Canadian Political Diversification Through Network Analysis

Kaitlin de Chastelain Finnigan
Computer Science, 572
University of Calgary
Calgary, Canada
kaitlin.dechastelain@ucalgary.ca 30044556

Noah Giustini
Computer Science, 572
University of Calgary
Calgary, Canada
noah.giustini1@ucalgary.ca 30041939

Luke Iremadze
Computer Science, 572
University of Calgary
Calgary, Canada
luke.iremadze@ucalgary.ca 10163614

***Abstract*—Canadian Members of Parliament and Activist Organizations Twitter accounts retweets were analyzed for the brief period leading up to the 2021 federal elections. The object was to identify echo chambers that plague the scene and to find remedies in the form of bridge nodes that would serve to deliver a broader view for the people following. We believe doing so would reduce confirmation bias that is a leading cause for echo chambers, leading to a less polarized society. We found six main communities in Canadian politics, which we applied centrality measures, and information diffusion techniques to answer three research questions: 1. Given a user's interactions with political parties' tweets on Twitter, can we identify feedback loops created as a result of these politics-based interactions? 2. Can we identify activist organizations as non-partisan or partisan based on their interactions? 3. Can we identify users who have a wide breadth of interaction with different political groups? Our research helped us answer these questions to a certain extent, we can identify where the feedback loops are, place Activist Organizations on a political spectrum, and identify ways to combat confirmation bias by following certain Twitter accounts based on a few measures. We think this should help combat polarization in society and spark an open dialogue with people with different opinions.**

Keywords—network science, Canadian politics, echo chambers

I. INTRODUCTION

Canadians, each federal election cycle, typically every four years, vote for people they want to represent them in the House of Commons. Members of the House of Commons are also known as Members of Parliament or MPs for short. Each MP is part of a political party, and after the election, the MPs whose affiliation with a political party form a majority vote for the Prime Minister, the party leader. [1]

However, politics involves competition and where there is competition, there are emotion. Politics have always been a heated subject through time everywhere globally, and Canada is no exception. As recently as 1980, Canada went through much drama on its construction of a new constitution, the Canadian Charter of Rights and Freedoms, where it was even willing to let go of the province of Quebec had they not come to terms. [2]

Unlike the 1980s, when people and politicians exchanged opinions in person or expressed it through press and television, today, we have social media as a far more prominent influencer due to its scale and reach. Furthermore, unlike the press and media, where news agencies incorporate different views, social media tailors information to each user. This ability to choose what to listen to feeds confirmation bias, which is the inclination to give preferential treatment to evidence supporting an existing belief. [3] This selective exposure to content is the primary driver of content diffusion and generates homogenous (one-dimensional) clusters, called echo chambers. [4] Echo chambers are thought to result from confirmation bias and social influence, which are all the ingredients for today's social media-driven politics. [5] We want to identify these echo chambers in Canadian politics, and for this, we will analyze Twitter accounts of MPs.

Twitter attracts a lot of political figures around the world, including Canadian MPs. It is a platform publicly open for everyone to see who is saying what and who likes what. The number of tweets per day has skyrocketed, from 5000 in February 2007 to almost 800 million as of today. [6] It would be impractical to analyze and look at each action individually; Thus, we would obtain a subset for our analysis. Once our data is gathered, we will apply network analysis to gain insights into Canadian politics.

We will be looking at MP Twitter accounts in our paper, but without analysis, we predict this alone will have echo chambers where the MP of a party retweets only what another MP has tweeted from the same party. Instead of researching the obvious, we would like to find ways to break this phenomenon, perhaps there are activist organizations that could help with this, or certain users on Twitter that we could identify and recommend our readers to follow for receiving a balanced news feed. Therefore, we will also be looking at including activist organizations in our analysis to find way to break out of the echo chamber.

Few papers seemed relevant to ours, where network analysis was used in a Canadian politics setting. "What the Hashtag" is a Canadian politics-based paper that looks at the use of a popular hashtag (#cdnpoli) on Twitter and draws conclusions as to what content it carries, although it paid no attention to echo chambers or was not focused on MPs alone. [7] Another looked at media coverage in the context of the Canadian climate discussion held between 2006-2010 but did not use Twitter/Social media, and its scope is specialized. [8] The third one tries to analyze feedback loops on Twitter and the spread of misinformation using bots, but in US politics; their main objective was finding which political side is guilty of spreading misinformation. [9] There has not been a study on feedback loops within a Canadian Politics sphere as far as our knowledge goes. The goal of this paper then is to identify feedback loops using the Twitter accounts of MPs. As previously mentioned, we predict that loops will form within each political party, we hope to identify activist organizations that would create bridges for Canadians to follow, and lastly investigate individual users who might also bridge the gap between different political parties.

II. DATASET DESCRIPTION

A. Raw Data

This data was obtained by requesting data from Twitter's API. We obtained a developer license to access the endpoints to obtain tweets from a user's timeline as well as the most recent 100 retweeting users of a tweet.

We obtained tweets from MPs in the Canadian Federal Legislature over the time period in the immediate aftermath of the election, specifically September 29, 2021 to October 13, 2021.

We also obtained tweets from major Canadian activist groups over the same period of time. The activist groups were sourced from Political Advocacy Groups in Canada Wikipedia page.

In addition to obtaining the tweets, for each tweet an MP or Activist Organization sent over the specified time period, we also obtained the most recent 100 users that retweeted each tweet. We collected the user ID's from them - we will refer to them throughout this paper as retweeters.

B. Dataset Cleaning & Wrangling

Due to the limitations of the retweeting users' endpoint, we could only obtain the latest 100 retweeting users of a tweet, however there were still tweets that had more than 100 retweets on them, which meant we were getting a subset of the users who interacted with that tweet. To avoid the situation of missing out on important connections we removed any MP or Activist Organization that had a tweet which got over 100 retweets over the time period of September 29, 2021, to October 13, 2021. This led to different amounts of MPs per party that tweets were collected from. Fig. 1 details the number of MPs per party that data was sourced from.

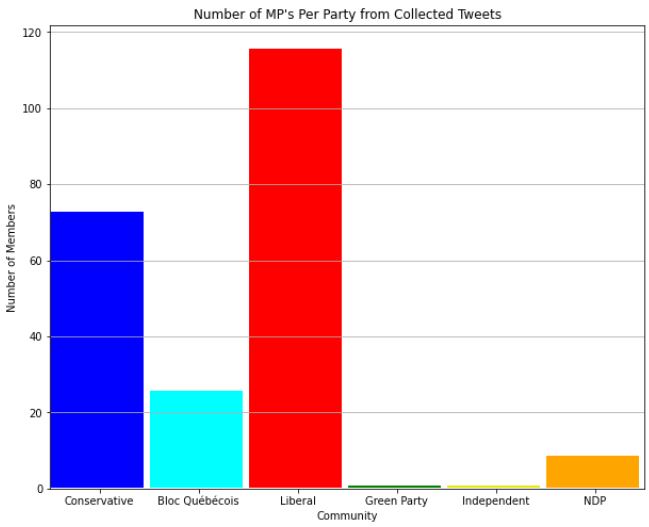

Fig. 1. Histogram of MP membership tweets were sourced from

C. Statistics of The Data After Cleaning

TABLE I. DATA COLLECTED SUMMARY

Characteristics	Total
Tweets Collected	4253
Distinct MPs	227
Distinct Activist Organizations	56
Distinct Retweeters of Activist Organizations	2772
Distinct Retweeters of MPs	6541

D. Network Construction

In constructing the network, we began by identifying the Twitter accounts of MPs and Activist Organizations that we wished to pull tweets from. This allowed us to pull a list of tweets that would be scraped to get the nodes of the network. For each Tweet from an MP or Activist Organization we got the users that retweeted their tweet and created an edge from the retweeter to a Member of Parliament (MP) or Activist Organization (AO). The edge was given a weight based on the number of interactions that existed from that user to the MP indicating interaction with multiple Tweets from the MP. This edge data was exported to a csv file with the source id, target id, and weight. Additionally, when pulling the data for the MPs in our network we pulled metadata for each node id pairing the ids with their Twitter name, political affiliation, and the name of the MP. This gave us the edge data required for the network, as well as node data containing metadata for each node.

E. Nodes

1) *Activist Organization (AO):* A group representing a cause in Canada. This Activist Organization may retweet other Activist Organizations or MP's.

2) *Members of Parliament (MP):* A member of a political party, has an affiliated party attached to this node. This MP may retweet other MP's or Activist Organizations

3) *Retweeter:* This is a user who has not been identified as an AO or MP, but rather is a user who retweets these other nodes.

F. Edges

Edges are directed, they represent: "Source Retweets X tweets of Target" where X is a weight attached describing how many times Source retweeted Target.

Fig. 2. Edge structure

G. Metadata

1) *Edge Weights:* An edge weight is described by the number of times a person has retweeted another person over the time period we sampled from. E.g. If Sally retweets 9 of Justin Trudeau's tweets then the edge weight representing her interaction would be 9.

III. METHODOLOGY

A. Network Collection

In collecting the network we first began by identifying the nodes that we would be studying for the network. This involved finding the Twitter accounts for all the MPs that are serving in office at the time of Sept. 2021, as well as taking a list of activist organizations and getting their twitter handles.

Once these handles were collected the MPs and activist organizations were filtered to ensure that we would not run into any Tweets that would exceed our capacity of 100 retweeters per tweet. This finally allowed us to collect a list of Tweet ids that could be scraped for information. With this list of Tweet ids a python script was setup using Tweepy to make Twitter API calls and collect all of the retweeters for each of the given Tweets in the id list. This data was recorded getting the user id of the user that retweeted the tweet, and the user id of the Tweet's creator. Finally this data was aggregated so that if a person interacted with an MP or activist organization multiple times it would be reflected in the weight of their edge rather than having a redundant number of edges in the network. This edge data with weights was finally written to an output file and this is the basis for the structure of the network.

In addition to finding the edges in the network we needed the metadata for the MPs and activist organizations that we have in our network. To do this, we took the list of MPs and activist organizations and got their id and paired it in rows such that they contained their political affiliation, Twitter handle and their name. Activist Organizations were given a political affiliation of 'Activist Organization' to identify them. This data was finally written to an output file giving us our basic metadata for the nodes in the network.

B. *Community Analysis*

Community analysis was done using the greedy_modularity_communities algorithm provided by NetworkX. We tried various resolution values to decrease the number of communities so we could have communities beyond only an MP or AO and their retweeters.

From there each node was given a community id according to the community they were placed in. These were changed into names to reflect the communities later on.

C. *Community Membership*

We have the information on whether a node is a Retweeter, MP, or AO by their metadata. An AO has the tag Activist Organization as its political affiliation and the MP has their party as their political affiliation, and Retweeters have null. We were able to plot the membership of a community by plotting a histogram [10] of the political affiliation disregarding those members that had null for their political affiliation.

D. *Accuracy of Communities*

We measured the accuracy of communities based on the MPs. We had the MPs political party affiliation, so we compared that to the community they were found to be in. We used a Jaccard similarity score and confusion matrix to compare these. The code was sourced from Cognitive Class' Machine Learning with python course. [11]

E. *Strength of Communities*

To find the strength between communities we used Pandas to add the communities of the source node and target node into the edge list.

TABLE II. STRENGTH OF COMMUNITIES

	target_politic	target_community	source_politic	source_community	Source	Target	Metadata
0	Liberal	Liberal	NaN	Liberal	866967099415113728	86384661	{"weight":1}
1	Liberal	Liberal	NaN	Liberal	47956279	86384661	{"weight":1}
2	Liberal	Liberal	NaN	Liberal	161562952	86384661	{"weight":1}
3	Liberal	Liberal	NaN	Liberal	709304733	86384661	{"weight":2}
4	Liberal	Liberal	NaN	Liberal	1223950760	86384661	{"weight":1}
...	...	...	...	...	...	...	...

From there we were able to count how many interactions happened between communities. Afterwards we used Excel to look at the interactions between the top 6 communities and created a normalized comparison of the interaction found in TABLE VI.

Additionally, we used Pandas to create a count of other communities the node interacted with (both as a source node and as a target node) Target Comm Interaction Count is the number of communities that people who retweet you are from. Source Comm Interaction Count is the number of communities that you retweet that you are not from.

F. *Finding Nonpartisan Communities*

From the community analysis we had a dataset of nodes that allowed us to see the number of communities that retweeted a user as well as the number of different communities that user retweeted.

We start by looking at the Activist Organizations that have 4 or more community interactions – in terms of how we structured the data this meant we were looking for AOs with a Target Comm Interaction Count of 3 or more.

TABLE III. NON-PARTISAN NODE

	Target Comm Interaction Count	id	community	political affiliation
8866	3.0	14079041	NDP	Activist Organization

There was only one organization. We then inspected the edges in Gephi to see if the nodes were from the 3 other political parties.

This was not the case, the three retweeting edges from other parties were from the Liberal, Conservative, and Unprofiled community 13, which consists of the Rideau institute and its retweeters which were not found to have a political affiliation (their work consists of foreign policy counts).

From there we looked at the Activist Organizations that had 2 other communities retweet them and repeated the process outlined above.

G. *Finding Partisan Communities*

From the community analysis we had a dataset of nodes that allowed us to see the number of communities that retweeted a user as well as the number of different communities that user retweeted.

We started by filtering for only Activist Organizations and then viewed those who's Target Community Interaction Count was 0. This indicates that only their community retweeted them.

H. *Identifying Bridging Nodes*

At this point in our analysis the nodes had the number of communities they interact with as a source and target as well as the betweenness centrality of the node itself. Using Pandas we were able to remove the nodes with 0 betweenness and select the top 99, top 90, and top 85 percent to look at as candidates for bridging. We were then able to filter out nodes that didn't interact with the number of communities specified in the criteria described in Identifying Community Bridging Users section.

I. *Identifying Activation Nodes:*

We used pandas to see which MPs interacted with other MPs by looking at the out degree of the MPs and then found the edges those nodes were included in and selected the targets from those. They were then grouped by community to form the activation sets.

J. *Information Cascade: Experiment 1*

The experiment starts with selecting two random nodes from the activation set from the community and then running experiments changing the number of iterations the information diffuses through to determine the cap of how long it takes to reach beyond the community.

In practice we found that nodes were being activated after time t=3 when running the cascade 1000 times.

A time cap of t=5 was used to ensure there was ample time for nodes to be activated.

Each cascade was repeated 1000 times and then the average activation time was taken and then rounded up.

This was done by running the information cascade method that was provided in class. (Notebook: Information-Cascade-Template)

This experiment was performed with 5 different sets of activation nodes, one set from each community that was in the overall activation node set.

K. Information Cascade Experiment 2

The experiment starts with selecting all nodes from the activation set from the community and then running experiments changing the number of iterations the information diffuses through to determine the cap of how long it takes to reach beyond the community.

In practice we found that nodes were being activated after time t=3 when running the cascade 1000 times.

A time cap of t=5 was used to ensure there was ample time for nodes to be activated.

Each cascade was repeated 1000 times and then the average activation time was taken and then rounded up.

This was done by running the information cascade method that was provided in class. (Notebook: Information-Cascade-Template)

This experiment was performed with 5 different sets of activation nodes, one set from each community that was in the overall activation node set.

1) Venn Diagram: Once we had all the nodes cascades, we used the venn diagram library pyvenn [12] to see how many individuals received information from more than one cascade.

L. Comparison to Null Model

We compared our real networks information cascade for each experiment type and each cascade by using a degree preserving null model that was directed and weighted. We then ran the same experiments as detailed above for each set of activation nodes on each null model. The python script we ran on these null models resulted in a csv for each random graph which had the average activation time for each activated node over the 1000 times the cascade was tried for each different set of activation nodes that were used. Here is a sample to illustrate how the data from each null model looked once finished:

TABLE IV. FINISHED NULL MODEL OUTPUT

Unnamed: 0	liberal_average_activation_time	conservative_average_activation_time	ndp_average_activation_time	lgbt_average_activation_time	bloq_average_activati
16014404	NaN	NaN	NaN	NaN	
20199202	NaN	1.0	NaN	NaN	
22849568	0.0	NaN	NaN	NaN	
23491400	1.0	NaN	NaN	NaN	
24990450	NaN	NaN	NaN	NaN	

We were then able to use pandas to get the average activation time for each node over all the null models. From there we determined the number of average number of activated nodes per cascade as well as the standard deviation and the maximum time a node was activated.

M. Null Model Generation

Null models were generated using a degree preserving null model for a directed weighted graph. We did this by adapting the connected double edge swap to preserve the direction of the edges, accounting for the same in and out degree. Original edges $u \rightarrow v$ and $x \rightarrow y$ become $u \rightarrow y$ and $x \rightarrow v$ the edge weight from $u \rightarrow v$ becomes the edge weight for $u \rightarrow y$ and edge weight from $x \rightarrow y$ becomes the edge weight for $x \rightarrow v$

N. Measuring Average Clustering and APL in Null models

For each null model we obtained the average clustering and average path length. We placed these in a Pandas Dataframe and then took the average and standard deviation.

O. Measuring Katz Centrality on Null Models

For each null model the Katz centrality of each node was calculated and then we placed them into a common pandas Dataframe indexed by each node id, we then averaged them to get the average Katz centrality over the null models for each node.

P. Centrality Measures

Centrality measures were computed using NetworkX by taking in the network as a graph and computing the desired centrality measure from it. This resulted in a dictionary where keys were the Id of each node, and its value consisted of its corresponding centrality measure value. The methodology for each centrality measure can be broken down as follows:

a) In-degree Centrality

The network is taken in as a directed graph and NetworkX's in_degree_centrality method was run over the directed graph. The resulting dictionary output is then converted to a Pandas dataframe and then exported to csv.

b) Out-degree Centrality

The network is taken in as a directed graph and NetworkX's out_degree_centrality method was run over the directed graph. The resulting dictionary output is then converted to a Pandas dataframe and then exported to csv.

c) Betweenness Centrality (Directed):

The network is taken in as a directed graph and NetworkX's betweenness_centrality method was run over the directed graph. The betweenness centrality was computed with k=None, and in both normalized and non-normalized forms. The resulting dictionary output is then converted to a Pandas dataframe and then exported to csv.

d) Betweenness Centrality (Undirected)

The network is taken in as an undirected graph and NetworkX's betweenness_centrality method was run over the undirected graph. The betweenness centrality was computed with k=None, and in both normalized and non-normalized forms. The resulting dictionary output is then converted to a Pandas dataframe and then exported to csv.

e) Katz Centrality

The network is taken in as a directed graph and NetworkX's katz_centrality_numpy method was run over the directed graph. The Katz centrality was computed with weight="Weight" in order to ensure that weights were considered in the computation. The output is then converted to a Pandas dataframe and then exported to csv.

The output csv files were then read into Gephi and merged with the table of nodes to pair each id with its corresponding centrality measures and other metadata. The table could then be exported from Gephi or analyzed in Gephi with the centrality measures paired with their respective Ids.

IV. Results

A. Basic Statistics

The network was comprised of 8984 nodes, and 14040 edges. Overall, the network is disconnected, containing 9 connected components, their sizes are: [8932, 10, 9, 8, 7, 4, 3]. The average path length of the largest connected component is 1.23 × 10-3 and its clustering coefficient is 9.61 × 10-3. This largest connected component was used as the size reference for the null models and is what the majority of the paper focuses on. The network displays properties of a scale free network as shown by the degree distributions.

1) Out-degree Distribution: The out degree had a fit of γ = 2.21 this shows that the out degree distribution follows the power law fairly well. The error was 0.012 according to the powerlaw python library's fit function

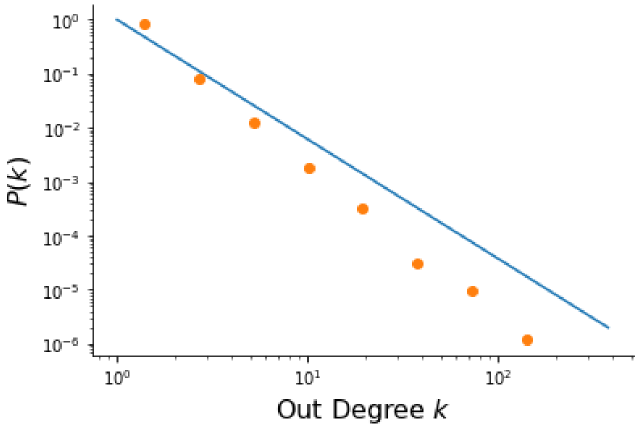

Fig. 3. Out-degree distribution

2) In Degree Distribution: The in-degree had a fit of γ = 10.81. When plotting this we can see the out-degree distribution doesn't follow the power law. The error was 0.103 according to the power-law python library's fit function. The reason this doesn't follow a power law is because the in-degree nodes are likely to have less variation because those users are MPs and AOs which are going to have high in degree because many people are retweeting them.

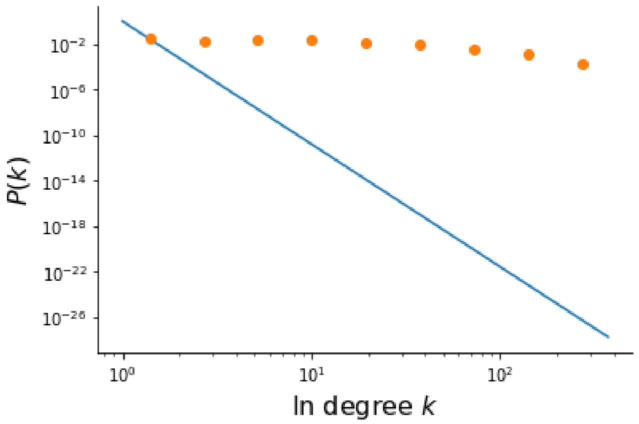

Fig. 4. In-degree centrality

3) Average Path Length: 1.23 × 10^{-3} We expected to find a few nodes that were retweeting more than one party, but for an individual to see the tweet of someone from the other end of a political party does not seem likely. The reason for this is our average path length of 1.23 × 10^{-3}, which is lower than our expectation. The likely reason for this is the introduction of activist organizations, where if another Retweeter also retweets another non-partisan activist organization it will provide them a quick path to another area of the network.

4) Average Clustering Coefficient: 9.61 × 10^{-3} We also anticipated a high level of clustering within the political parties, where MPs were retweeting one another to spread their messages and consequently circulating tweets around the party but, our clustering coefficient of 9.61 × 10^{-3} and the visualizations show that we have many single degree nodes which could be contributing to the lower overall clustering coefficient than expected.

B. Null Model Comparison

TABLE V. below provides the mean and variance for both average clustering and average path length for our two null models we will be comparing with throughout the paper. As a reminder the original data has the following characteristics:

- Clustering Coefficient 9.61 × 10^{-3}
- Average Path Length 1.23 × 10^{-3}

TABLE V. Null Models

Measure	Directed Weighted Degree Preserving Model	Undirected Unweighted Degree Preserving Model
<Clustering Coefficient>	1.51 x 10^{-3}	1.96 x 10^{-2}
σ (Clustering Coefficient)	2.38 x 10^{-4}	1.07 x 10^{-3}
<APL>	2.31 x 10^{-3}	3.28
σ (APL)	1.33 x 10^{-3}	1.47

TABLE V. shows a comparison between the two sets of null models that we generated to compare our network. We can see that the directed weighted models came out with very similar results while the undirected unweighted models came out with much different statistics than the rest. This is expected since the variable that is being changed here is the network being undirected, which would completely change how the clustering and APL are computed for the network. Our real network has a notably higher clustering coefficient and a shorter APL indicating that the network has hub nodes that are consistent with a scale free network which shortens the average path length than those generated in the null models and that the structures we are seeing in our real network are not consistent with those in a random network.

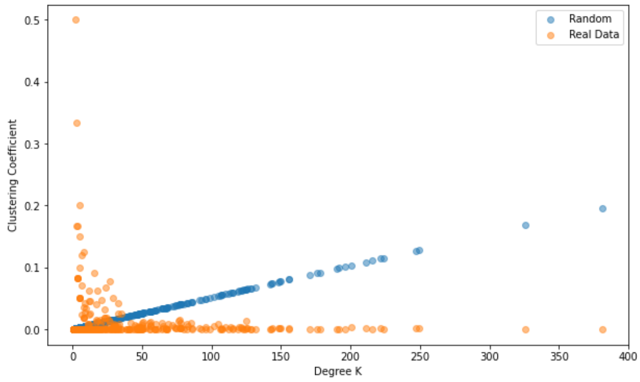

Fig. 6. Degree vs clustering coefficient for real and null model data

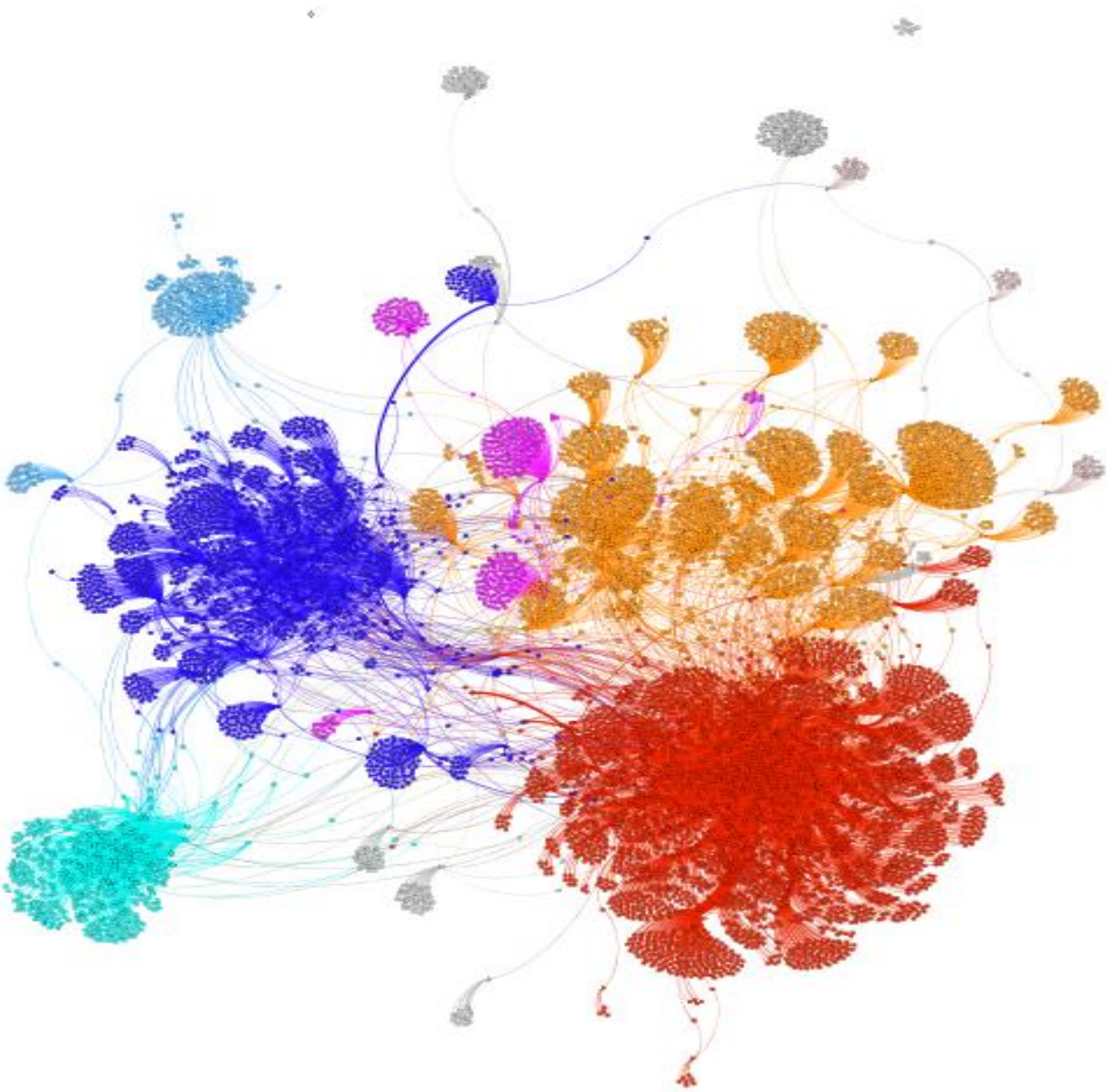

Fig. 5. Visualization of full network coloured according to community

Fig. 6 shows that in our null models there is a correlation between the degree of the node and its clustering coefficient. This is not shown in our network where we have nodes with a low degree, but high clustering.

C. Communities with Greedy Algorithms

Nodes in the network have been coloured according to their assigned community and are coloured as follows:

- Liberal community
- Conservative community
- NDP community
- Bloc Quebecois community

LGBTQ2S+ community
Conservative Activist community

Other communities that were too small to be profiled are coloured in shades of grey. Additionally, nodes in the network are given a different shape according to what the node represents. Specifically, a triangle represents an Activist Organization, a hexagon represents a Member of Parliament, and a circle represents a retweeter. This colouring and shaping guide will be consistently used in the remainder of the visualizations that are shown.

The community analysis that was run was a greedy algorithm supplied by NetworkX – this resulted in a modularity of 0.66, using a resolution value of 0.155 to minimize and encapsulate the main political parties. This allowed us to avoid the problem found where the community would consist of the MP or AO and their followers. This resulted in 21 different communities.

1) Liberal Community

The liberal community consists of Liberals and an independent MP, who was formerly a member of the Liberal party. The Activist Organizations detailed here are mostly student organizations or educational initiatives for international politics. A histogram of membership is shown in Fig. 7.

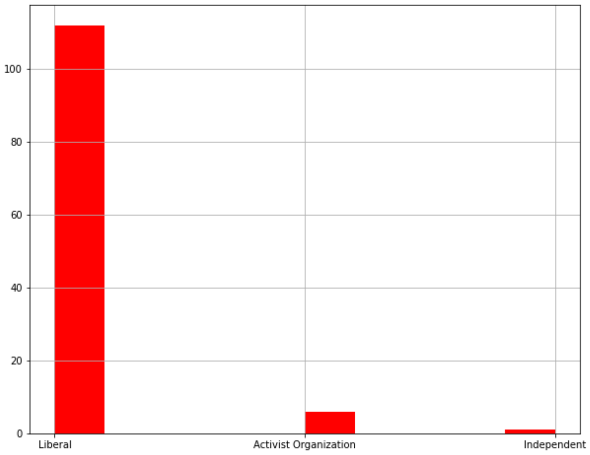

Fig. 7. Liberal community histogram

2) NDP Environmental Community

This community is named the NDP Environmental Community due to its strong affiliation with the cause. This community consists of all NDP & Green party MP's as well as a significant number of activist organizations, as can be seen in detail in Fig. 8. This is consistent as the Green party and NDP party have similar values on environmental issues. The activist organizations in this community follow three common traits: Social Justice, Environmental Activism, and Indigenous Representative accounts. All three of these traits commonly reflect left leaning policy preference.

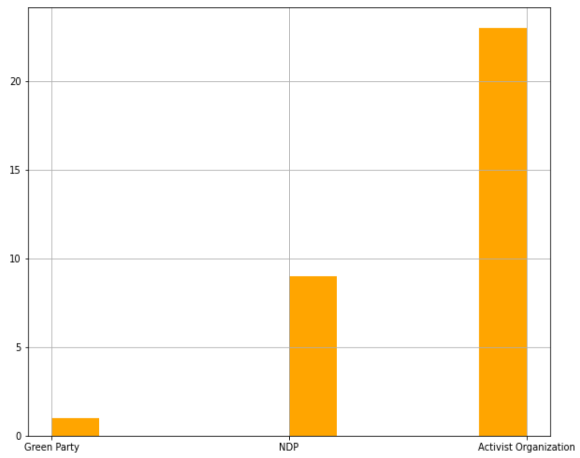

Fig. 8. NDP community histogram

3) Conservative Community

The Conservative Community detailed in Fig. 9 consists of mainly Conservative MP's as well as one individual from the Bloc Quebecois, Simon-Pierre Savard-Tremblay. The activist organizations in this community are ifnaliance, CANZUK, Worldvisioncan, cjpac, and CIJAinfo. Ifnalliance is the Independent First Nations Alliance, and CANZUK looks to promote international relations between Canada, Australia, New Zealand and the UK. For the purposes of the individuals who interact with these organizations they are considered more right leaning.

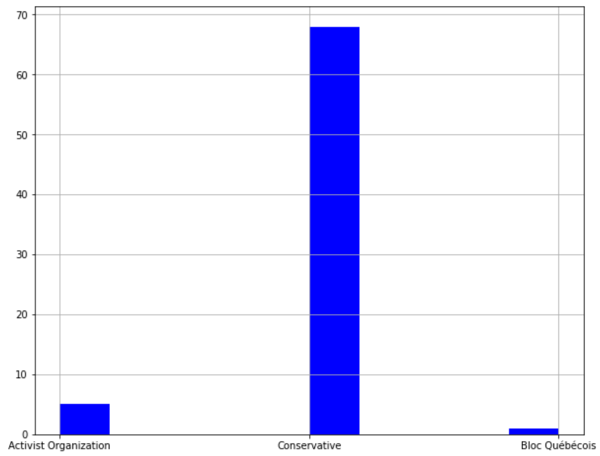

Fig. 9. Conservative community histogram

4) Bloc Quebecois Community

The Bloc Quebecois Community consists of only the MPs from the Bloc Quebecois party, as shown in Fig. 10. We believe the cause is twofold, activist groups are English speaking and are less likely to interact with Francophone users, and/or the views of the Bloc are so closely aligned with conservatives that activist groups that are right leaning may have been grouped in with Fig. 9 instead.

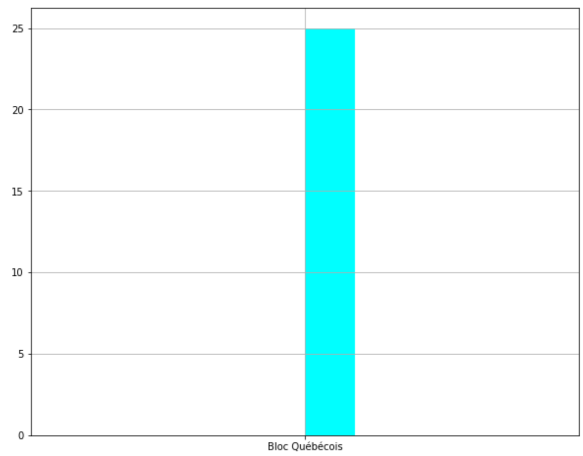

Fig. 10. Bloc Québécois

5) LGBTQ2S+ Civil Liberties Community

The LGBTQ2S+ Civil Liberties Community consists of only activist organizations as detailed in Fig. 11. The five groups that create this community are: CdnBankers, RainbowRailroad, egalecanada, cancivlib, YouCanPlayTeam. Every organization aside from CdnBankers are for the advancement of minorities, specifically LGBTQ2S+. This suggests that the members of this community value the advancement of civil liberties over party values.

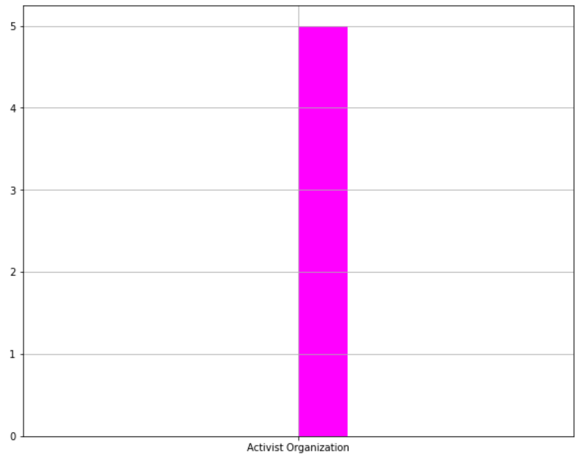

Fig. 11. LGBTQ2S+ community histogram

6) Conservative Activists Community

The Conservative Activists Community consists of activist organizations and one conservative MP, Dan Mazier. The major difference between this and Fig. 12 community are the users who interact with Dan Mazier interact more with the Justice Centre for Constitutional Freedom, and the Campaign Life Coalition. This community is likely more representative of these values these activist groups profess, (which tend to be further right than some conservatives) than that of the actual party.

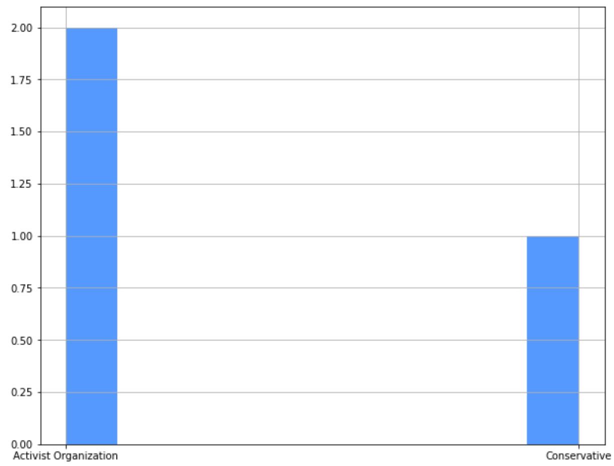

Fig. 12. Conservative activist community histogram

7) Community Size

Fig. 13 displays the membership counts for each identified community. This shows us that 39.6% of retweeters were in the liberal community, indicating there is a lot of people engaging directly with the liberal MP's. It should be noted that of the collected MPs, 51% of them were liberal as seen in Fig. 1. This could be a contributing factor to overall size of the liberal community.

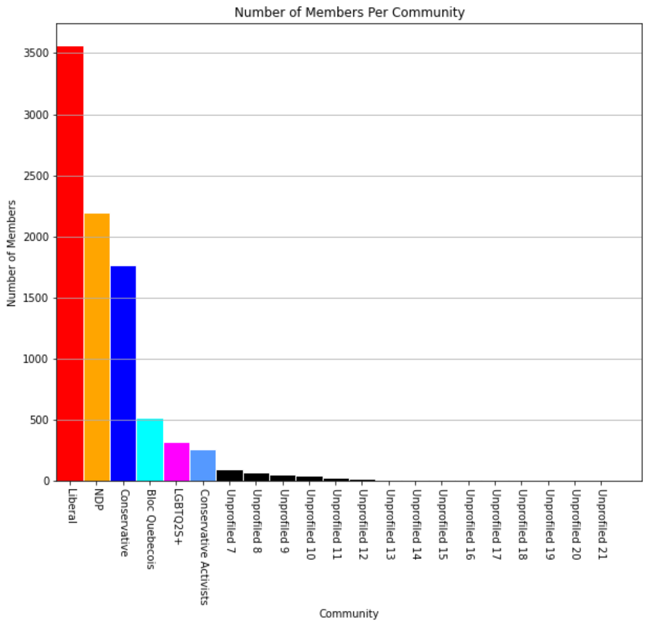

Fig. 13. Community size distribution

8) Accuracy of Communities.

Looking at the MPs that we have the political affiliation for, the Jaccard similarity score was computed to be 0.97. This score indicates that the political affiliation predicted by the community analysis have a 97% overlap with the true political affiliation.

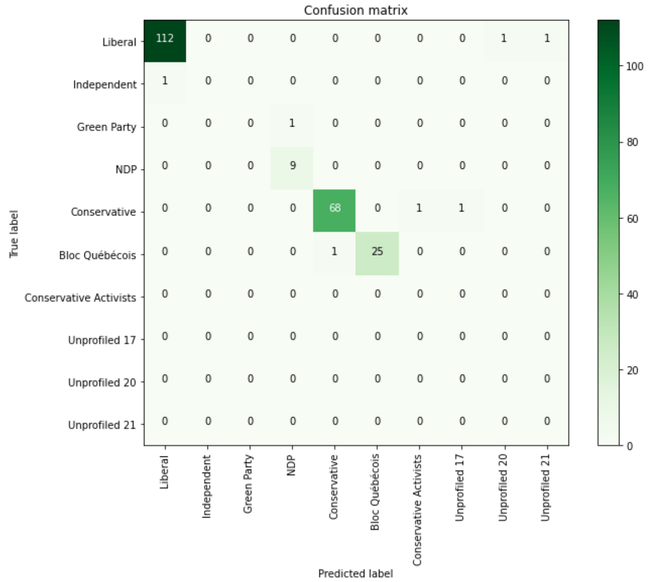

Fig. 14. Confusion matrix of MP's political affiliation vs predicted community

Fig. 14 is a confusion matrix which tells us how many MPs were miscategorized in the community in terms of which party they belong to. There are very few instances in which an MP is miscategorized outside of their community and only one instance where an MP has been mislabeled in terms of which party they belong to.

9) Intercommunity Interaction

We wanted to show how interactive different communities were with other communities. It was not enough for us to take the total number of edges between different communities and show the number, because that would skew the results towards the community that has more members (i.e., The liberal community). Thus, we wanted to show true representation of how each community interacts with other communities on a whole. Fig. 15 shows the results and to our surprise the Conservative community has the best outreach to other communities.

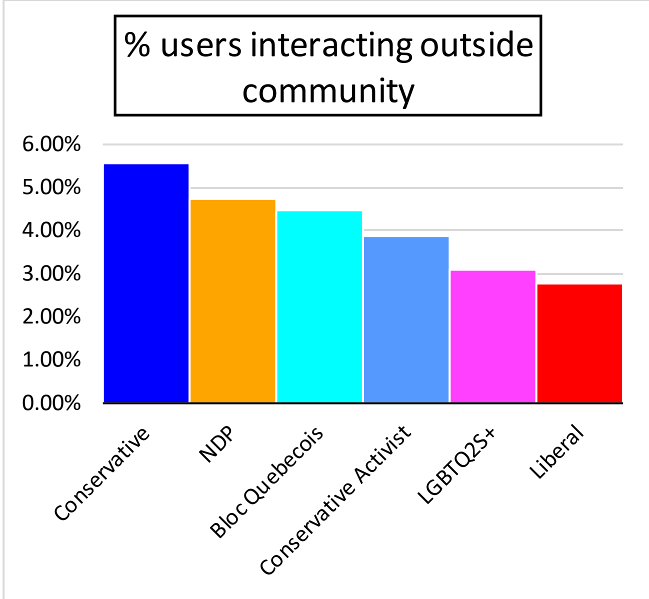

Fig. 15. The % of users interacting outside community

10) Strength Between Communities

An analysis was performed calculating how many edges there were between the communities to indicate the strength of connection between them. TABLE VI. details the percentage of edges that exist between two communities over all edges between communities.

11) Notable Individuals

There are some nodes that are retweeted by four or more communities. TABLE VII. details what the communities are that interact with that node.

TABLE VI. INTERCOMMUNITY CONNECTIONS

Communities	Liberal	NDP	Conservative	Bloc Quebecois	Conservative Activist	LGBTQ2S+
Liberal						
NDP	3.26%					
Conservative	1.90%	1.82%				
Bloc Quebecois	0.32%	1.56%	0.92%			
Conservative Activist	0.00%	0.12%	0.64%	0.13%		
LGBTQ2S+	0.31%	0.44%	0.10%	0.00%	0.00%	

TABLE VII. NOTABLE INDIVIDUALS

Node Twitter Handle	Node ID	Community	Connected Communities
@CharlieAngusNDP	215632349	NDP	Liberal, Conservative, LGBTQ2S+, Conservative Activists
@alexboulerice	196717787	NDP	Liberal, Conservative, LGBTQ2S+, Bloc Québécois

There are some nodes that retweet multiple communities. TABLE VIII. details what the communities that node retweets.

TABLE VIII. NOTABLE INDIVIDUALS RETWEETING FROM MULTIPLE COMMUNITIES

Node Twitter Handle	Node ID	Community	Connected Communities
@ProChristopher	596787864	Liberal	Liberal, Conservative, LGBTQ2S+, Bloc Québécois
@davidakin	12034642	Conservative	Liberal, NDP, Bloc Québécois
@trueintegrity87	711980580813066240	NDP	Liberal, Conservative, Unprofiled 9

12) Activist Organization Partisanship

a) Partisan Activist Organizations: Activist Organizations that were retweeted by more communities than their own: There were 20 activist organizations that were found to have retweeters only in their community, where we were able to label 11 as partisan. Results can be found in under appendix in TABLE IX.

The reason only 11 could be labeled as partisan was because 9 of them came from communities with no inherent political affiliation.

From TABLE IX. we can see that there are communities which aren't connected to the largest component. As such, we cannot say that these organizations are partisan or non-partisan as there is no data indicating their retweeters lean one way or the other on the political spectrum.

EXAMPLE: UNPROFILED COMMUNITY 18 WHICH CONSISTS OF @GENSQUEEZE AND THEIR RETWEETERS AND UNPROFILED COMMUNITY 19 WHICH CONSISTS OF @REALWOMENCANADA AND THEIR RETWEETERS WERE NOT PART OF THE LARGEST CONNECTED COMPONENT. @GENSQUEEZE CLAIMS TO BE NON-PARTISAN. WE CANNOT CONFIRM THAT THEY ARE NONPARTISAN AS WE HAVE NO METADATA ON THE RETWEETERS THEMSELVES AND THE RETWEETERS ARE NOT RETWEETING ANY OTHER MP'S OR ACTIVIST ORGANIZATIONS.

Additionally, we cannot say anything about those organizations within a community that is connected to the large component but has no retweeters outside of their community. It could be said that if a retweeter of an Activist Organization also retweets another party then the Activist Organization would be leaning one way or another; however, when we're looking at partisanship, we're looking for groups who have direct interaction with a particular ideology expressed by a party, and so if the retweeters of an Activist Organization only retweet the community the activist organization represents, it does not necessarily indicate any political leaning.

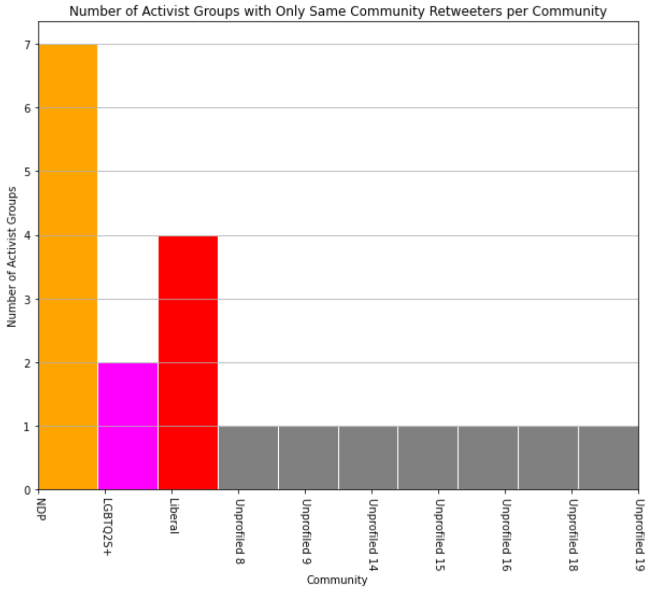

Fig. 16. Number of activist groups

b) Nonpartisan Activist Organizations

There did not exist an organization where all political communities retweeted them.

We were able to identify 3 nonpartisan organizations, with the criteria that they were retweeted by 2 other political communities.

TABLE IX. NONPARTISAN ACTIVIST ORGANIZATION

Twitter Handle	User ID	Users Community	Communities that Retweet User
@DavidSuzukiFDN	14079041	NDP	Liberal, Conservative
@CIJAinfo	21410801	Conservative	Liberal, NDP
@OpenMediaOrg	47452040	NDP	Liberal, Conservative

D. Disassortative Network

The network is disassortative. This means that our high degree nodes are more likely to connect with low degree nodes than high degree nodes. The degree assortativity coefficient r of the entire network is -0.48 which shows us the network is disassortative. TABLE X. details the assortativity coefficients of each community.

TABLE X. ASSORTATIVITY COEDDICIENT R

Community	Assortativity Coefficient r
Conservative	-0.46
NDP	-0.49
Liberal	-0.54
LGBTQ2S+	-0.77
Bloc Quebecois	-0.63
Conservative Activists	-0.80

These initial r value indicate how well the MPs and Activist Organizations in each community connect to each other. A higher r value indicates there are more MPs and AOs connected to one another. The reason for this is MPs and AOs were more likely to have more regular users as retweeters (that were lower degree) than other MPs or AOs.

E. Centrality Measures

A colouring guide for the following visualizations is shown in IV.C. Additionally, triangles represent activist organizations, circles represent normal users, and hexagons represent MPs. The visualizations of the network has been filtered to show only nodes with degree 10 or more.

1) *In-Degree Centrality*

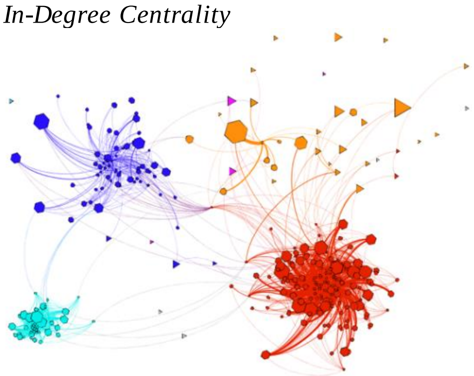

Fig. 17. In-degree centrality visualization

The size of each node shows the in-degree centrality of the node.

a) What it means in the context of our network: By looking at the nodes with high in degree centrality we can see which MPs are interacted with most, indicating the most popular MPs and Activist Organizations.

TABLE XI. IN-DEGREE CENTRALITY TOP 10

Id	Political Affiliation	In Degree Centrality
215632349	NDP	0.0422
14079041	Activist Organization	0.0364
739149720	Conservative	0.0275
22849568	Liberal	0.0275
45976740	Activist Organization	0.0250
1707636642	Liberal	0.0247
29754743	Liberal	0.0238
20199202	Liberal	0.0236
17823761	NDP	0.0219
40550119	Liberal	0.0216

2) *Out-Degree Centrality*

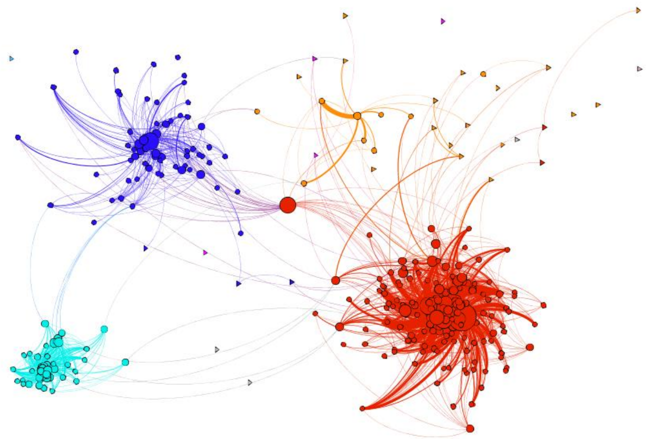

Fig. 18. Out-degree centrality visualization

The size of each node shows the out-degree centrality of the node.

a) What it means in the context of our network:

Nodes with high out-degree centrality help us identify which retweeters are retweeting a lot of different MP's and Activist Organizations. Notably, the nodes with the highest out-degree centrality are the ones with no political affiliation, regular Twitter users.

TABLE XII. OUT-DEGREE CENTRALITY TOP 10

Id	Political Affiliation	Out Degree Centrality
16272844	Unknown	0.0117
1223950760	Unknown	0.0071
1303704709123969024	Unknown	0.0064
2978018901	Unknown	0.0064
1440509126216347656	Unknown	0.0061

1446352422272528385	Unknown	0.0057
89505108	Unknown	0.0050
425710711	Unknown	0.0042
265743906	Unknown	0.0040
2434162616	Unknown	0.0037

3) Undirected Betweenness Centrality

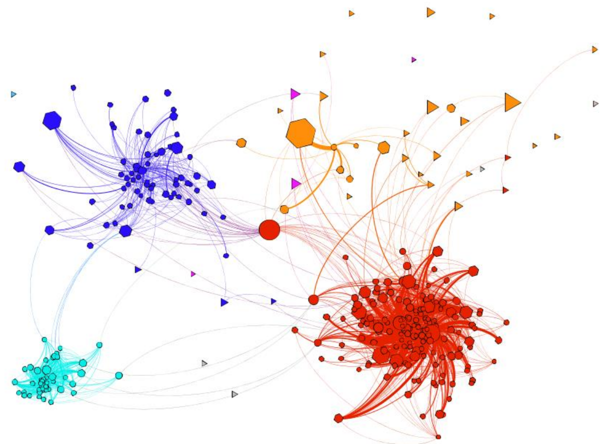

Fig. 19. Undirected betweenness centrality visualization

The size of each node shows the betweenness centrality (undirected) of the node.

a) What it means in the context of our network:

The undirected betweenness centrality gives insight as to what nodes are important in connecting groups together. Nodes with a high betweenness centrality are the ones that are more likely to interact with, or be interacted with multiple different communities

TABLE XIII. BETWEENNESS CENTRALITY TOP 10

Id	Political Affiliation	Betweenness Centrality
215632349	NDP	4958323.66
1446352422272528385	Unknown	3781095.48
16272844	Unknown	3097217.45
14079041	Activist Organization	3091532.81
739149720	Conservative	2776032.75
1707636642	Liberal	2610868.39
1303704709123969024	Unknown	2164853.33
22849568	Liberal	2048458.57
277332906	Activist Organization	193345.00
45976740	Activist Organization	1922583.14

4) Directed Betweenness Centrality

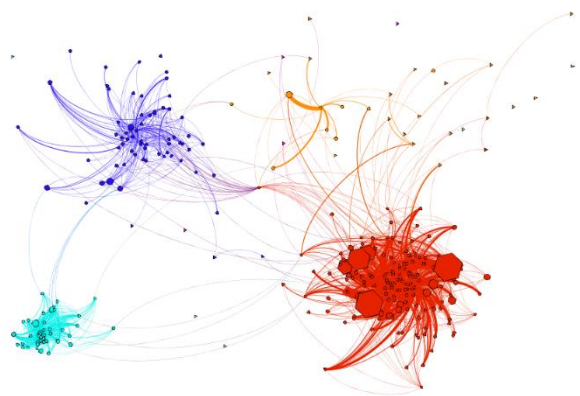

Fig. 20. Directed betweenness centrality visualized

The size of each node shows the betweenness centrality (directed) of the node.

a) What it means in the context of our network:

The undirected betweenness centrality gives us an idea of what MPs or activist groups are working to connect between groups in the network. Given that the edge direction in our network goes from Retweeter, the directed betweenness will show us the hub nodes of a given community (political group).

TABLE XIV. DIRECTED BETWEENNESS CENTRALITY TOP 10

Id	Political Affiliation	Betweenness Centrality (Directed)
40550119	Liberal	8130
45848808	Liberal	7838.83
234550882	Liberal	6193.67
205786669	Liberal	4131.58
22849568	Liberal	3741.5
2344419362	Liberal	1993.83
780323935	Liberal	1564.5
1144289639924207616	Liberal	1550.33
276713213	Liberal	1435
1342125115383939073	Bloc Québécois	1246.42

5) Katz Centrality

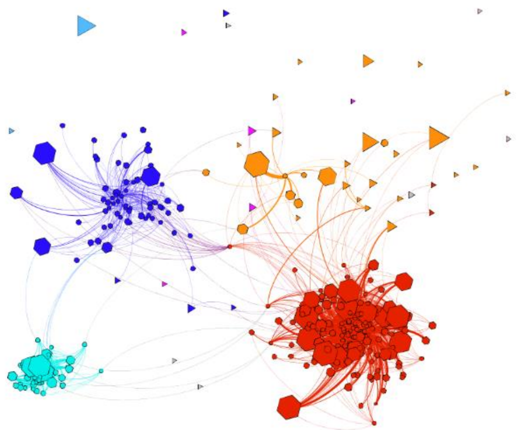

Fig. 21. Katz centrality visualization

The size of each node shows the Katz centrality of the node.

a) Comparison to Null Model: The Katz centrality of our real network increases far less than it does in the random null models. What we would expect to see from a random network is the Katz centrality increasing linearly with the degree of the node. In our real network we see a very slight increase in the Katz centrality of a node as its degree increases, however this is nowhere near the expected behavior in a random network. This tells us that our real network does not follow the structure of a random network and the Katz centralities are dependent on some factor of the structure of the network rather than being random.

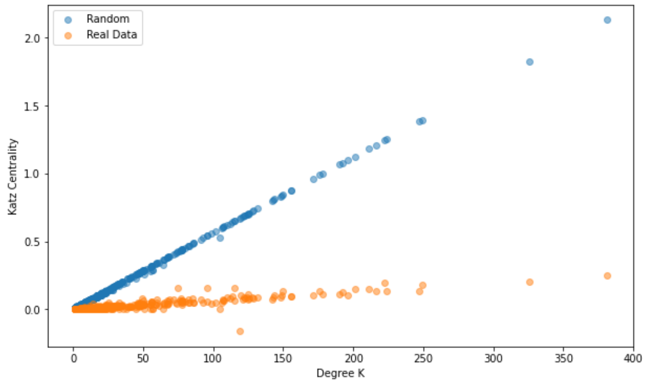

Fig. 22. Katz centrality against the null model

b) What it means in the context of our network:

The Katz centrality can give us some insight as to what nodes are influential in the network. In this case the nodes with a high Katz centrality are the nodes that are influential upon other nodes in the network. Among these nodes is primarily MPs and some activist organizations.

TABLE XV. KATZ CENTRALITY TOP 10

Id	Political Affiliation	Katz Centrality
215632349	NDP	0.245
14079041	Activist Organization	0.204
1707636642	Liberal	0.199
22849568	Liberal	0.184
631988630	Liberal	0.161
1318671397560979456	Liberal	0.157
1899063048	Liberal	0.156
20199202	Liberal	0.151
40550119	Liberal	0.151
739149720	Conservative	0.136

F. *Identifying Community Bridging Users*

When we had considered only the centrality measures of the network, we had trouble distinguishing which nodes were truly important in diversifying the network and helping to bridge the gaps between parties. By introducing community analysis, we can see the divides in the network and by also taking our centrality measures into account we can start to figure out what nodes are bridging the gaps between communities.

When we are looking at the nodes in this network, we see a small subset of the nodes which have a very high betweenness centrality and are well connected with other communities in the network. From these nodes, we can begin to identify bridge nodes that may help an individual diversify their twitter feed without traveling too far outside of their local community. To figure out what these bridge nodes are we must properly define what a bridge node is and how we can identify them. They are defined as follows:

1) Strong Bridges

Strong bridge nodes are identified as either:

- An MP or activist organization who has 4 communities retweeting them
- A user that retweets 4 communities and has a high betweenness (in the top 99% non-inclusive of 0 betweenness)

Table showing regular bridge examples can be seen in TABLE XXIII. under the appendix.

2) Regular Bridges

Regular bridge nodes are identified either:

- An MP or AO who has 3 communities retweeting them and a relatively high betweenness (top 70% - non-inclusive of 0 betweenness)
- A user who retweets 3 communities and has high betweenness (top 95% - non-inclusive of 0 betweenness)

Table showing regular bridge examples can be seen in TABLE XXIV. under the appendix.

3) Weak Bridges

Weak bridge nodes are identified either:

- An MP or AO who has 2 communities retweeting them and a high betweenness
- A user who retweets at least 2 communities and has very high betweenness (top 90% - not inclusive of 0 betweenness)

Table showing regular bridge examples can be seen in TABLE XXV. under the appendix.

G. *Information Diffusion*

1) Initial Starting Nodes

Initial starting nodes were chosen under the following criteria:

The node is an Activist Organization or MP who has been retweeted by another Activist Organization or MP. This resulted in 91 potential activation nodes from 5 different communities.

2) Experiment Summary

We proposed two experiments. The first was to mimic normal tweets in which two random nodes in the initial activation node set were identified and activated to see what communities information would pass through normally. The second experiment was where the starting nodes were all in on an announcement, and we disseminate information through them to find out who sees the message. This could mimic an announcement that all MPs were supposed to send out, but some forgot until they saw it on Twitter so just retweeted their fellow party member.

These two experiments were run with 5 of the 6 profiled communities as that's where the activation nodes were from. These were done to find out how many people could see a tweet and what percent of those who saw it were outside of the originating community.

a) Experiment 1

TABLE XVI. RESULTS OF COMMUNITY EXPERIMENT 1

Experiment		Bridge		
		Strong	Regular	Weak
Liberal	Retweeters	ProChristopher	0	johangreg 44thParliament
Liberal	MP's & AO's with wide demographics	0	0	SeamusORegan Carolyn_Bennett karinagould HonAhmedHussen
Conservative	Retweeters	0	trueintegrity87	44thParliament
Conservative	MP's & AO's with wide demographics	0	0	AlexRuff17
NDP	Retweeters	ProChristopher	0	0
NDP	MP's & AO's with wide demographics	Alexboulerice CharlieAngusNDP	0	GrandCouncilT3
LGBTQ2S+	Retweeters	ProChristopher	0	0
LGBTQ2S+	MP's & AO's with wide demographics	0	0	egalecanada
Bloc Québécois	Retweeters	0	0	0
Bloc Québécois	MP's & AO's with wide demographics	0	0	0

TABLE XVII. DIFFUSION TABLE

Diffusion	Number of Initial Activation Nodes	Total Number of Activated Nodes	Maximum Time at which a Node was Activated	Communities Impacted in Diffusion
Liberal	2	145	2	Liberal, Conservative, NDP, Bloc Quebecois, LGBTQ2S+, Unprofiled 10, Unprofiled 11
Conservative	2	155	2	Liberal, Conservative, NDP
NDP	2	340	2	Liberal, Conservative, NDP, Bloc Quebecois, LGBTQ2S+
Bloc Quebecois	2	230	3	Bloc Quebecois, Liberal, Conservative
LGBTQ2S+	2	189	2	LGBTQ2S+, NDP, Liberal

The liberal party is impacted in every community diffusion. NDP and Conservatives are only impacted 80% of the cascades, and Bloc Quebecois is only affected in 60% of the cascades. Interestingly, information that was diffused from the LGBTQ2S+ group was only diffused into itself the NDP and Liberal communities. This indicates that members of these communities are interacting more with the ideas from the LGBTQ2S+ community. The Liberal cascade was able to reach Unprofiled community 10 and 11. Unprofiled community 10 mainly consists of those who interact with charity and environmental groups.

From TABLE XVI. We can see the users ProChristopher and 44thParliament were cascaded upon in multiple experiments. Specifically, when running the LGBTQ2S+, NDP and Liberal information was cascaded to ProChristopher who is a strong bridge node in the network. The Liberal and Conservative cascade experiments saw information cascaded to 44thParliament who acts as a weak bridge node in the network.

TABLE XVIII. COMPARISON TO NULL MODEL DIFFUSION

Diffusion	Number of Activated Nodes on Real Network	Null Model: Average Number of Activated Nodes	Null Model: Standard deviation of Number of Activated Nodes
Liberal	145	9.25	1.50
Conservative	155	3.20	0.55
NDP	340	4.43	0.84
Bloc Quebecois	2	6.85	1.12
LGBTQ2S+	2	3.24	0.64

We can see here that the random networks vary drastically, this is because in the random model the activation node is not always going to link to a large hub to disseminate beyond its list, in fact its highly likely it could just link to other degree 2 or one nodes. This shows that the information cascade is not a result of the degree distribution but of the meaningful link to the hub nodes and beyond that the real network provides.

b) Experiment 2 Results

TABLE XIX. EXPERIMENT 2 RESULTS

Experiment		Bridge		
		Strong	Regular	Weak
Liberal	Retweeters	ProChristopher	davidakin trueintegrity87	johangreg friendscb 44thParliamen
Liberal	MPs & Activists Organizations with wide demographics		OmarAlghabra PamDamoff FP_Champagne	SeamusORegan HedyFry Carolyn_Bennett MarciIen karinagould stbstvdan AHousefather HonAhmedHussen DLeBlancNB PatrickBWeiler
Conservative Cascade	Retweeters		trueintegrity87	CIJAinfo 44thParliament
Conservative Cascade	MPs & Activists Organizations with wide demographics		gerarddeltell Eric_Melillo	PierrePaulHus GaryAVidal AlexRuff17
NDP	Retweeters	ProChristopher		GrandCouncilT3
NDP	MPs & Activists Organizations with wide demographics	alexboulerice CharlieAngusNDP		
LGBTQ2S+ Cascade	Retweeters	ProChristopher		egalecanada
LGBTQ2S+ Cascade	MPs & Activists Organizations with wide demographics			
Bloc Québécois Cascade	Retweeters	ProChristopher		
Bloc Québécois Cascade	MPs & Activists Organizations with wide demographics			

TABLE XX. COMPARISON TO NULL MODEL

Diffusion	Number of Initial Activation Nodes	Total Number of Activated Nodes	Maximum Time at which a Node was Activated	Communities Impacted in Diffusion
Liberal	49	2702	2	Liberal, Conservative, NDP, Bloc Quebecois, Unprofiled 10, Unprofiled 11, Unprofiled 8, LGBTQ2S+
Conservative	18	951	2	Liberal, Conservative, NDP, Bloc Quebecois, LGBTQ2S

				+, Conservative Activists
NDP	9	1123	3	Liberal, Conservative, NDP, Bloc Quebecois, LGBTQ2S+, Unprofiled 13
Bloc Quebecois	13	476	2	Liberal, Conservative, Bloc Quebecois
LGBTQ2S+	2	184	2	Liberal, NDP, LGBTQ2S+

As detailed in TABLE XX. , the liberal party is impacted in every community diffusion. NDP, Conservatives, and Bloc Quebecois are only impacted 80% of the cascades. With adding more initial activation nodes two more groups were able to be reached in the Liberal cascade, one more community from the NDP cascade and the Conservative cascade was able to reach the LGBTQ2S+ and Conservative Activists Community. The communities impacted from both the Bloc Quebecois and LGBTQ2S+ cascades did not change, indicating no other communities interacted with them or if there were, their interaction count (i.e. edge weight) was not sufficient for them to be activated.

The users ProChristopher and 44thParliament were cascaded upon in multiple experiments. Specifically, when running the LGBTQ2S+, NDP and Liberal information was cascaded to ProChristopher who is a strong bridge node in the network. The Liberal and Conservative cascade experiments saw information cascaded to 44thParliament who acts as a weak bridge node in the network.

We expected @johangreg to be activated in both the liberal and NDP cascades as well, however due to the fact they only retweeted the NDP community once they were not activated and didn't receive the information.

a) Overlapping Diffusions

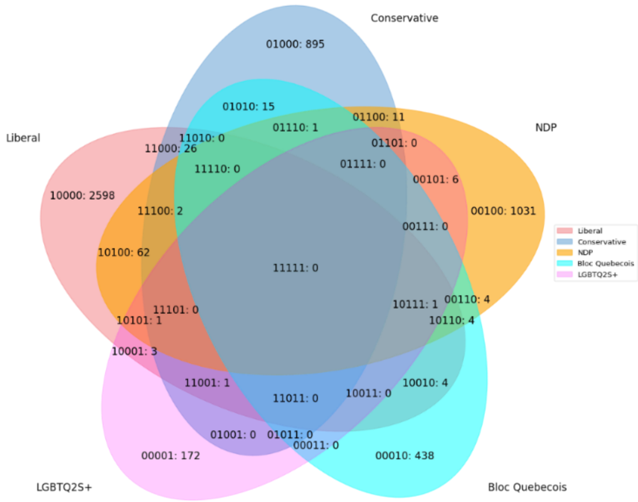

Fig. 23. Experiment 2 diffusion overlaps

Fig. 23 describes how many members received common information from the different diffusions. The sequence of 1's and 0's in the labels indicate which sets are included in the section and the number on the other side of the ':' indicates how many people received information from all those diffusions. As an example, Label 10101:1 means that there was 1 person who received information starting from the Liberals, NDP, and the LGBTQ2S+ Community.

Fig. 23 demonstrates that some individuals are receiving information from more than one source in these cascades. Since each user only belongs to one community, we know that some users are receiving information from outside their community. Fig. 23 further illustrates what pairs of information are likely to reach the same people. An example of this is more individuals received information from both the Liberals and NDP than Liberals and Conservatives. This gives us an indication that there are more nodes retweeting both the Liberal and NDP party than the Liberals and Conservatives.

Fig. 23 does not encapsulate the communities of those impacted. (i.e., there could be a user from the NDP who was not activated in the NDP diffusion but was activated in both the Liberal and conservative).

Looking at Fig. 23 we notice there is an individual person who received information from all cascades except the information originating from the conservative community. This person was @ProChristopher who was identified to be a strong bridge node from the Liberal Community.

b) Information cascade legend

The following visualizations display the diffusion of information given different activation nodes:

Initial activation nodes
Nodes activated at Time t=1
Nodes activated at Time t=2
Nodes activated at Time t=3

These visualizations below provide us with some insight as to how the information travels mostly in the community but branches out towards other communities.

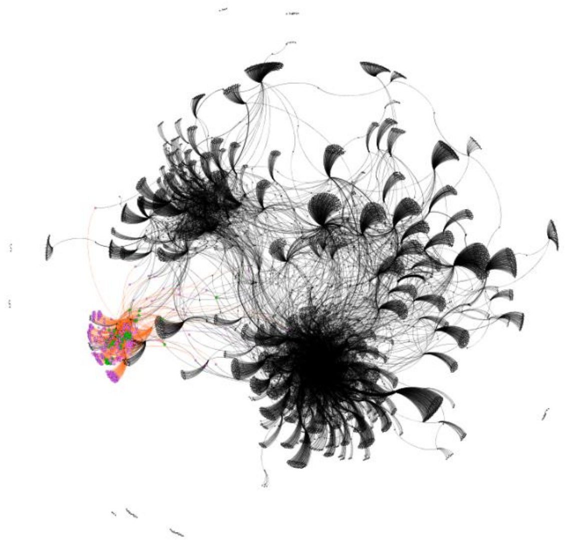

Fig. 24. Bloc Québécois Information Cascade

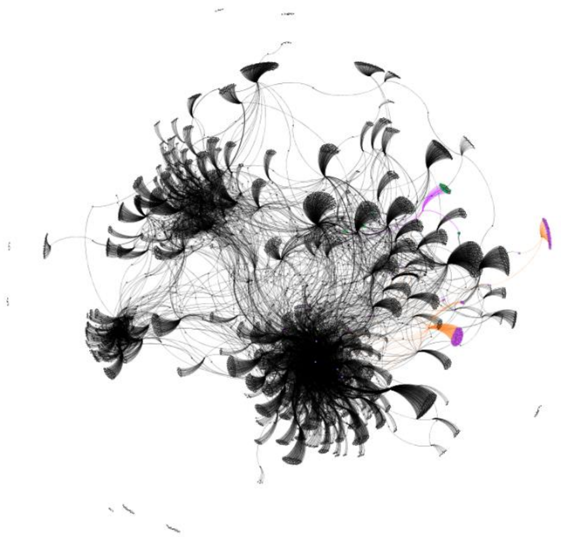

Fig. 25. LGBTQ2S+ information cascade

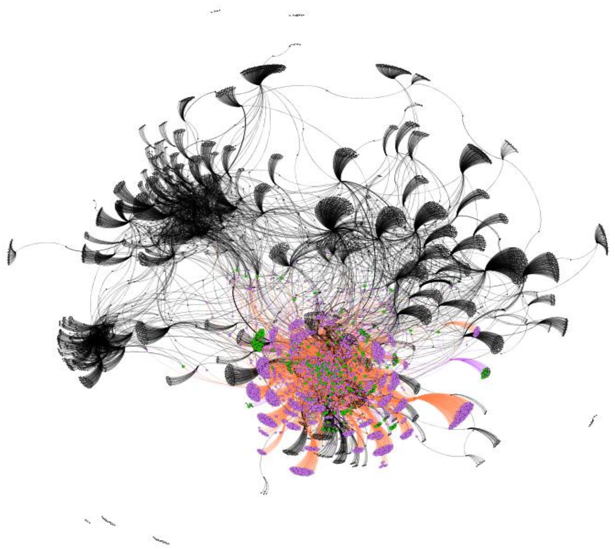

Fig. 26. Liberal information cascade

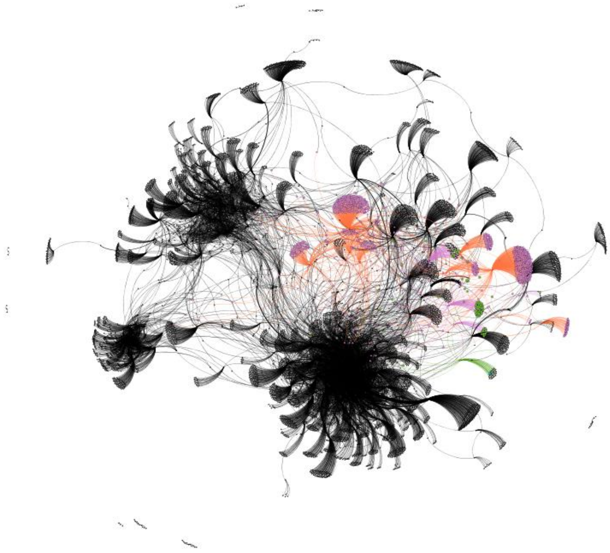

Fig. 27. NDP information cascade

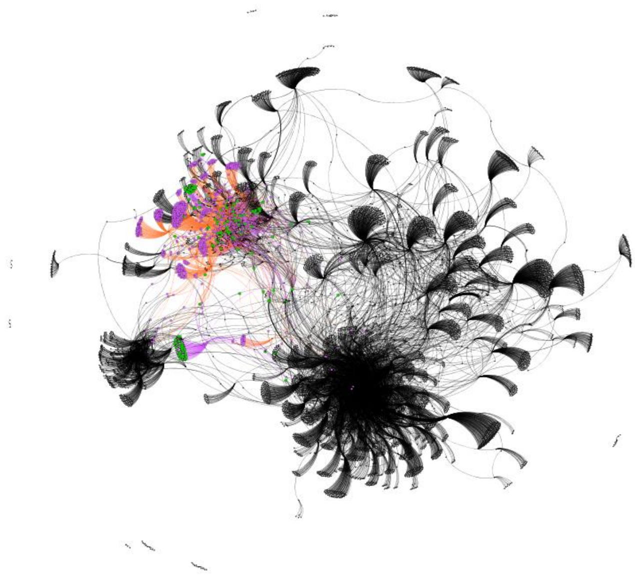

Fig. 28. Conservative information cascade

TABLE XXI. COMPARISON TO NULL MODEL

Diffusion	Number of Activated Nodes on Real Network	Null Model: Average Number of Activated Nodes	Null Model: Standard deviation of Number of Activated Nodes
Liberal	2702	87	3.76
Conservative	951	31	2.27
NDP	1123	17	1.56
Bloc Quebecois	476	34	2.52
LGBTQ2S+	184	3	0.64

Like the previous experiment the random models yielded a much lower activation number due to the fact the that there's no guarantee the activation node will connect to a large hub node.

V. Discussion

A. Research Question 1

To identify feedback loops we first clustered all our data into different communities as shown in Fig. 5. Here we selected the top 6 communities as the other 15 were too small to be of significant importance. These 6 distinct communities are actively recirculating tweets inside them, which is a suspect for designating it as an echo chamber due to the high amounts of confirmation bias and social influence. [4] While this shows where the echo chambers are we might take interest in who the biggest offenders are.

We define breaking the echo chamber as interaction with other communities. For one community to interact with another shows there is attempted to go against confirmation bias, even though social influence is not going anywhere. To quantify the degree to which this happens we normalized the out-of-community interactions with the total number of members in that community, which is shown in Fig. 15. We can see that the Conservative community strives the best against confirmation bias at a 5.57% out-of-community interaction rate, while the Liberal community is almost half as effective with a 2.78% rate. Referencing TABLE XV. Katz centrality top 10 we could see a potential cause for the low out-of-community interaction rate, whereby the top 10 list on Katz centrality is dominated by 7 Liberal community members, showing a very strong influence that is preventing them to break this echo chamber, this influence is also confirmed in TABLE XI. , where half of the list (5) are Liberal community members.

B. Research Question 2

We were able to identify activist organizations as partisan by observing if the activist organization tweets were only retweeted by members of the same community, and that community was identified to be political based on the retweeters connection to a MP. The only AO's identified this way were from the Liberal and NDP community and none from the conservative community. This indicated to us that we needed to look beyond just the retweeters to identify political leaning of Activist Organizations.

Another way to determine the leaning of an Activist Organization was to look at the community it was placed in and the strongest connections it has to other communities. When looking at the LGBTQ2S+ community we can see there is a higher percentage of connections between the Liberal and NDP party than the Conservatives or Bloc Quebecois from TABLE VI. ; using the strength of communities, we can identify the activist organizations as leaning towards the Liberal and NDP values. This also accounts for the AO's in the Conservative Activists community, where we can see the Conservative Activist Community interacting with the Conservative community proportionally more than the NDP and Bloc Quebecois communities. Thus, it can be stated that the AO's in the Conservative Activists community lean towards the Conservative party values while the LGBTQ2S+ community leans towards the values of the NDP and Liberal party.

In terms of non-partisanship, we found that none of the two AO's from our six communities interacted with the other four non-AO communities equally; however, there were three AO's outside the six communities identified in TABLE IX. Nonpartisan activist organization, showing that the three AO's interacted with three political party communities with the exception of Bloc Quebecois (explained in limitations), hence we labelled them as non-partisan. While these AO's interacted with multiple communities, they were still heavily dominated by their own community, which could be an indication that while non-partisan they still profess a view congruent with a particular ideology.

C. Research Question 3

The betweenness centrality we found when we undirected the graph pointed to nodes that could spread information well throughout the network. By undirecting the network, it allowed us to see which nodes are potential spreaders of information to other sections of the network, based on a mutual retweeting relationship. The hopes were finding nodes that didn't necessarily have a high out-degree centrality, but were integral in providing information between two communities, so that if someone were to follow them, they would get exposure to a diverse set of tweets.

Once we had our bridges identified, we were able to test and see which nodes were being cascaded into, from different starting communities. The findings of these experiments are twofold. First, the retweeter bridges that were activated in multiple cascades prove a connection between the two communities; that is, if another node from another community was activated then we would be seeing an example of the cross-community interaction, leading to breaking confirmation bias. Secondly, the MPs and AOs with diverse followings allowed us to see if information was disseminated to the more diverse sections of parties. In finding the MP and AO bridges with diverse followings, it allowed us to see if information was being spread to more neutral areas, creating potential for information to move into other communities.

The first experiment allowed us to see if normal twitter interactions (where a tweet with information starts at one or two sources) resulted in particular users outside the community being activated and received the information.

We found through our first experiment that the amount of interaction with another community is integral. Information cannot flow through to another community if the connection isn't strong enough, which is to say the users aren't really retweeting another community, leading to a lot the information being likely to stick in the community. As an example, we had an expectation that @johangreg from the liberal community would be in the LGBTQ2S+ and NDP cascade, but due to the fact that they barely interacted with other parties, the information never actually made it out to them and they were only identified as receiving information when it originated in the community they were from. This was true for both our first and second experiment where we were looking to see if the information being spread by multiple people at once would impact the number of individuals that information was cascaded to. Individuals would be given more chances to be activated and thus we thought this would increase community spread. Overall, we saw more diffusion to more communities when more users were sending out the same information at the start.

Interestingly, not many people identified as bridging candidates were impacted by information outside of their community. This points to the importance of the edge weights and how if you retweet a group more than once in a 3-week period the more likely you are to retweet that community's tweet and disseminate that information. Ultimately, with the information cascade we were able to show that information starting out in one community reached other communities through some of the bridge nodes we identified. If one were to follow the bridge users, we identified they would be more likely to get information from more than one perspective.

D. Limitations

Some of the biggest MP's were part of the Liberal party, which meant we could not collect sufficient data to be considered for analysis. Omitting big names like these could have had an impact in showing a low out-of-community interaction rate in Fig. 15. It is worth mentioning that other party's MP's were also omitted for the same reasons, but to a lesser extent as they are not the majority.

We suspect Bloc Quebecois were isolated in terms of their interactions outside their party. The times they were interacting outside their party it was with other MPs, not AO's. This is likely because the list of AO's was sourced from Wikipedia [13], which detailed Canadian AO's but, did not include francophone AO's.

Although our research shows AO's retweets as a group demonstrates partisan and non-partisan political favouritism, which answers our 2nd research question, there are plenty of limitations to this method in terms of finding key nodes to break echo chambers. We could have used out-degree centrality instead of betweenness centrality to find these bridge nodes that provide multiple perspectives. Combined with community structure of selecting nodes with interactions with more than one community would also have provided a good basis for selecting important diversifying nodes. We could have potentially missed out on some nodes by omitting this method.

E. Future considerations

1. Having metadata on our activist organizations and retweeters would allow us to track influence and homophily within our network. In further research if demographic information is available that would be a strong area to investigate.
2. Capacity for handling the retweeters of retweeters. In our current approach we were limited in fully understanding the value of an individual retweeters, as we didn't know how their retweet was spread. This was due to Twitter API license limitations we had, which prevented us from getting 2nd degree retweeters or more.
3. Including French Activist Organizations to get better representation of French Canada.

VI. Conclusion

Initial analysis of the network revealed that the network was disassortative, containing many nodes of degree 1. This came as a result of not being able to track users that are retweeting retweets. Through our analysis of the network, we have been able to identify communities within it, and assign users to these communities according to their interactions with MPs and activist organizations. In doing so we have been able to assess the partisanship of activist organizations and identify organizations that do not fall into any of the typical political leanings. We were able to identify potential sources of feedback loops, however we could not truly identify a feedback loop taking place without assessing the retweeters of a retweet. Finally, we were able to identify users that have a wide breadth of interaction in the network and users that are good bridges between communities. Additionally, we showed that information originating from different community sources was able to reach these bridge nodes showing that they truly do have a wide breadth of interaction.

VII. Appendix

A. Source Code

The source files can be accessed through a shared GitHub repository at: https://github.com/Noah-Giustini/twitterbot

B. Other Tables

TABLE XXII. 20 PARTISAN ACTIVIST ORGANIZATIONS

Twitter Handle	ID	Community
@UNACanada	142731825	Liberal
@CASAACAE	42702616	Liberal
@ccrweb	65666257	Liberal
@CAUS	189349553	Liberal
@TOpublicspace	29545977	NDP
@TransportAction	109144163	NDP
@OptionConso	79260190	NDP
@OCAPtoronto	247412217	NDP
@TOenviro	59686058	NDP
@TI_Canada	3409346380	NDP
@AbortionRights	37044121	NDP

TABLE XXIII. Strong bridge examples

Id	Political Affiliation	Community Id	Betweenness Centrality	Source Community Interactions	Target Community Interactions
196717787	NDP	2	8.16×10^5	0	4
215632349	NDP	2	4.96×10^6	0	4
596787864	None	1	1.32×10^6	4	0

TABLE XXIV. REGULAR BRIDGE EXAMPLES

Id	Political Affiliation	Community Id	Betweenness Centrality	Source Community Interactions	Target Community Interactions

1707636642	Liberal	1	2.61 x 10^{6}	0	3
205786669	Liberal	1	1.01 x 10^{6}	0	3
343059330	Conservative	3	5.46 x 10^{5}	0	3
36133644	Conservative	3	1.44 x 10^{6}	0	3
20199202	Liberal	1	1.88 x 10^{6}	0	3
14079041	Activist Organization	2	3.09 x 10^{6}	0	3
256360738	Conservative	3	4.34 x 10^{5}	0	3
737359208945844224	Conservative	3	1.23 x 10^{6}	0	3
24990450	Bloc Québécois	4	5.96 x 10^{5}	0	3
12034642	None	3	210175.77	3	0
711980580813066240	None	2	736622.89	3	0

TABLE XXV. WEAK BRIDGE EXAMPLES

Id	**Political Affiliation**	**Community Id**	**Betweenness Centrality**	**Source Community Interactions**	**Target Community Interactions**
2891740872	Liberal	1	8.26 x 10^{5}	0	2
40550119	Liberal	1	1.61 x 10^{6}	0	2
61521038	Liberal	1	1.06 x 10^{6}	0	2
22849568	Liberal	1	2.05 x 10^{6}	0	2
25127782	Liberal	1	1.42 x 10^{6}	0	2
1646334073	Liberal	1	7.56 x 10^{5}	0	2
25813888	Activist Organization	1	3.80 x 10^{5}	0	2
1162219997977960449	Liberal	1	3.16 x 10^{5}	0	2
234550882	Liberal	1	2.95 x 10^{5}	0	2
341866567	NDP	2	4.27 x 10^{5}	0	2
16220555	Green Party	2	7.49 x 10^{5}	0	2
21410801	Activist Organization	3	8.19 x 10^{5}	0	2
412708728	Conservative	3	9.06 x 10^{5}	0	2
1957288050	Conservative	3	2.90 x 10^{5}	0	2
1143229947932229632	Bloc Québécois	4	8.13 x 10^{5}	0	2
85428184	Liberal	1	4.72 x 10^{5}	0	2
47338701	Liberal	1	1.38 x 10^{6}	0	2
1942107584	Liberal	1	6.43 x 10^{5}	0	2
720579941184757760	Liberal	1	3.32 x 10^{5}	0	2
1390883714	Liberal	1	2.75 x 10^{5}	0	2
593944500	Liberal	1	3.15 x 10^{5}	0	2
277332906	Activist Organization	2	1.93 x 10^{6}	0	2
24913022	Activist Organization	2	1.02 x 10^{6}	0	2
377588094	Activist Organization	2	4.16 x 10^{5}	0	2
47452040	Activist Organization	2	1.09 x 10^{6}	0	2
2715275551	Activist Organization	2	6.85 x 10^{5}	0	2
17823761	NDP	2	1.52 x 10^{6}	0	2
266855812	NDP	2	9.09 x 10^{5}	0	2

175259033	NDP	2	5.14 x 10^5	0	2
1382491059247255554	Conservative	3	4.22 x 10^5	0	2
172004509	Conservative	3	2.07 x 10^5	0	2
1086084557009575936	Conservative	3	7.06 x 10^5	0	2
335361064	Bloc Québécois	4	5.81 x 10^5	0	2
2800741820	Bloc Québécois	4	2.84 x 10^5	0	2
211564038	Activist Organization	5	1.40 x 10^6	0	2
119925381	Activist Organization	5	1.40 x 10^6	0	2
45976740	Activist Organization	6	1.92 x 10^6	0	2
3299711869	Activist Organization	8	2.38 x 10^5	0	2
158095776	Activist Organization	9	4.88 x 10^5	0	2
218952640	Activist Organization	10	3.31 x 10^5	0	2
40246161	Activist Organization	11	2.33 x 10^5	0	2
16272844	None	1	3.10 x 10^6	2	0
1446352422272528385	None	1	3.78 x 10^6	2	0

REFERENCES

[1] Government of Canada, "Discover Canada - Federal Elections - Canada.ca," *Government of Canada*, 2012. https://www.canada.ca/en/immigration-refugees-citizenship/corporate/publications-manuals/discover-canada/read-online/federal-elections.html (accessed Nov. 17, 2021).

[2] E. G. West and S. L. Winer, "The Individual, Political Tension, and Canada's Quest for a New Constitution," *Can. Public Policy / Anal. Polit.*, vol. 6, no. 1, p. 3, 1980, doi: 10.2307/3550064.

[3] C. Metzgar, "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," *Prof. Saf.*, vol. 58, no. 9, p. 44, 2013.

[4] M. Del Vicario *et al.*, "The spreading of misinformation online," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 3, pp. 554–559, 2016, doi: 10.1073/pnas.1517441113.

[5] M. Del Vicario, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi, "Modeling confirmation bias and polarization," *Sci. Rep.*, vol. 7, no. July 2016, pp. 1–9, 2017, doi: 10.1038/srep40391.

[6] InternetLiveStats.com, "Twitter Usage Statistics - Internet Live Stats." https://www.internetlivestats.com/twitter-statistics/ (accessed Nov. 18, 2021).

[7] T. A. Small, "What the hashtag?: A content analysis of Canadian politics on Twitter," *Inf. Commun. Soc.*, vol. 14, no. 6, pp. 872–895, 2011, doi: 10.1080/1369118X.2011.554572.

[8] M. C. J. Stoddart, D. B. Tindall, J. Smith, and R. Haluza-Delay, "Media Access and Political Efficacy in the Eco-politics of Climate Change: Canadian National News and Mediated Policy Networks," *Environ. Commun.*, vol. 11, no. 3, pp. 386–400, 2017, doi: 10.1080/17524032.2016.1275731.

[9] W. Chen, D. Pacheco, K. C. Yang, and F. Menczer, "Neutral bots probe political bias on social media," *Nat. Commun.*, vol. 12, no. 1, Dec. 2021, doi: 10.1038/s41467-021-25738-6.

[10] Student, "python matplotlib histogram specify different colours for different bars [duplicate]," *Stackoverflow*, 2018. https://stackoverflow.com/questions/49290266/python-matplotlib-histogram-specify-different-colours-for-different-bars (accessed Dec. 10, 2021).

[11] Cognitiveclass.ai, "Machine Learning with Python," *Cognitive Class*, 2021. https://courses.cognitiveclass.ai/courses/course-v1:CognitiveClass+ML0101ENv3+2018/course/ (accessed Dec. 10, 2021).

[12] jeremycarroll Tctianchi, JosephRyanPeterson, TC-MCDEX, orangeSi, "pyvenn." 2020, [Online]. Available: https://github.com/tctianchi/pyvenn.

[13] Wikipedia, "Category:Political advocacy groups in Canada," *Wikimedia Foundation Inc.*, 2019. https://en.wikipedia.org/wiki/Category:Political_advocacy_groups_in_Canada (accessed Oct. 15, 2021).