



Analysis of how Canada and the top 5 Canadian  
Banks are celebrating Canada's 150<sup>th</sup> anniversary  
using social media (Twitter).

---

---

*Student Name: Hira Fatima*  
*Course Title: DS8006 – Social Media Analytics*

*Assignment: Final Project*  
*Due Date: April 3, 2017 (11:59 PM)*

Hira Fatima  
Master of Data Science and Analytics | RYERSON UNIVERSITY

# Table of Contents

---

1. SUMMARY.....	2
2. PROBLEM STATEMENT & DATA SELECTION.....	3
Real World Problem? .....	4
3. PREVIOUS WORK / LITERATURE REVIEW .....	5
4. DATA COLLECTION & PREPROCESSING .....	6
5. ANALYSIS.....	8
#Canada150 .....	8
Top Canadian Banks.....	11
6. CONCLUSION.....	12
Use of this Analysis in Real World by Decision Maker.....	12
Lessons Learned .....	12
7. REFERENCES .....	14
8. APPENDIX A - #Canada150 .....	15
9. APPENDIX B - #Make150Count (RBC).....	19
10. APPENDIX C - #TDCommonGround (TD BANK) .....	22
11. APPENDIX D - #StandForCanada (CIBC) .....	25
12. APPENDIX E – Additional Stop Words .....	29

# 1. SUMMARY

---



The aim of this project is to assess how Canadians are celebrating Canada's 150<sup>th</sup> anniversary using social media. Social media does a great job at connecting everyone together. Just by using #Canada150, individuals, organizations, government officials, and anyone and everyone who wants to join the conversation can become a part of the extensive network. As we see in this project, the text/content analysis and social network analysis conveniently revealed the most influential people/organizations in the network for #Canada150. It allowed us to understand what people are talking about and what is most important to them by their way of liking or retweeting tweets. It should be no surprise that Justin Trudeau was the most influential individual in the network for #Canada150 followed by Canada's official handle, @Canada and Library and Archives Canada. The only for-profit organization that had the highest in-degree and out-degree centrality in the network was WestJet Airlines.

The analysis was further extended to assess the top 5 Canadian banks to evaluate how actively they are involved on social media to celebrate Canada's 150<sup>th</sup> anniversary. Only 3 of the top 5 banks currently have a designated hashtag and a campaign for Canada's 150<sup>th</sup> anniversary; therefore, the scope of this analysis was revised to comparing these 3 banks only, namely, RBC, TD and CIBC. Since it is also CIBC's 150<sup>th</sup> anniversary, CIBC has been the most active on Twitter and has the most interaction with Canadians on social media (Twitter) as opposed to other banks.

Twitter was used as the main social media platform for this project and the analysis was performed using R and Netlytic.

## 2. PROBLEM STATEMENT & DATA SELECTION

---

The goal of this project is to perform text and social network analysis on the official hashtag being used for Canada's 150th anniversary to answer questions such as:

What kind of people/organizations are involved in this network?

Who is leading the network?

Who are the most influential accounts in this network?

How does the network structure look like?

Since everyone all over Canada is using this official hashtag, is everyone talking about the same topic or is there a higher number of sub-cliques talking about different things?

The Government of Canada's [website](#) confirmed that [#Canada150](#) is being used on Twitter as the official hashtag to celebrate its 150th anniversary and therefore this hashtag was used to retrieve tweets for this project.

In addition to the analysis of this main hashtag, the top 5 banks in Canada were also studied to assess how these financial institutions are celebrating Canada's anniversary using social media, specifically, Twitter and which one of these is most active on social media. Websites for these 5 Canadian banks namely, TD, RBC, Scotiabank, CIBC and BMO were screened to find if they have any designated hashtags. Of these, only [TD](#), [RBC](#) and [CIBC](#) have official designated hashtags and initiatives to celebrate

Canada's 150th anniversary (as of Mar 31, 2017); the other two banks (Scotiabank and BMO) do not have any official hashtags announced yet.

The following table summarizes the hashtags for each bank and their message for Canadians.

Canada #Canada150	Make the most of your Canada 150! <a href="#">LINK</a>
TD BANK #TDCommonGround	"We're celebrating Canada's 150th anniversary by supporting 150 grassroots initiatives. Join us by sharing and participating in the #TDCommonGround Project." <a href="#">LINK</a>
RBC #Make150Count	"Let's Make Canada's 150th Anniversary Count. Tell us how you will #Make150Count. We're inspiring thousands of young Canadians across the country to make a difference in their communities by giving them \$150, no strings attached, and asking them to do something great with it." <a href="#">LINK</a>
CIBC #StandForCanada	"In 2017, we celebrate the 150th anniversary of both Canada and CIBC. We've been standing for all Canadians since 1867. Given this epic occasion we're collecting the thoughts and feelings of Canadians by creating a digital time capsule. We invite you to help shape the narrative by creating a personalized meme using #StandForCanada." <a href="#">LINK</a>

## Real World Problem?

The first part of this analysis can be used to summarize how Canada is celebrating its 150<sup>th</sup> anniversary using social media, who are the most influential individuals/organizations, how actively are the Canadian's involved, which organizations are most active and what everyone has been talking about the most. The second part of this analysis can be used to compare the top 5 banks and understand how strong their initiatives are for Canada's 150<sup>th</sup> anniversary on a social platform (Twitter). The analysis will also reveal the involvement and popularity of these banks amongst Canadians.

### 3. PREVIOUS WORK / LITERATURE REVIEW

---

The main reference for this project was used from the article called “Enabling Community Through Social Media” written by Gruzd A. and Haythornthwaite C. in 2013. This article provides an in-depth social network analysis of an online community formed using #hscmca [1]. In this project, similar analysis was performed on four hashtags. This included analyzing the name network graph, discovering any leaders or influential people/organizations based on their in-degree and out-degree centrality, and actor/professional roles [1]. A similar article called “The Multiple Facets of Influence: Identifying Political Influentials and Opinion Leaders on Twitter” written by Dubois, E., & Gaffney, D. in 2014 provides a comparable analysis of content and social network analysis to identify influential individuals [2]. One more article used as a reference that provided parallel information was written by Xu, W. W., Sang, Y., Blasiola, S., & Park, H. W in 2014, called “Predicting opinion leaders in Twitter activism networks: The case of the Wisconsin recall election”. An important concept taken from this article was the finding that tweets posted by organizations often have greater influence than those by individual users as we see in the case of WestJet airlines (discussed further in the Analysis section) [3].

Expert blogs were used for technical reference to formulate the R code including but not limited to Andreas Ose Marthinussen’s blog called, “Analyzing the US election using Twitter and Meta-Data in R” for text and content analysis [4]; and, Julia Silge’s blog called “Ten Thousand Tweets” for creating graphs and charts. [5]. Examples from Plotly website were also used as a reference to create some charts [6]. Code from a blog published by Martin in 2015 called “Download, manage and visualize Twitter data with R” was used as a reference to plot tweets on world/Canada’s map for this project [7]. Lastly, Netlytic was used for social network analysis [8].

## 4. DATA COLLECTION & PREPROCESSING

---

Twitter was used as the main social media platform for this project because it was used by the Government of Canada as well as the banks. Although the Government of Canada also used other social media platforms, some of the banks only used Twitter and therefore it made sense to use Twitter only to keep things consistent. For instance, TD and RBC have only publicised a hashtag for Twitter. CIBC on the other hand is using Twitter, Facebook and Instagram for its campaign.

R was used for text analysis and Netlytic [8] was used for the network analysis. 1 week's worth of tweets were collected approximately for the same timeframe (Mar 24 to April 1) from both the sources (R and Netlytic) to maintain consistency of the analysis in this project. Eight datasets were collected for the following hashtags as listed below:

Search Term	Dataset Description	Tweet Count (in R)	Tweet Count (in Netlytic)
#Canada150 lang:en	This hashtag was used to retrieve tweets related to Canada's 150th anniversary as this is the official hashtag being used. Although there is a separate designated hashtag for French (#Canada150e), the language was still restricted to English only so to avoid any noise in data.	19767*	7471
#Make150Count lang:en	Royal Bank of Plaza (RBC) is using this hashtag as their official hashtag to celebrate Canada's 150th anniversary.	58	73
#TDCommonGround lang:en	Toronto Dominion (TD) Bank is using this hashtag as their official hashtag to celebrate Canada's 150th anniversary.	61	76
#StandForCanada lang:en	Canadian Imperial Bank of Commerce (CIBC) is using this hashtag as their official hashtag to celebrate Canada's 150th anniversary.	418	485

Same analysis was performed on all four hashtags. First off, the datasets were converted to a data frame format for further processing. The tweets were then cleaned up to get rid of punctuations, special characters and other noise in the data. For text analysis, stop words were also removed and the list of stop words was modified for each dataset separately to remove unnecessary words in each dataset. The list of stop words is provided in the Appendix E.

\*The dataset for #Canada150 was very large and was giving issues with memory allocation in R. For that reason, the number of tweets used to find the most frequent words was limited to 10,000 in R.



## 5. ANALYSIS

### #Canada150<sup>1</sup>

The text and content analysis of #Canada150 revealed that Canadians are actively involved in celebrating Canada's 150<sup>th</sup> anniversary as shown by the number of tweets rendered for #Canada150. More than 62% of the tweets were retweets confirming that Canadians are busy sharing and retweeting posts with this



hashtag (Ref: FIGURE 2 in Appendix A). As shown in FIGURE 1 (Appendix A), there was a spike in tweets on April 1. This spike was caused by Justin Trudeau's tweets regarding his favourite beer as shown here. This tweet became viral quickly leading to a spike in the overall number of tweets

between April 1<sup>st</sup> and 2<sup>nd</sup>. Justin Trudeau has been actively using #Canada150 to show his spirit for Canada's 150<sup>th</sup> anniversary. This is also evident in further analysis where he has the highest in-degree centrality in this network. Justin Trudeau also has the highest number of liked post in this network as shown in FIGURE 3 (Appendix A).

Diameter	• 22
Density	• 0.000221292911785
Reciprocity	• 0.034353529044347
Centralization	• 0.127080085212171
Modularity	• 0.893755184064952

FIGURE 4 (Appendix A) shows a high-level (zoomed-out) view of the name network for #Canada150. The macro network measures given on the left here demonstrate that this network has high modularity and low centralization. This is in line with the general

<sup>1</sup> Refer to Appendix A that has all the charts and graphs for #Canada150

trend that if a network graph has high modularity, it has low centralization as shown in FIGURE 5 (Appendix A). This is because although everyone is mentioning #Canada150 in the network, different cliques are talking about different things depending on their interest. The network was analyzed at different zoom levels to look for top individuals/organizations with the highest in-degree and out-degree centrality (Ref: FIGURE 6 and 7 in Appendix A). The table below summarizes the top accounts with highest in-degree and out-degree centrality.

#	User Name	Description	Why Influential	Selection Criteria
1	justintrudeau	Account run by the 23rd Prime Minister of Canada and staff	This is the account of the Prime Minister of Canada and it is no surprise that it has a high in-degree centrality for #Canada150.	In-degree
2	canada	Showcasing Canada to the world - français: <a href="http://international.gc.ca/socialmedia">@AuCanada</a> <a href="http://international.gc.ca/socialmedia">http://international.gc.ca/socialmedia</a>	This is the official handle of Canada and it makes sense that this handle has a high in-degree centrality as #Canada is often included in the tweets	In-degree
3	WestJet	<a href="#">#OwnersCare</a> and we're online and available to help 24/7. Please send personal info via direct message only. Tweet your travels with <a href="#">#WestJet</a> and <a href="#">#MostCanadian</a>	WestJet claims to be the most Canadian airline and has been actively involved with celebrating Canada's 150 <sup>th</sup> anniversary.	In-degree
4	libraryarchives	Library and Archives Canada preserves your documentary heritage. Terms of use: <a href="http://goo.gl/EjGP2Q">http://goo.gl/EjGP2Q</a> En Français : <a href="#">@BiblioArchives</a>	This is the official handle of Library and Archives Canada and has been sharing posts about Canada on a regular basis as a way of celebrating Canada's 150 <sup>th</sup> anniversary.	In-degree
5	Canada150music	Celebrating Canada's 150th Anniversary! <a href="#">#Searchlight</a> <a href="#">#Canada150Music</a> <a href="#">#CanadaMusic150</a> <a href="#">#Canada150th</a> <a href="#">#Canada150</a> <a href="#">#CMW2017</a> <a href="#">#JunoAwards</a> <a href="#">#Canada</a> <a href="#">#TuneOfTheDay</a>	At 11:40 PM EST on December 31st, 2016 Jason & Mandi LaFauci registered Canada150 Music via Facebook, Twitter, Instagram and YouTube. More info at <a href="http://canada150thmusic.ca">canada150thmusic.ca</a> This is an initiative by two individuals to celebrate Canada's 150 <sup>th</sup> anniversary by cherishing Canadian music.	In-degree
6	Canbloghouse	Hi I'm Sandy! <a href="#">#SocialMedia</a> <a href="#">@BConnectedConf</a> ~ <a href="#">#Travel</a> <a href="#">#Tourism</a> Writer <a href="#">@TourOutaouais</a> ~ Freelance	Sandy claims to be one of the biggest bloggers in Canada who writes about Canada. As a passionate Canadian, it only	Out-degree

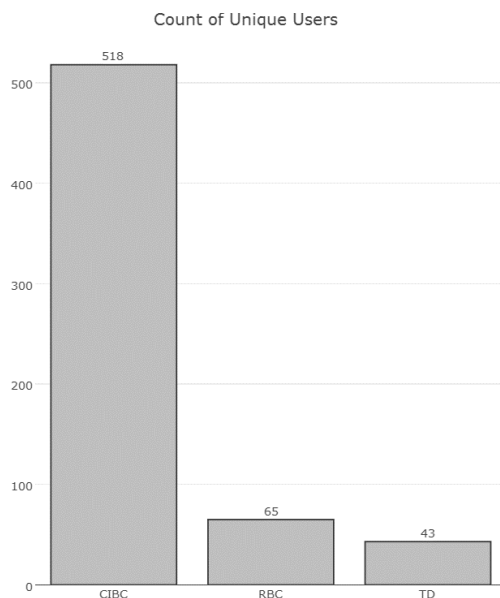
		Writer ~ #Ottawa2017 #Canada150 #Canada #Canadian	makes sense that she would have a high out-degree centrality as she mentioned #Canada150 many times in her tweets.	
7	Canada150music	Celebrating Canada's 150th Anniversary! <a href="#">#Searchlight</a> <a href="#">#Canada150Music</a> <a href="#">#CanadaMusic150</a> <a href="#">#Canada150th</a> <a href="#">#Canada150</a> <a href="#">#CMW2017</a> <a href="#">#JunoAwards</a> <a href="#">#Canada</a> <a href="#">#TuneOfTheDay</a>	Canada150music also has a high out-degree centrality as it references #Canada150 very often as it shares music on a regular basis.	Out-degree
8	Heatherrankinme	Singer-Songwriter. ECMA/Juno Nominated Solo Album, A Fine Line. Member of The Rankin Family, Actor & Co-Owner of The Red Shoe Pub. Booking: jeff@jmatalent.com	As a musician and a passionate Canadian, Heather is also actively involved on twitter with celebrating Canada's 150 <sup>th</sup> anniversary.	Out-degree
9	cheynezmommy	Paula Lambert Loves the Hip, loves Canada, the MN Vikings, have eclectic tastes in following what interests me. Happy to be here.	Paula Lambert is also another music fan and passionate Canadian celebrating Canada's 150 anniversary.	Out-degree
10	WestJet	#OwnersCare and we're online and available to help 24/7. Please send personal info via direct message only. Tweet your travels with #WestJet and #MostCanadian	WestJet is actively involved in building its image as the Most Canadian airline. It therefore makes sense that it is actively tweeting about #Canada150.	Out-degree
11	Johangreg	Green Liberal • Sports fanatic • #LGBTQ • #Aspergers • #cdnpoli • tweet from the heart and soul, and usually inappropriately ~\_(\`ツ)\_/~ johansengreg@hotmail.com	Greg is another passionate Canadian actively celebrating Canada's 150 <sup>th</sup> anniversary.	Out-degree

As mentioned before, it is no surprise that Justin Trudeau and Canada's official account have the highest in-degree centrality followed by Library and Archives Canada. Other than Canadian enthusiasts, bloggers and musicians, there was one only one organization that stood out the most in the network celebrating Canada's 150<sup>th</sup> anniversary and it was WestJet Canada airlines. WestJet has the highest in-degree as well as out-degree centrality in this network. The organization is using this opportunity to advertise

themselves as the most Canadian airline and their endeavours seems to be paying off as their tweets are receiving high reciprocity.

## Top Canadian Banks

In addition to the analysis of #Canada150, the top 5 Canadian banks were also analyzed to compare how each of them is celebrating Canada's 150<sup>th</sup> anniversary and who is in the lead so far on social media (specifically Twitter). Only Twitter was used in this analysis because it was the common social media platform used by 3 out of 5 top banks. Two banks (BMO and Scotiabank) do not have any social media campaigns for Canada's anniversary yet, and only CIBC is using Instagram and Facebook in addition to



Twitter. RBC and TD are only using a hashtag for their campaigns. Tweets were collected during the same period to maintain consistency for comparison. The graph here shows the count of unique users who posted in each of these networks. It is very clear from the graph that CIBC has the most involvement as opposed to any other banks. This year marks the 150<sup>th</sup> anniversary of Canada and CIBC both. Since it is a huge achievement for CIBC as well, the bank is very

involved in this celebration as they are calling everyone to #StandForCanada. Although this year is also BMO's 200<sup>th</sup> anniversary, they have not announced any official campaign for Canada's anniversary yet.

The social network analysis of these three banks reveal that the name networks are very sparse for RBC and TD as there weren't enough tweets rendered. CIBC has the best and most engaged network of all; again, confirming that out of all banks, only CIBC is most actively involved in celebrating Canada's 150<sup>th</sup> anniversary on Twitter with substantial public engagement. The Appendices B to D include a detailed analysis for each of these banks.

## 6. CONCLUSION

---

### Use of this Analysis in Real World by Decision Makers

This analysis can be used in the real world by the Government of Canada officials to analyze and assess how their campaign is resonating with Canadians. It can help them evaluate who are the most involved individuals or organizations. Perhaps on the day of Canada's anniversary, a similar analysis could be run to evaluate the tweets throughout the year to recognize and award individuals and organizations who have been most actively involved. This would not only provide encouragement, but also send a strong message to Canadians that government officials care and appreciate them, their interest and their involvement. Furthermore, the Canadian government can use this analysis to evaluate how each of their initiatives are doing. If a certain initiative is not attracting public attention, then the government can take proactive measures to make improvements. On the other hand, if certain parts of the campaign turn out to be very popular among Canadians, then, that is also important information which can be used to plan future events. This analysis also provides a complete overview of Canada's presence on social media (Twitter) and can be used as a reference in the later years.

### Lessons Learned

The most important lesson learned from this assignment was that the timing is crucial when analyzing social media data. For instance, depending on the period for which tweets are retrieved, the explanation of the analysis can be very different as the network graphs and tweet content changes over time. In the beginning of this analysis, tweets were extracted at the beginning of March. At that time, the graphs and charts revealed a different story based on the events that happened in that time frame. When the analysis was repeated with a different set of tweets at the end of March, the tweets painted a much different picture. The event analyzed in this project is still going on and therefore this is an interim analysis. For

a complete analysis of #Canada150, it would be more relevant to keep collecting tweets at least until the end of July 30<sup>th</sup> (1 month after the end of Canada's anniversary) to cover all events and big stories that might happen during this time frame.

## 7. REFERENCES

---

- [1] H. C. Gruzd A, "Enabling Community Through Social Media," *J Med Internet Res* 2013;15(10):e248, 2013.
- [2] E. & G. D. Dubois, "The multiple facets of influence: identifying political influentials and opinion leaders on Twitter.," *American Behavioral Scientist*, 58(10), 1260-1277., 2014.
- [3] W. W. S. Y. B. S. & P. H. W. Xu, "Predicting opinion leaders in Twitter activism networks: The case of the Wisconsin recall election.," *American Behavioral Scientist*, 58(10), 1278-1293., 2014.
- [4] A. O. Marthinussen, "Analyzing the US election using Twitter and Meta-Data in R," 29 July 2016. [Online]. Available: <https://www.r-bloggers.com/analyzing-the-us-election-using-twitter-and-meta-data-in-r/>.
- [5] J. Silge, "Ten Thousand Tweets," 8 December 2015. [Online]. Available: <http://juliasilge.com/blog/Ten-Thousand-Tweets/>.
- [6] Plotly, "Bar Charts in R," 30 Mar 2017. [Online]. Available: <https://plot.ly/r/bar-charts/>.
- [7] Martin, "Download, manage and visualize twitter data with R," 21 3 2015. [Online]. Available: <http://www.gis-blog.com/download-manage-and-visualize-twitter-data-with-r/>.
- [8] A. Gruzd, "Netlytic," 2017. [Online]. Available: <https://netlytic.org/home/>.

## 8. APPENDIX A - #Canada150

FIGURE 1

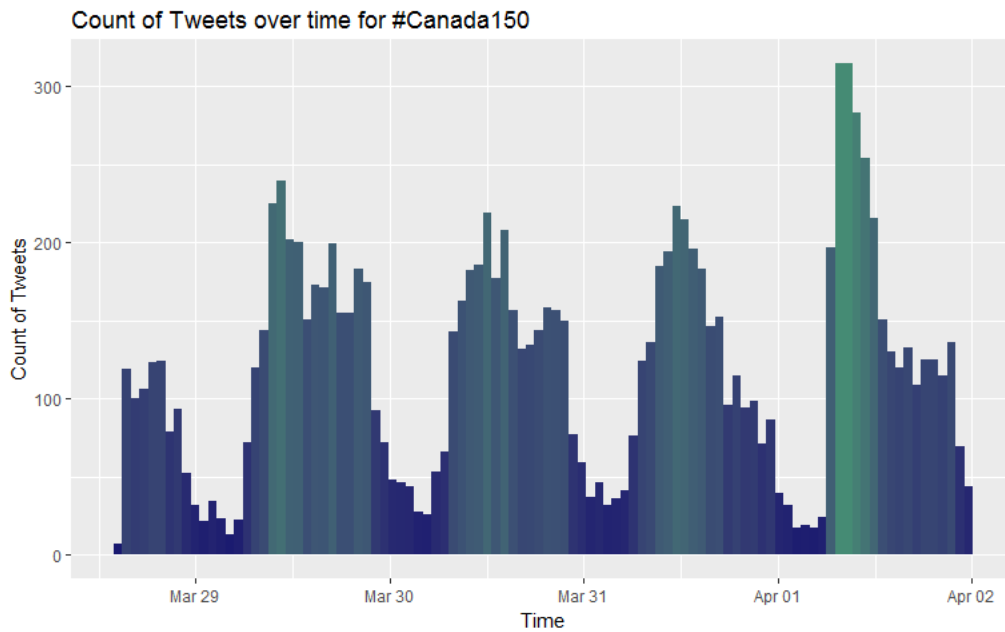


FIGURE 2

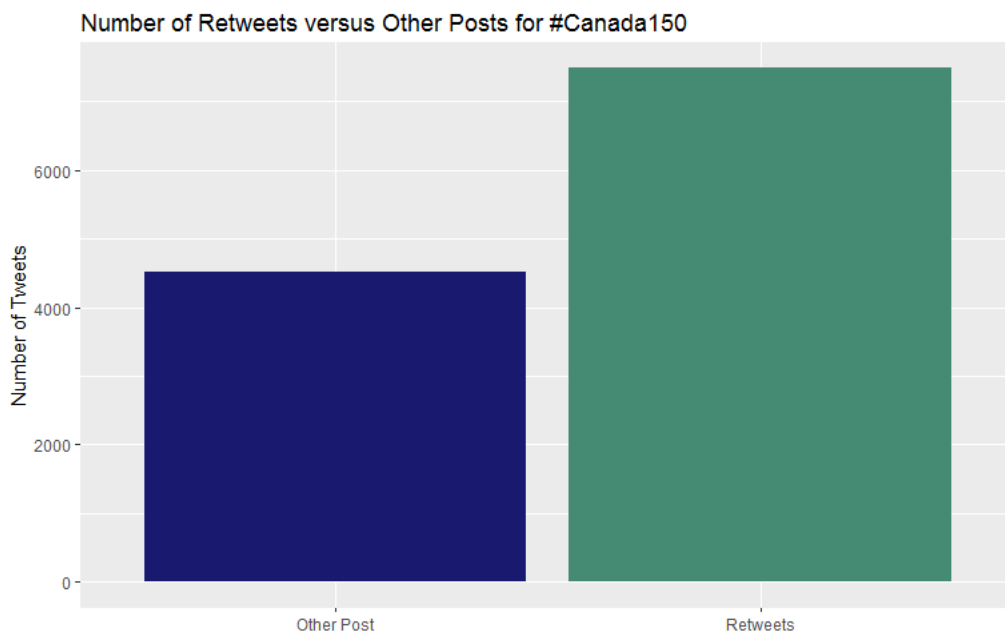




FIGURE 3

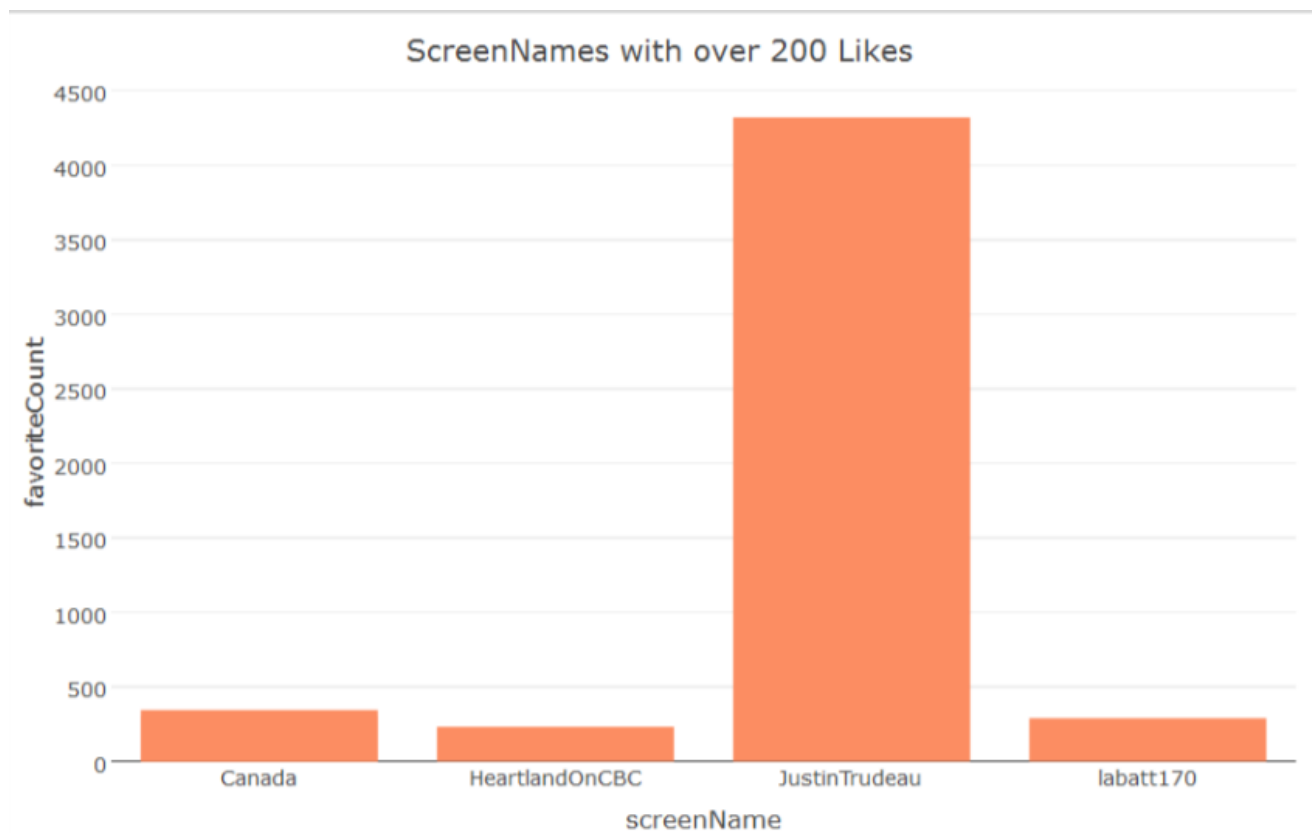


FIGURE 4

0% Zoom (with and without Edges)

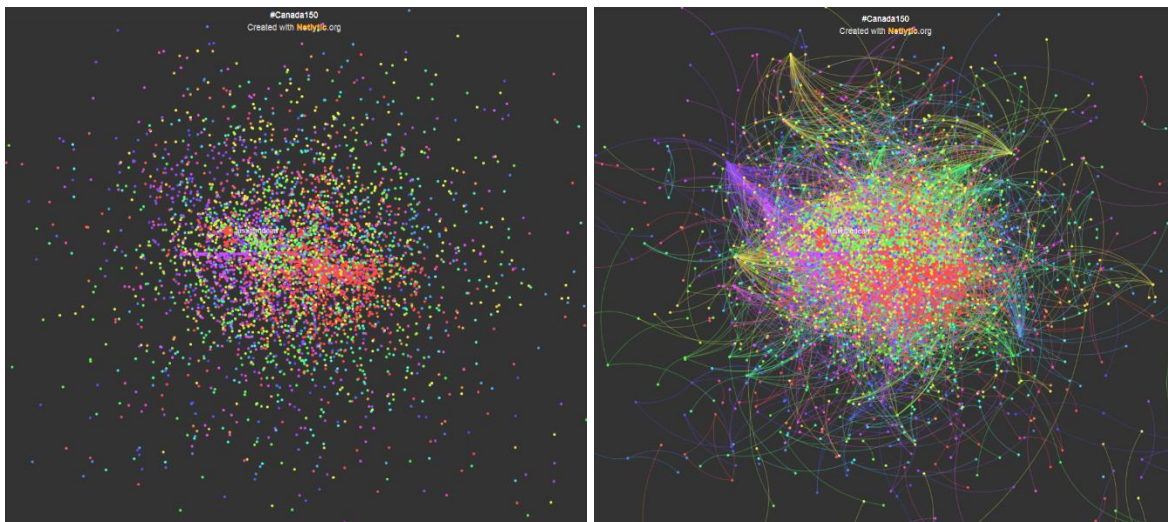


FIGURE 5

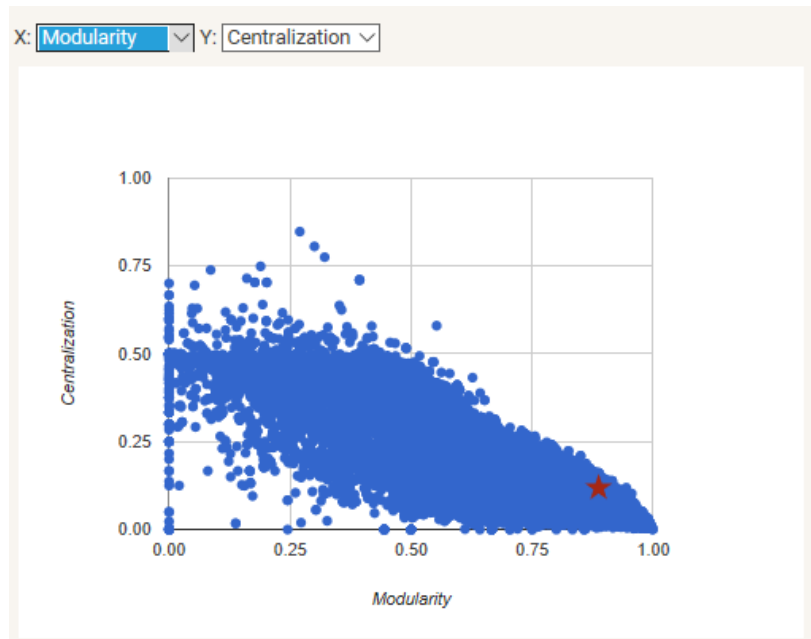


FIGURE 6

9% Zoom with Edges (In-degree)

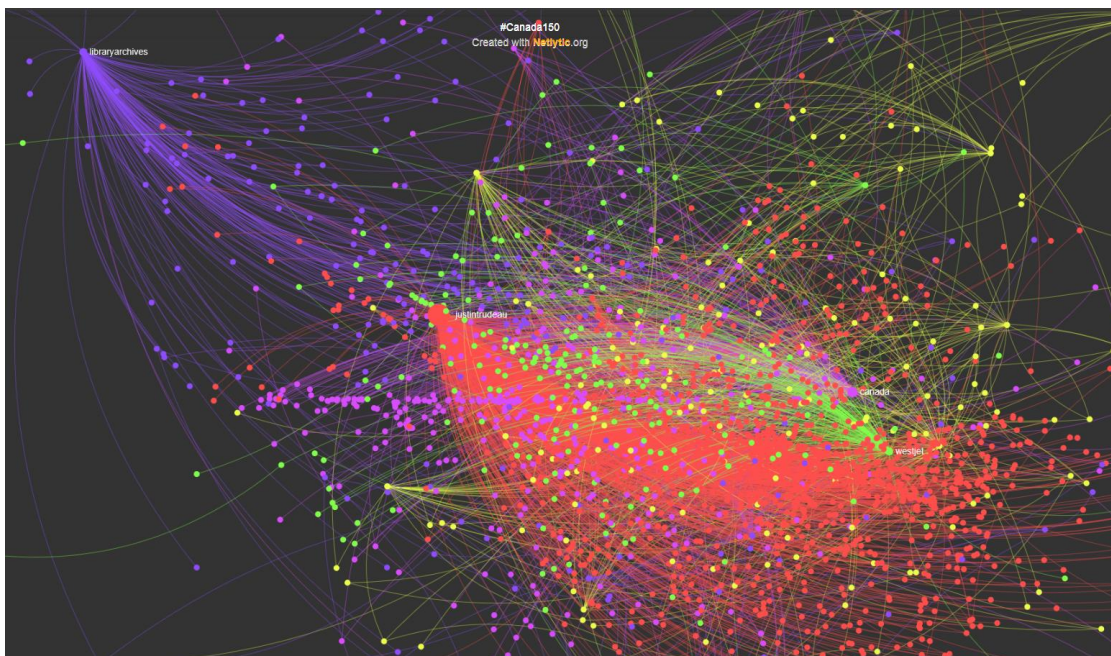


FIGURE 7

6% Zoom with Edges (Outdegree)

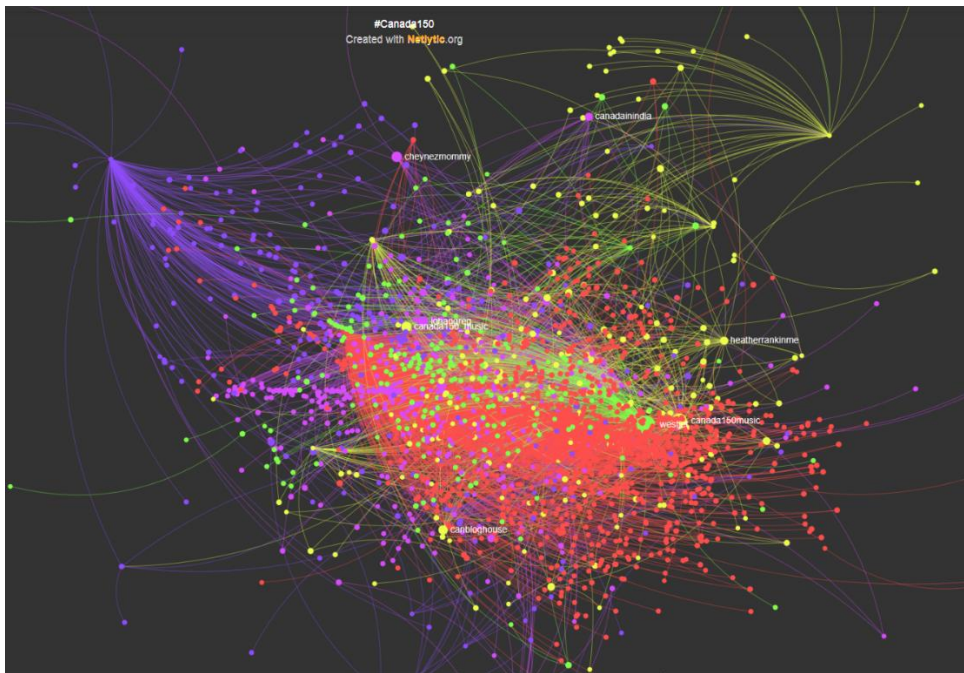
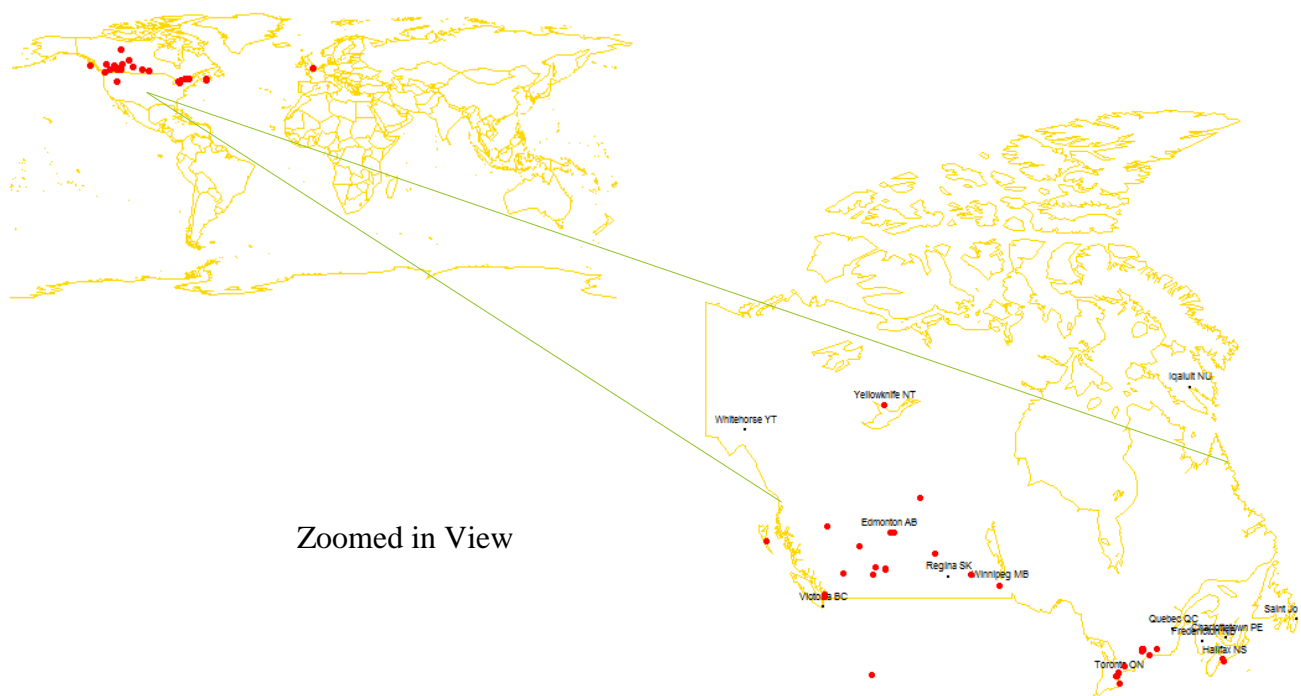
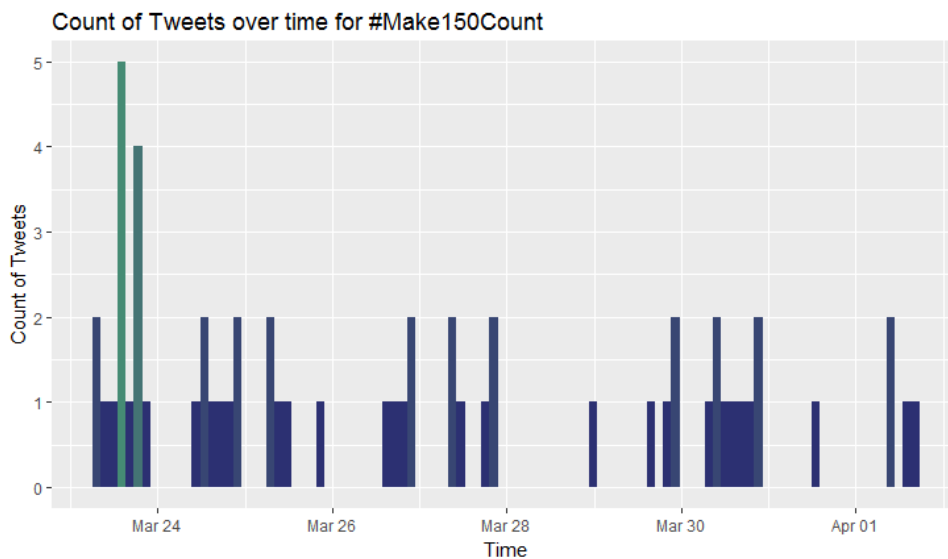


FIGURE 8

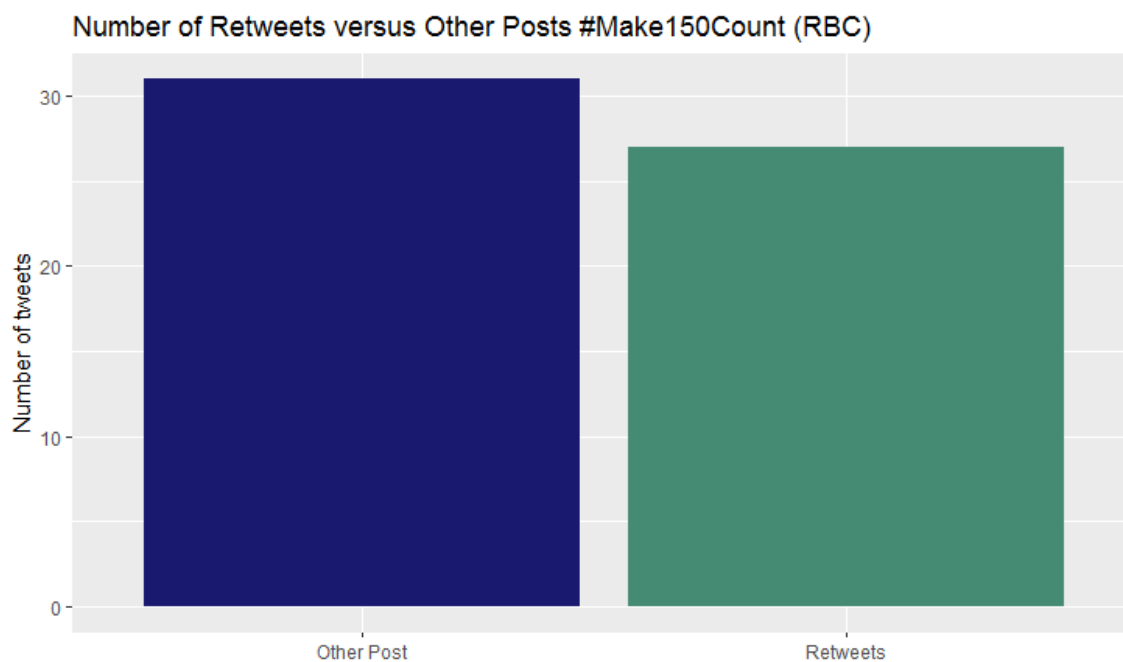


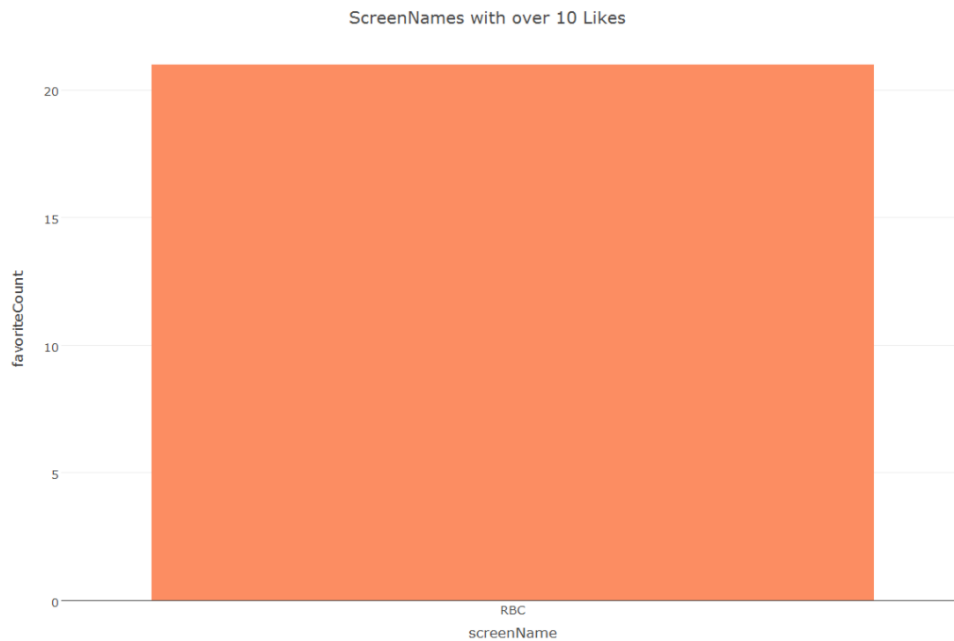
## 9. APPENDIX B - #Make150Count (RBC)



The graph here show that RBC does not have a lot of involvement and activity on Twitter. The number of tweets is very sparse, the count of retweets is low, and, no other usernames appeared

in the 3<sup>rd</sup> chart for most liked tweets except RBC's own handle. The last figure shows a very sparse network with a diameter value of only 3. This confirms that it has a very small network.

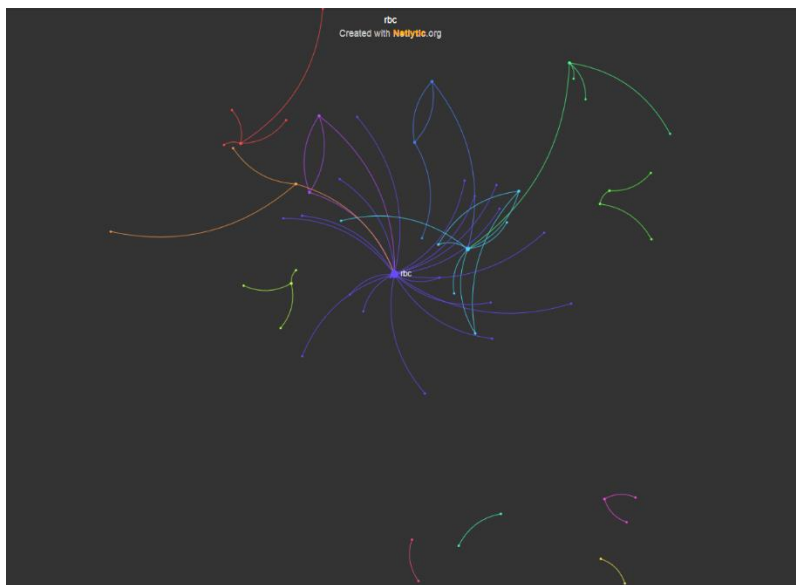
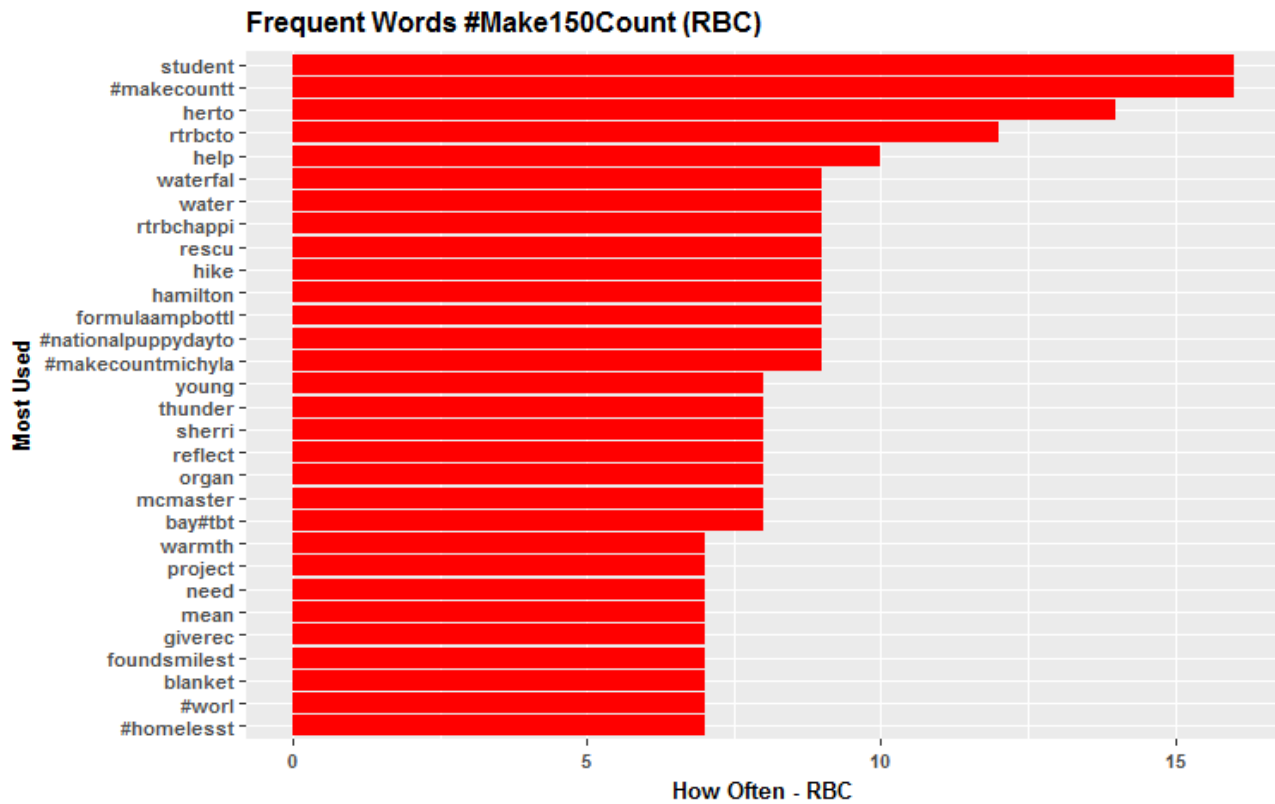




## RBC Word Cloud



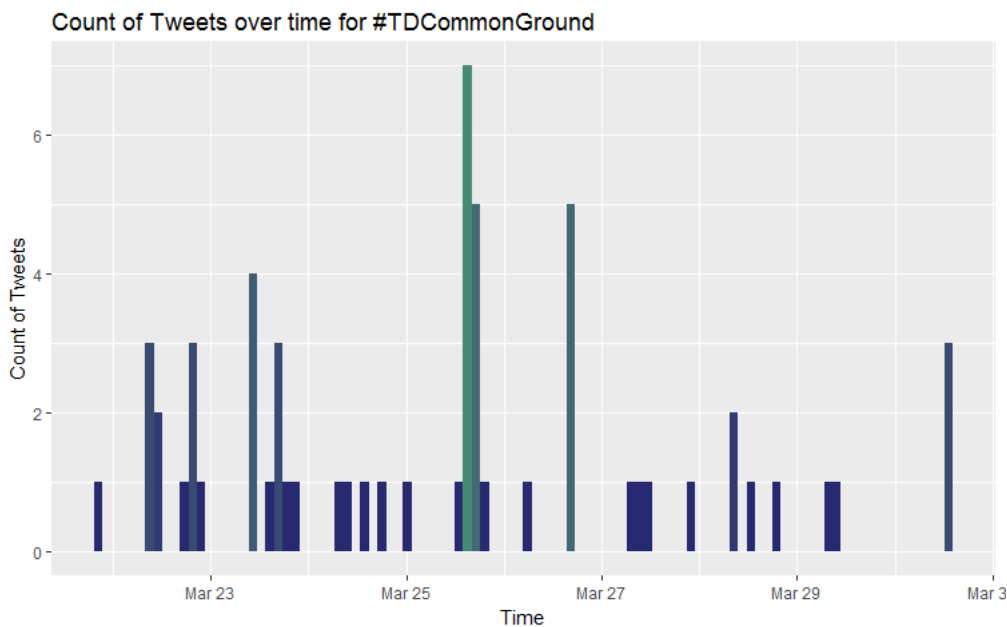




Network Graph for RBC

Diameter:	3
Density:	0.015819209039548
Reciprocity:	0.071428571428571
Centralization:	0.167738164815897
Modularity:	0.725833333333333

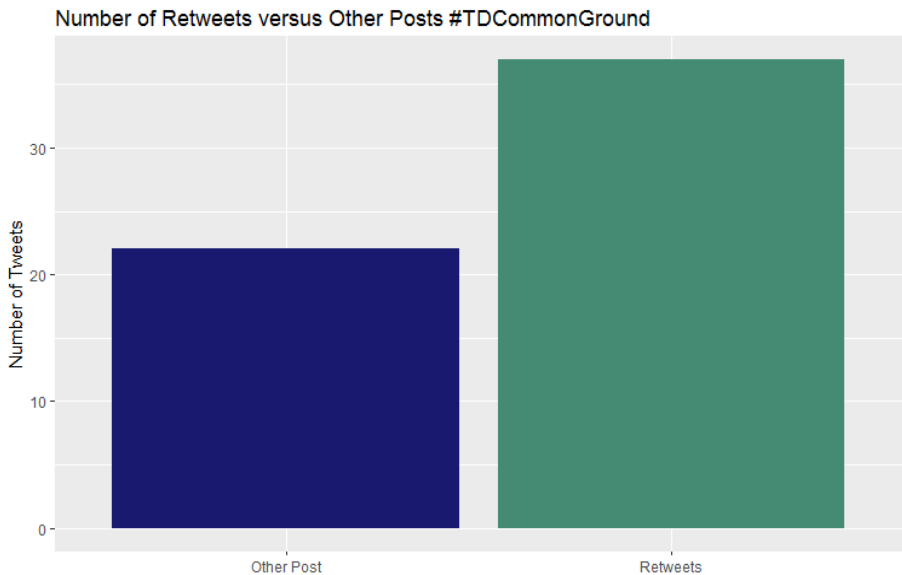
# 10. APPENDIX C - #TDCCommonGround (TD BANK)

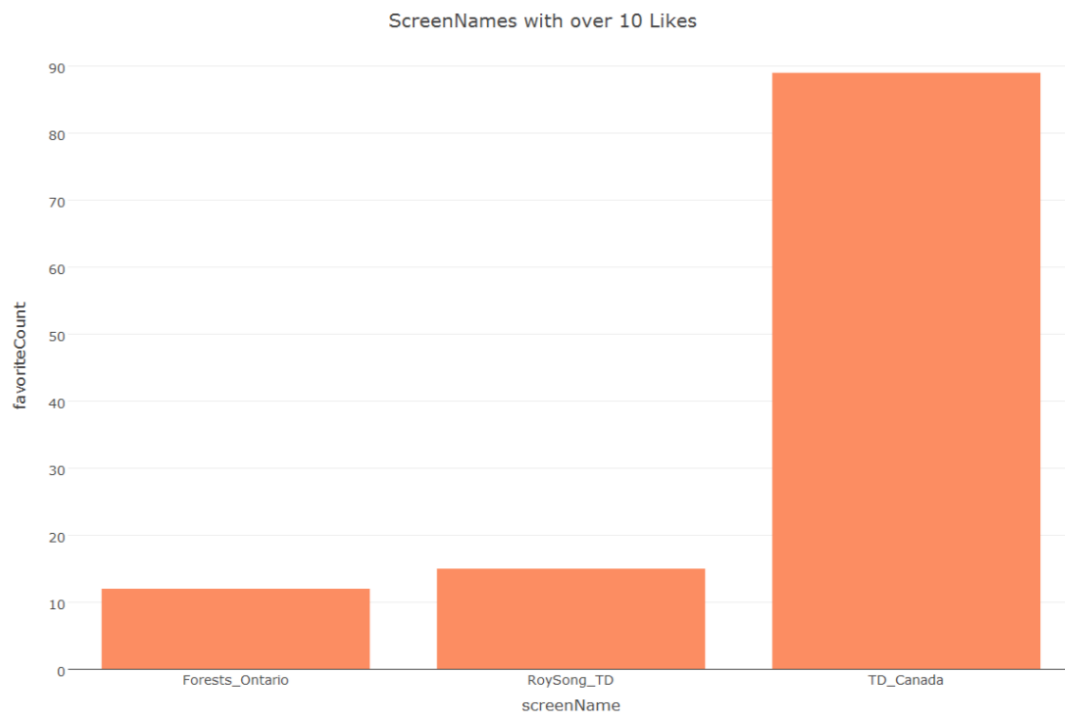


Although TD also has a very small number of tweets and a very sparse network with a diameter of only 3 (like RBC), it still managed to attract users to retweet its posts leading to a higher

number of retweets versus other post.

In the third chart that contains screennames whose tweets received the most likes, two other accounts appeared in addition to TD’s own account; one is of a loyal TD employee and the other is the official handle of Forests Ontario. It makes sense that Forests Ontario’s account appears in TD’s network because TD’s initiative for Canada’s 150<sup>th</sup> anniversary is to plant 150 trees across Canada which it wouldn’t not be able to do without Forest Ontario’s help.

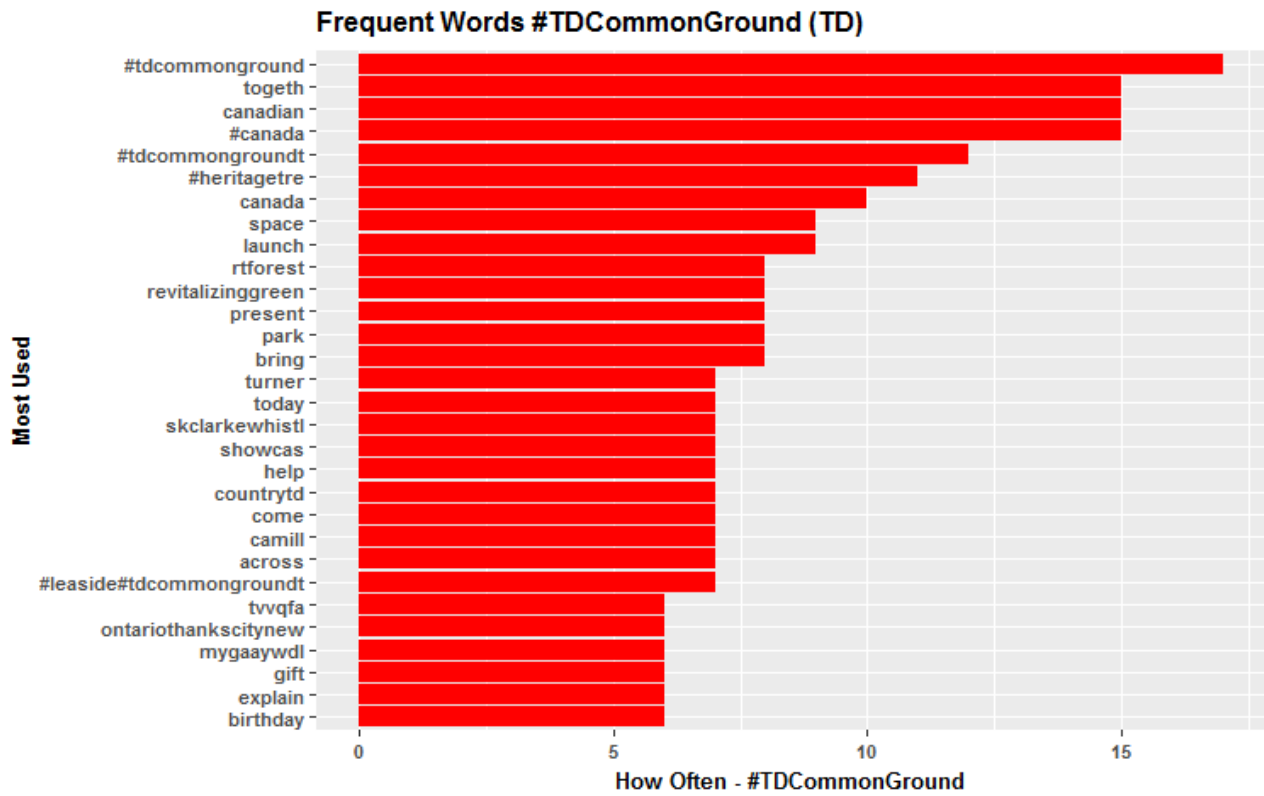




## TD Word Cloud

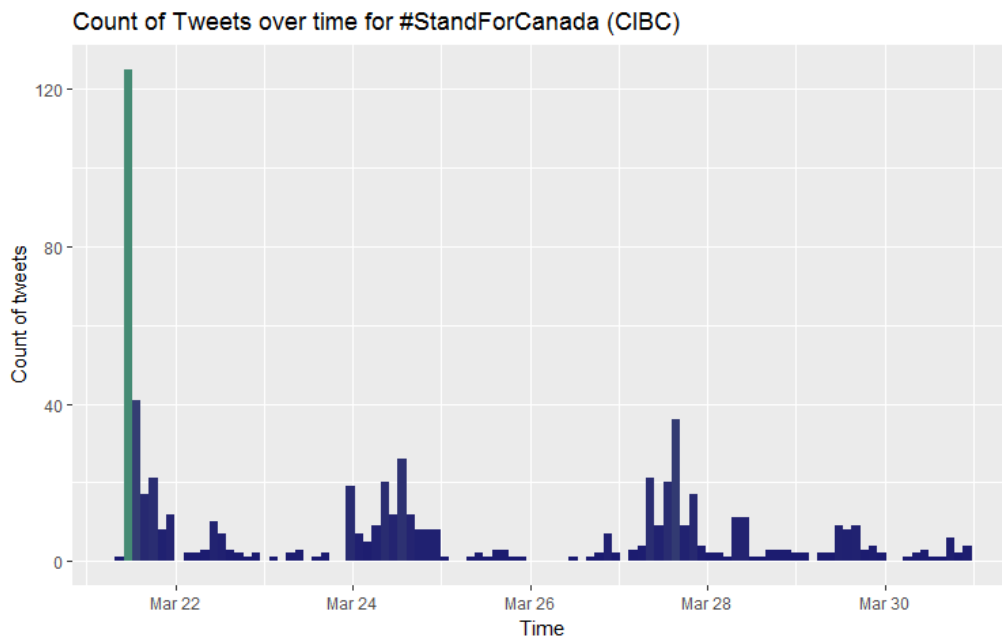






Diameter:	3
Density:	0.027469316189363
Reciprocity:	0.085106382978723
Centralization:	0.096491228070175
Modularity:	0.759659746251442

## 11. APPENDIX D - #StandForCanada (CIBC)



The graphs for CIBC's #StandForCanada campaign demonstrate that CIBC, of all the banks, is most successful in attracting interest of Canadians in terms of reciprocity with their tweets. The first

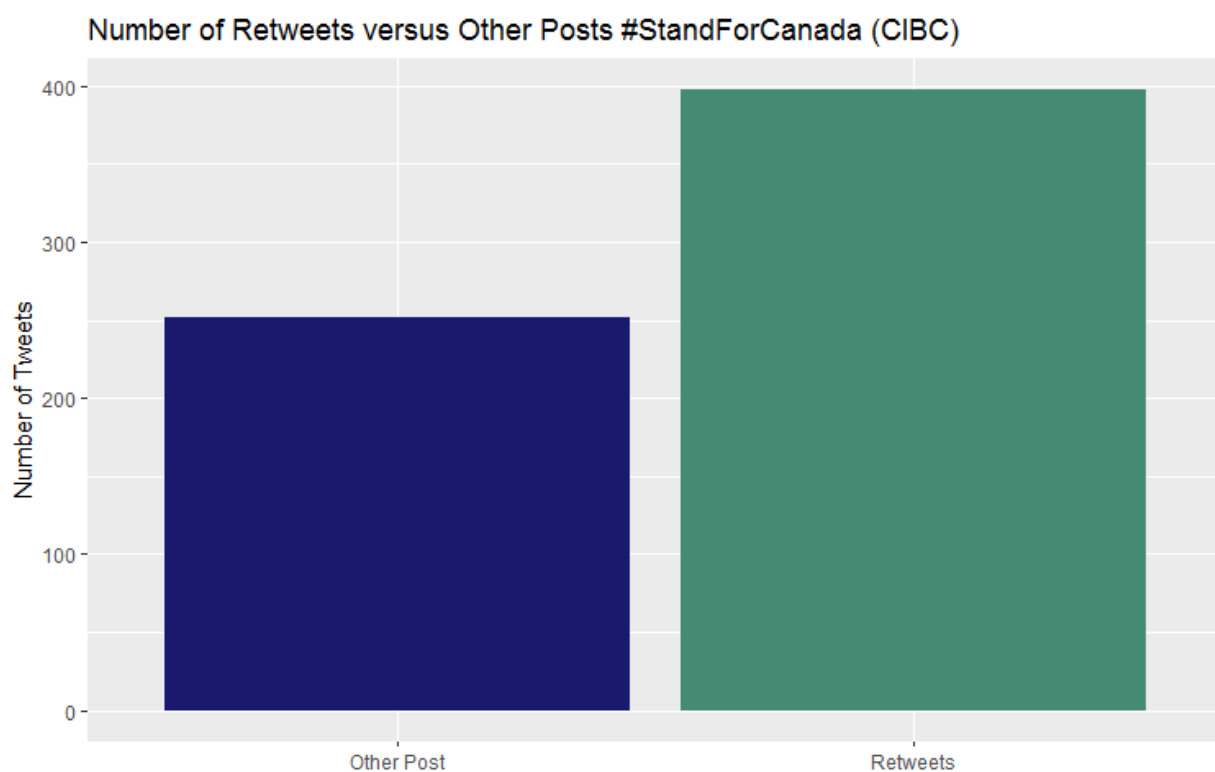
graph here shows a spike in the count of tweets between Mar 21<sup>st</sup> and 22<sup>nd</sup>. This was in response to a video that CIBC posted on March 21 as shown below.

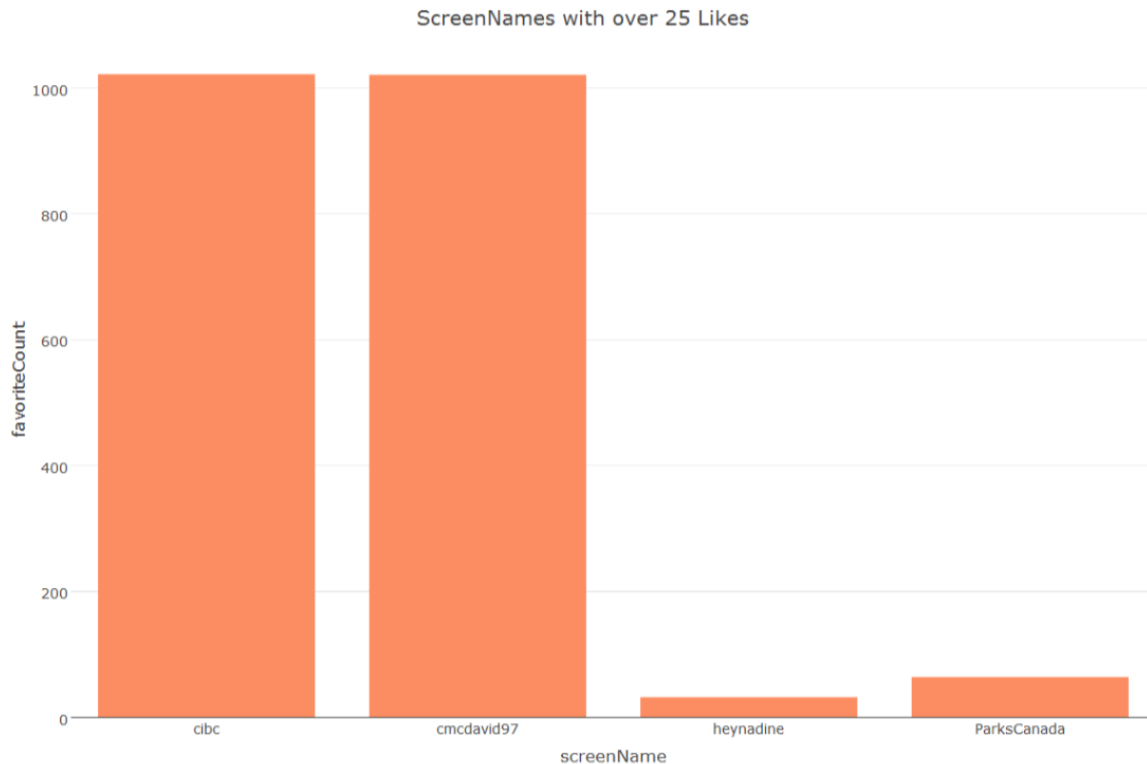


CIBC collaborated with Connor McDavid who is a Canadian ice hockey centre and captain for the Edmonton Oilers of the National Hockey League. The video gained instant popularity and hence there was a huge spike in the number of likes received

by @cibc and @connordavid97 user handles. The third graph shows these screennames which attracted the highest number of likes for their tweets.

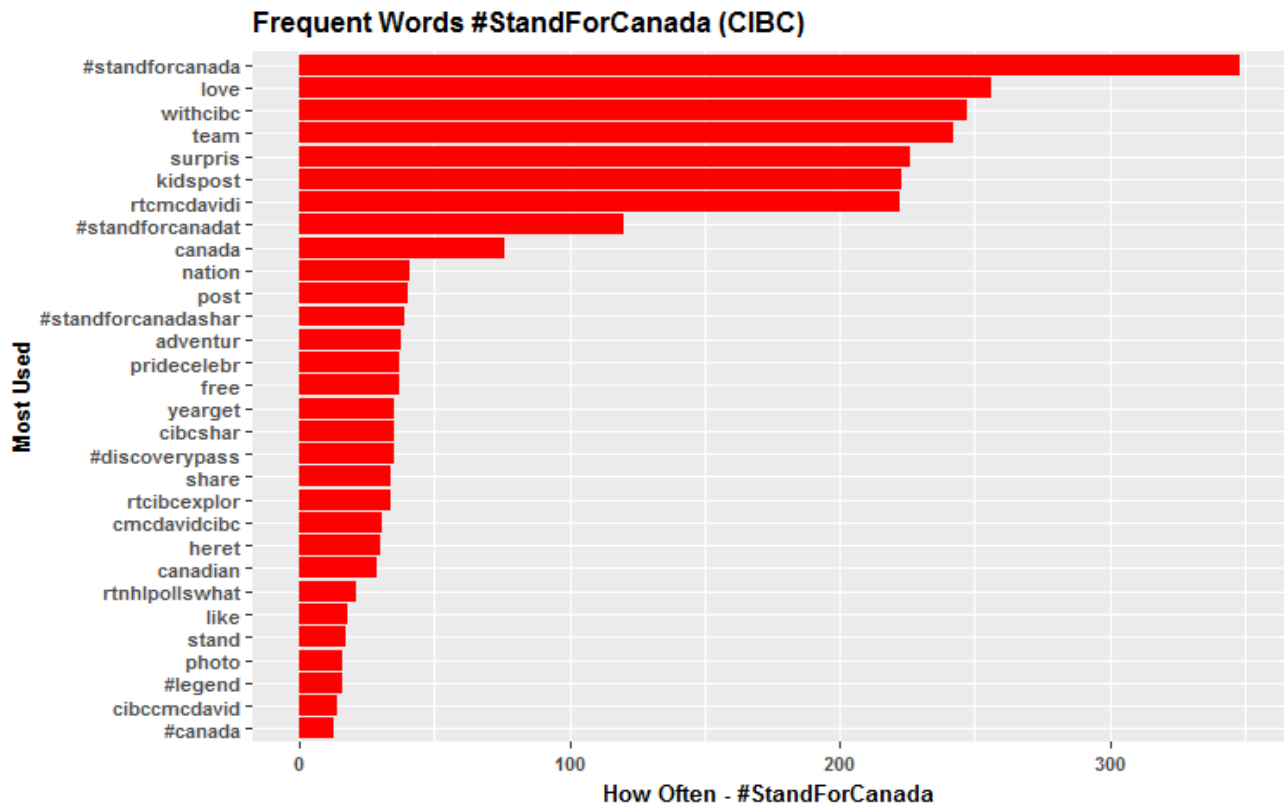
Of the three banks being analyzed here, CIBC has the densest network with a diameter of 5 and the least value for graph modularity indicating that all users in the network are mostly talking about the same topic and reciprocating with each other.





CIBC Word Cloud





Diameter: 5

Density: 0.004642580798762

Reciprocity: 0.029914529914530

Centralization: 0.2810366170187

Modularity: 0.495700291105956

## 12. APPENDIX E – Additional Stop Words

---

DATASET	STOP WORDS
CANADA150	("canada", "#canada", "#canadat", "canada", "mari", "canadian", "canadaa", "wait", "took", "check")
RBC	("will", "make", "want", "fake", "back", "can", "vgddftdg", "rbc", "#rbc", "makecount", "make", "lead", "local", "kymtttqx", "give")
TD	"will", "make", "want", "fake", "back", "can", "jfdcxldst", "jfdcxlds", "td", "#td", "#tdcommonground")
CIBC	"will", "make", "want", "fake", "back", "can", "gffdsrgf", "aswaasd", "cibc", "#cibc", "#StandForCanada")