes	3
Chandra Arya	3
Chris Bittle	3
Chrystia Freeland	2
Churence Rogers	1
Colin Fraser	1
Dan Ruimy	2
Dan Vandal	1
Darrell Samson	1
Darren Fisher	1
Darshan Singh Kang	3
David Graham	1
David Lametti	2
Deb Schulte	2
Diane Lebouthillier	2
Dominic Leblanc	1
Doug Eyolfson	1
Emmanuel Dubourg	3
Emmanuella Lambropoulos	2
Eva Nassif	2
Fayçal El-khoury	2
Filomena Tassi	2
Francesco Sorbara	2
Francis Drouin	1
Francis Scarpaleggia	3
Frank Baylis	2
François-philippe Champagne	2

Gagan Sikand	3
Gary Anandasangaree	3
Geng Tan	3
Geoff Regan	3
Ginette Petitpas Taylor	2
Gordie Hogg	3
Greg Fergus	3
Gudie Hutchings	1
Harjit S Sajjan	2
Hedy Fry	2
Hunter Tootoo	1
Iqra Khalid	3
James Maloney	1
Jane Philpott	1
Jati Sidhu	1
Jean Yip	3
Jean-claude Poissant	2
Jean-yves Duclos	2
Jennifer O'connell	2
Jim Carr	2
Joe Peschisolido	2
John Aldag	2
John Mckay	3
John Oliver	3
Jonathan Wilkinson	1
Joyce Murray	2
Joël Lightbound	1
Judy Sgro	3

Julie Dabrusin	1
Julie Dzerowicz	3
Justin Trudeau	2
Kamal Khera	3
Karen Ludwig	2
Karen Mccrimmon	1
Karina Gould	2
Kate Young	3
Ken Hardie	3
Ken Mcdonald	1
Kent Hehr	1
Kevin Lamoureux	1
Kim Rudd	1
Kirsty Duncan	3
Kyle Peterson	3
Larry Bagnell	3
Larry Maguire	1
Lawrence Macaulay	1
Linda Lapointe	3
Lloyd Longfield	3
Majid Jowhari	3
Marc G Serré	1
Marc Garneau	2
Marc Miller	2
Marco Mendicino	3
Marie-claude Bibeau	2
Mark Eyking	1
Mark Gerretsen	1

Mark Holland	3
Marwan Tabbara	3
Mary Ng	2
Maryam Monsef	2
Maryann Mihychuk	1
Matt Decourcey	3
Michael Levitt	3
Michael Mcleod	1
Michel Picard	2
Mike Bossio	1
Mona Fortier	3
Mélanie Joly	2
Nathaniel Erskine-smith	3
Navdeep Bains	2
Neil Ellis	1
Nick Whalen	1
Nicola Di Iorio	2
Omar Alghabra	1
Pablo Rodriguez	2
Pam Damoff	3
Pam Goldsmith-jones	3
Patty Hajdu	2
Paul Lefebvre	1
Peter Fonseca	3
Peter Fragiskatos	3
Peter Schiefke	2
Pierre Breton	2
Rachel Bendayan	3

Raj Grewal	3
Raj Saini	3
Ralph Goodale	2
Ramesh Sangha	3
Ramez Ayoub	2
Randeep Sarai	3
Randy Boissonnault	2
Richard Hébert	2
Rob Oliphant	3
Robert D. Nault	1
Robert-falcon Ouellette	1
Rodger Cuzner	1
Ron Mckinnon	2
Ruby Sahota	3
Rémi Massé	2
Salma Zahid	3
Scott Brison	2
Scott Simms	1
Seamus O'regan	1
Sean Casey	1
Sean Fraser	1
Shaun Chen	3
Sherry Romanado	2
Sonia Sidhu	3
Stephen Fuhr	3
Steve Mackinnon	2
Stéphane Lauzon	2
Sukh Dhaliwal	2

Sven Spengemann	3
Terry Beech	3
Terry Duguid	2
Terry Sheehan	2
Tj Harvey	1
Vance Badawey	1
Wayne Easter	1
Wayne Long	1
William Amos	1
Yasmin Ratansi	3
Yves Robillard	3
Yvonne Jones	1

11 Appendix 4: The number of Principal Components

In this thesis the representations were reduced from 64 dimensions to 2 dimensions using principal component analysis. This decision was made for two reasons. First, reducing the number of dimensions was necessary to allow k-means clustering to be used on the data. Limiting the dimensionality of the data is necessary due to the curse of dimensions, a phenomenon where the distance between points becomes meaningless as the dimensionality of the data increases. Second, reducing the representations to 2 dimensions allows to visualize them in 2 dimensions which facilitates interpretation. Analyzing a 2 dimensional scatter plot is an often used technique when analyzing embeddings.

Principal component analysis allows to keep a subspace of the data that explains a large part of the variance. Some heuristic techniques exist to determine how many components to keep when analyzing the data. in figure 25, I use a scree plot of the eigenvalues of the first 20 components of the data to observe whether there is an optimal number of principal components to keep in the current data. Ideally, one would include components until there is a sharp drop in the eigenvalue, indicating that adding extra components does not explain a lot more variance. In this case the drop is linear across the observed eigenvalues. There is no clear cutoff that helps decide how many principal components to keep.

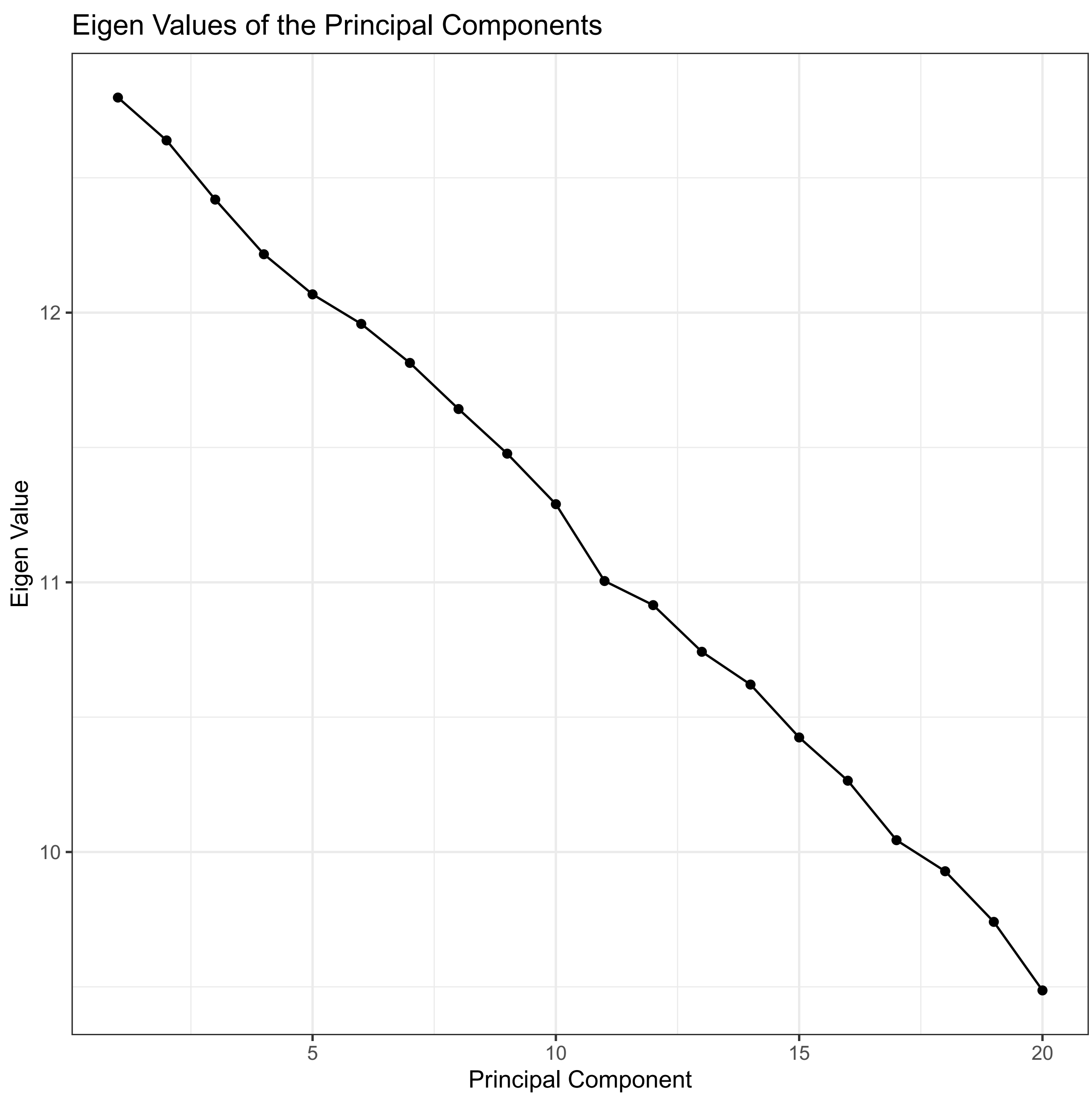


Figure 25: Eigen Value for the first 20 Principal Component.