Twitter and the Moscow City Duma Election: Data Mining and Cluster Analysis

EPPS 6323: Knowledge Mining

November 19, 2019

Party Affiliation and Social Media

2019 Moscow City Duma Election

- Duma – legislative body of Moscow
- Multiparty system
- 45 constituencies
- Moscow City Election Commission (MCEC)
- Election day – 9/8
- Protests during registration procedures for candidates

Russian Twitter

- Social media platform
- Usage differs based on society and culture
- Interaction between political figures and general public
- Not all political figures have Twitter accounts

Research Statement

- To establish if Twitter users can be grouped based on similarity of party affiliation or the type of twitter interaction (such as favorites, replies, retweets).
- To identify if Twitter users with specific party affiliation have similar twitter interaction (such as favorites, replies, retweets).

Hypothesis:

- If Twitter users have a certain party affiliation or a level of participation in Twitter interactions, then they are more likely to communicate in similar groups on the social media platform.

Data

Collection

- GetOldTweets3 query search
- Time range – 6/5/19 to 9/9/19
- Word: elections
- Limitations:
    - Word choice
        - Data availability
    - No geographic range
    - Cyrillic script

Manipulation

- Import into RStudio
- Subset and filter data to make dummy variables
- Pivot table to process data by reorganizing it
- Remove outliers
- Reshape data frame
- 10 parties, 2 independent associations
- 3 forms of Twitter activity

Machine Learning: Supervised vs. Unsupervised

K-Nearest Neighbors

- Identify training and test set to determine similarity of cases
- Similarity measure is used to classify new cases

Limitation:

- jitter package no longer available
- Unable to use supervised machine learning approach

K-Means Cluster Analysis

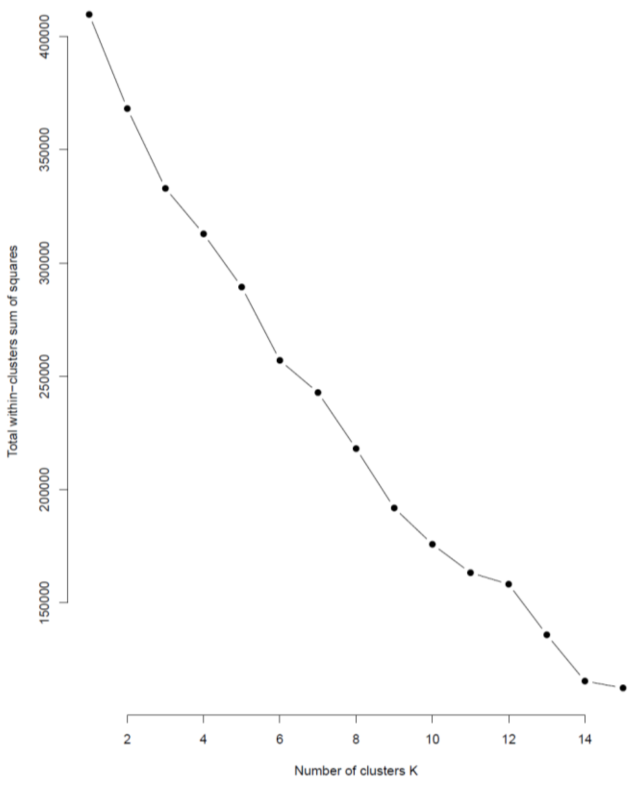
- Identify similarity among users in terms of party affiliation or Twitter interactions
- Test several clusters → need to determine optimal clusters for dataset

Limitation:

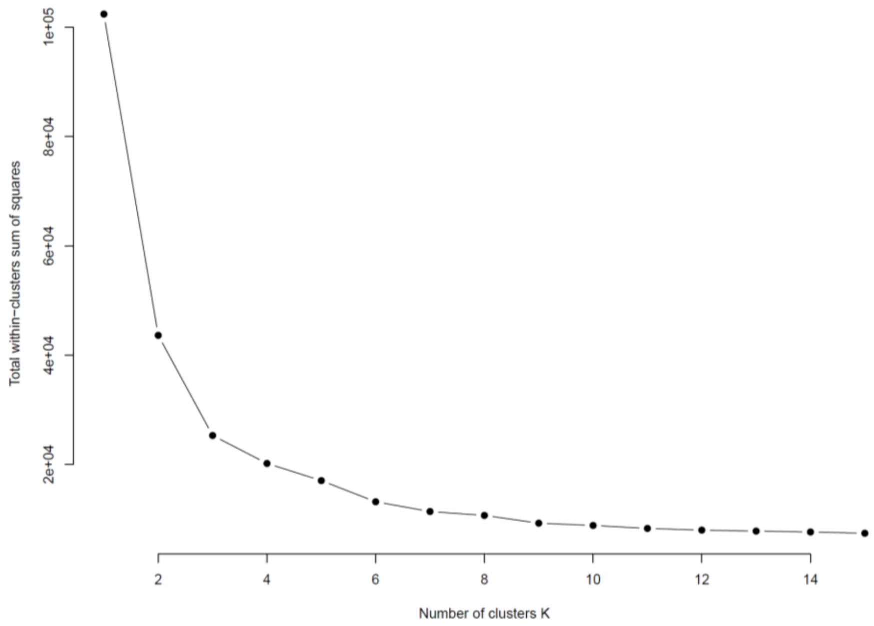
- Vector memory exhausted
- Determining optimal clusters
    - Unable to conduct Gap Statistic Method

Determine Optimal Clusters: Elbow Method

Party Affiliation

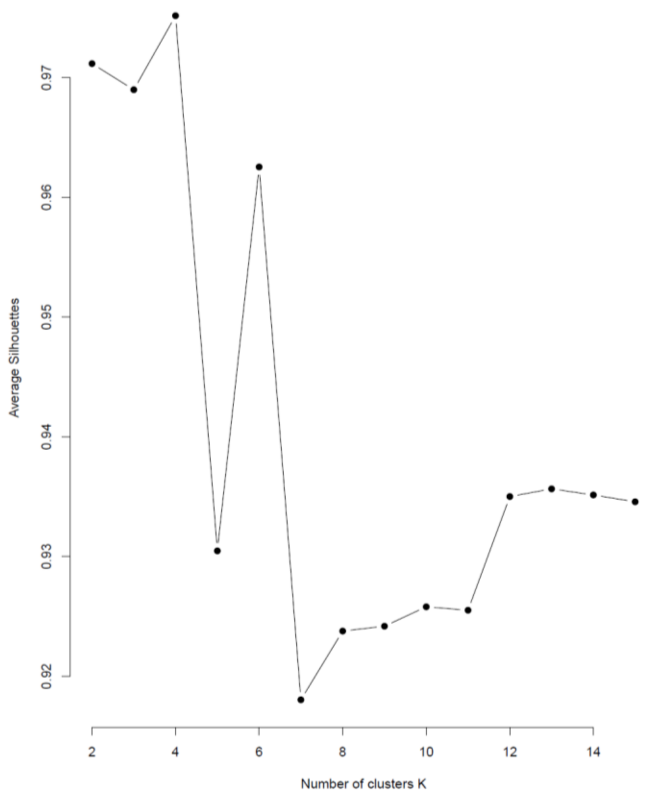


Twitter Interaction



Determine Optimal Clusters: Silhouette Method

Party Affiliation

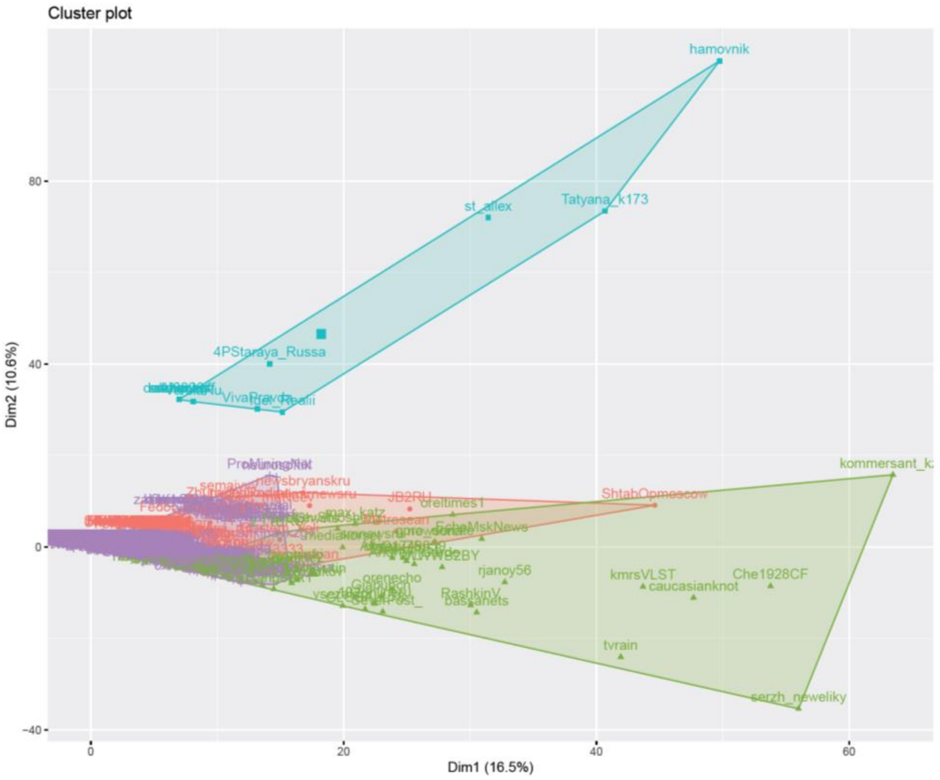


Twitter Interaction

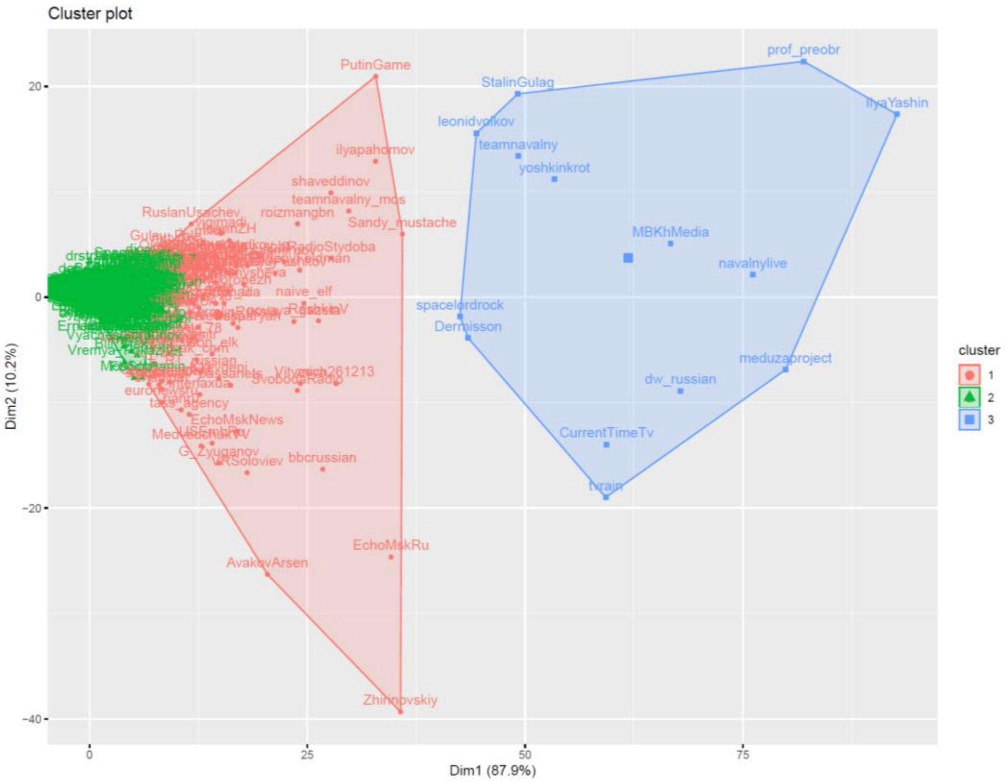
- Unable to generate plot due to warnings related to iteration

Final Cluster Analysis

Party Affiliation: K-means clustering with 4 clusters of sizes 58, 79, 11, 33996



Twitter Interaction: K-means clustering with 3 clusters of sizes 104, 34026, 14



Cluster Analysis

- Party Affiliation
  - Divided the dataset into similar subsets of users based on their Twitter activity in relation to political parties
    - To, mentions, hashtags, Twitter text
  - Political parties do not create directly evident groupings
    - May be based on political ideologies of parties
- Twitter Interactions
  - Divided the dataset into similar subsets of used based on their Twitter activity in relation to Twitter interactions
    - Replies, retweets, favorites
  - Clusters may not necessarily depend on type of Twitter interaction but the amount of interaction

Predictive Analytics

- Incorporate less well-known political parties to determine if they match a specific cluster
  - Unable to currently test this due to issue in data → can no longer use regular expression to filter key words identifying which users participate in discussions concerning specific political parties

Linear Regression

DV: Replies

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.4876	0.1591	9.35	0.0000
A_Just_Ru	37.4548	3.2256	11.61	0.0000
Civilian_Power	14.0935	17.2862	0.82	0.4149
Communists_Ru	-27.7988	4.6793	-5.94	0.0000
CPRF	5.5669	0.2998	18.57	0.0000
Greens	-4.3092	9.7637	-0.44	0.6590
Gudkov	48.6038	1.2004	40.49	0.0000
LDPR	10.1818	1.2250	8.31	0.0000
Navalny	6.5013	0.4152	15.66	0.0000
Party_Growth	-18.1144	5.4120	-3.35	0.0008
Rodina	-3.7764	2.6534	-1.42	0.1547
United_Ru	3.6568	1.0769	3.40	0.0007
Yabloko	1.7393	1.2735	1.37	0.1720

DV: Retweets

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.4924	0.6499	6.91	0.0000
A_Just_Ru	69.8062	13.1804	5.30	0.0000
Civilian_Power	12.6179	70.6354	0.18	0.8582
Communists_Ru	-103.5382	19.1207	-5.41	0.0000
CPRF	15.5966	1.2249	12.73	0.0000
Greens	-4.2336	39.8967	-0.11	0.9155
Gudkov	147.8964	4.9050	30.15	0.0000
LDPR	8.4048	5.0058	1.68	0.0932
Navalny	23.7525	1.6968	14.00	0.0000
Party_Growth	-35.8408	22.1147	-1.62	0.1051
Rodina	4.5866	10.8426	0.42	0.6723
United_Ru	32.0585	4.4004	7.29	0.0000
Yabloko	-5.9885	5.2038	-1.15	0.2498

Linear Regression

DV: Favorites

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	16.1993	2.5785	6.28	0.0000
A_Just_Ru	255.2879	52.2919	4.88	0.0000
Civilian_Power	67.8433	280.2386	0.24	0.8087
Communists_Ru	-383.0996	75.8592	-5.05	0.0000
CPRF	47.7409	4.8598	9.82	0.0000
Greens	-44.1782	158.2858	-0.28	0.7802
Gudkov	503.4869	19.4601	25.87	0.0000
LDPR	6.9817	19.8600	0.35	0.7252
Navalny	98.3626	6.7317	14.61	0.0000
Party_Growth	-82.5939	87.7378	-0.94	0.3465
Rodina	-18.9286	43.0169	-0.44	0.6599
United_Ru	133.5473	17.4581	7.65	0.0000
Yabloko	-2.5350	20.6457	-0.12	0.9023

DV: Combined Twitter Interactions

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	22.1793	3.3138	6.69	0.0000
A_Just_Ru	362.5489	67.2041	5.39	0.0000
Civilian_Power	94.5547	360.1553	0.26	0.7929
Communists_Ru	-514.4366	97.4923	-5.28	0.0000
CPRF	68.9045	6.2456	11.03	0.0000
Greens	-52.7211	203.4247	-0.26	0.7955
Gudkov	699.9871	25.0096	27.99	0.0000
LDPR	25.5683	25.5236	1.00	0.3165
Navalny	128.6164	8.6515	14.87	0.0000
Party_Growth	-136.5491	112.7583	-1.21	0.2259
Rodina	-18.1184	55.2842	-0.33	0.7431
United_Ru	169.2626	22.4367	7.54	0.0000
Yabloko	-6.7842	26.5333	-0.26	0.7982

Conclusion

- Uncover possible patterns and structure of data
  - influences predictive analytics
- Disadvantages of cluster analysis:
  - Number of clusters need to be specified
  - Outliers may have impacted analysis
- Linear regression:
  - Briefly demonstrates that active Twitter participation by political parties results in more Twitter interactions
- Constrained by data
  - Unable to determine presence of political polarization or conduct sentiment analysis

Bibliography

Bhadane, Chetashri, Hardi Dalal, and Heenal Doshi. 2015. “Sentiment Analysis: Measuring opinions.” *Procedia Computer Science* 45: 808-814.

Bodrunova, Svetlana S. 2013. “Fragmentation And Polarization Of The Public Sphere In The 2000s: Evidence from Italy and Russia.” *Global Media Journal-German Edition* 3(1).

Cha, Meeyoung, Fabrício Benevenuto, Hamed Haddadi, and Krishna Gummadi. 2012. “The world of connections and information flow in twitter.” *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 42(4): 991-998.

Colton, Timothy J., and Michael McFaul. 2000. “Reinventing Russia's Party of Power: “Unity” and the 1999 Duma Election.” *Post-Soviet Affairs* 16(3): 201-224.

Conover, Michael D., Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. “Political polarization on twitter.” In *Fifth international AAAI conference on weblogs and social media*: 89-96.

Fiorina, Morris P., and Samuel J. Abrams. 2008. “Political polarization in the American public.” *Annu. Rev. Polit. Sci.* 11(June): 563-588. doi: 10.1146/annurev.polisci.11.053106.153836

Frye, Timothy. 2002. “The Perils Of Polarization: Economic performance in the postcommunist world." *World politics* 54(3): 308-337.

Jackson, Linda A., and Jin-Liang Wang. 2013. “Cultural Differences In Social Networking Site Use: A comparative study of China and the United States.” *Computers in Human Behavior* 29(3): 910-921. doi: 10.1016/j.chb.2012.11.024

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning with Applications in R*. New York City, NY: Springer.

Kolossov, V. A. 1997. “Political Polarization At The National And The Intra-Urban Levels: The role of Moscow in Russian politics and socio-political cleavages within the city.” *GeoJournal* 42(4): 385-401.

Kunicova, Jana, and Thomas Frederick Remington. 2008. “Mandates, Parties and Dissent: Effect of Electoral Rules on Parliamentary Party Cohesion in the Russian State Duma, 1994—2003.” *Party Politics* 14(5): 555-574.

McAllister, Ian, and Stephen White. 2008. “‘It's the Economy, Comrade!’ Parties and Voters in the 2007 Russian Duma Election.” *Europe-Asia Studies* 60(6): 931-957.

Tan, Chenhao, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. 2011. “User-level sentiment analysis incorporating social networks.” In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (August): 1397-1405. doi: 10.1145/2020408.2020614

Thames Jr, Frank C. 2001. “Legislative Voting Behaviour In The Russian Duma: Understanding the effect of mandate.” *Europe-Asia Studies* 53(6): 869-884.

Yardi, Sarita, and Danah Boyd. 2010. “Dynamic Debates: An analysis of group polarization over time on Twitter.” *Bulletin of science, technology & society* 30(5): 316-327. doi: 10.1177/0270467610380011