Twitter and the Moscow City Duma Election: Collecting Data

EPPS 6302: Methods of Data Collection and Production

November 21, 2019

Introduction

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 or political parties have Twitter accounts

Research Statement

- Original focus: political polarization and sentiment analysis
- Explore whether twitter interactions are impacted by discussions surrounding specific political parties

Research Statement:

Determine whether amount of tweets or discussions concerning different political parties impacts social media platform connections (such as favorites, replies, retweets).

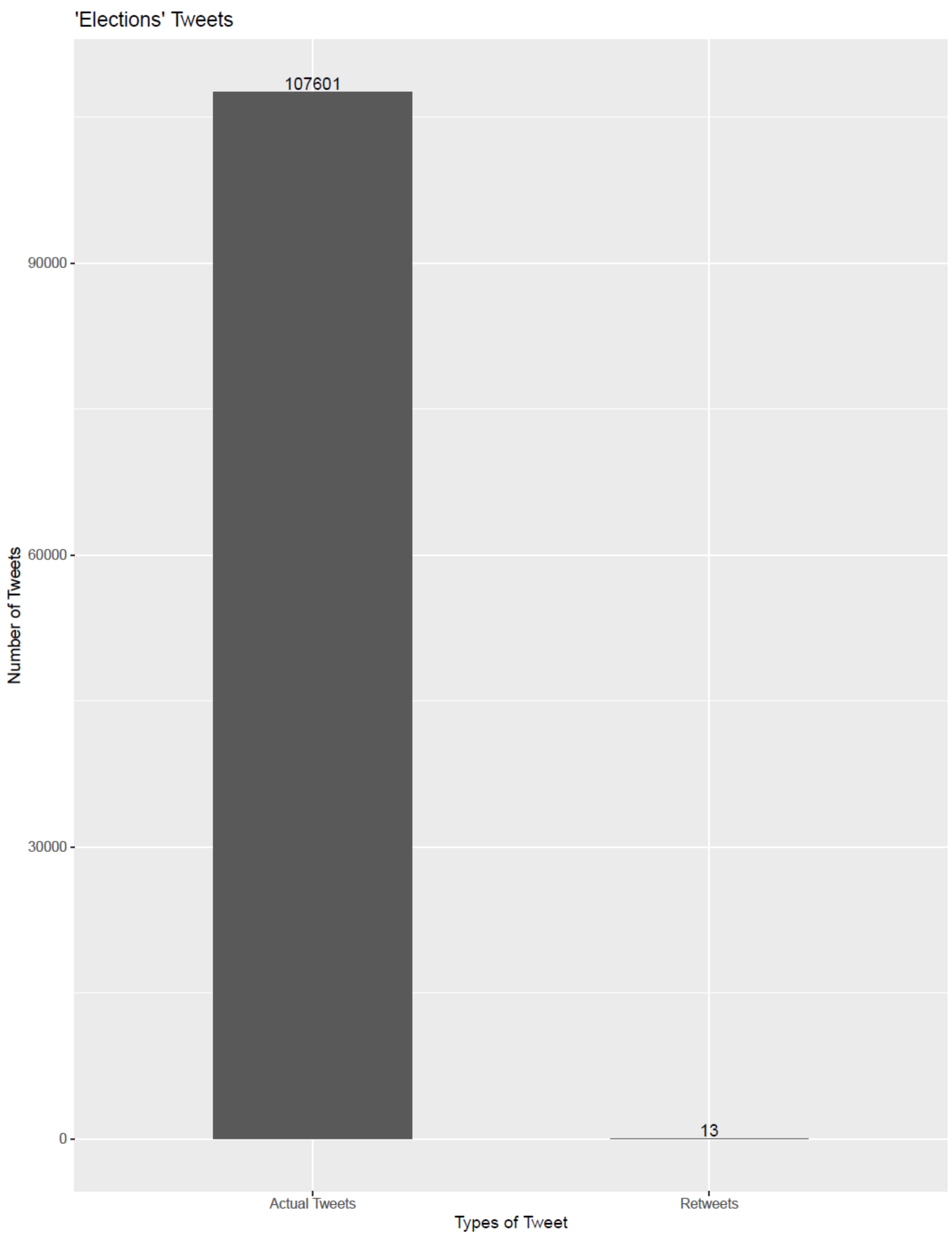
Data Collection Methods

- GetOldTweets3 query search
  - Find Tweets using the word: elections
  - Quantitative and qualitative data
- Time range – 6/5/19 to 9/9/19
  - Twitter API not an option due to limit on time range
- Limitations:
  - Query search of specific election related terms
  - Geographic range

Next step: understand data

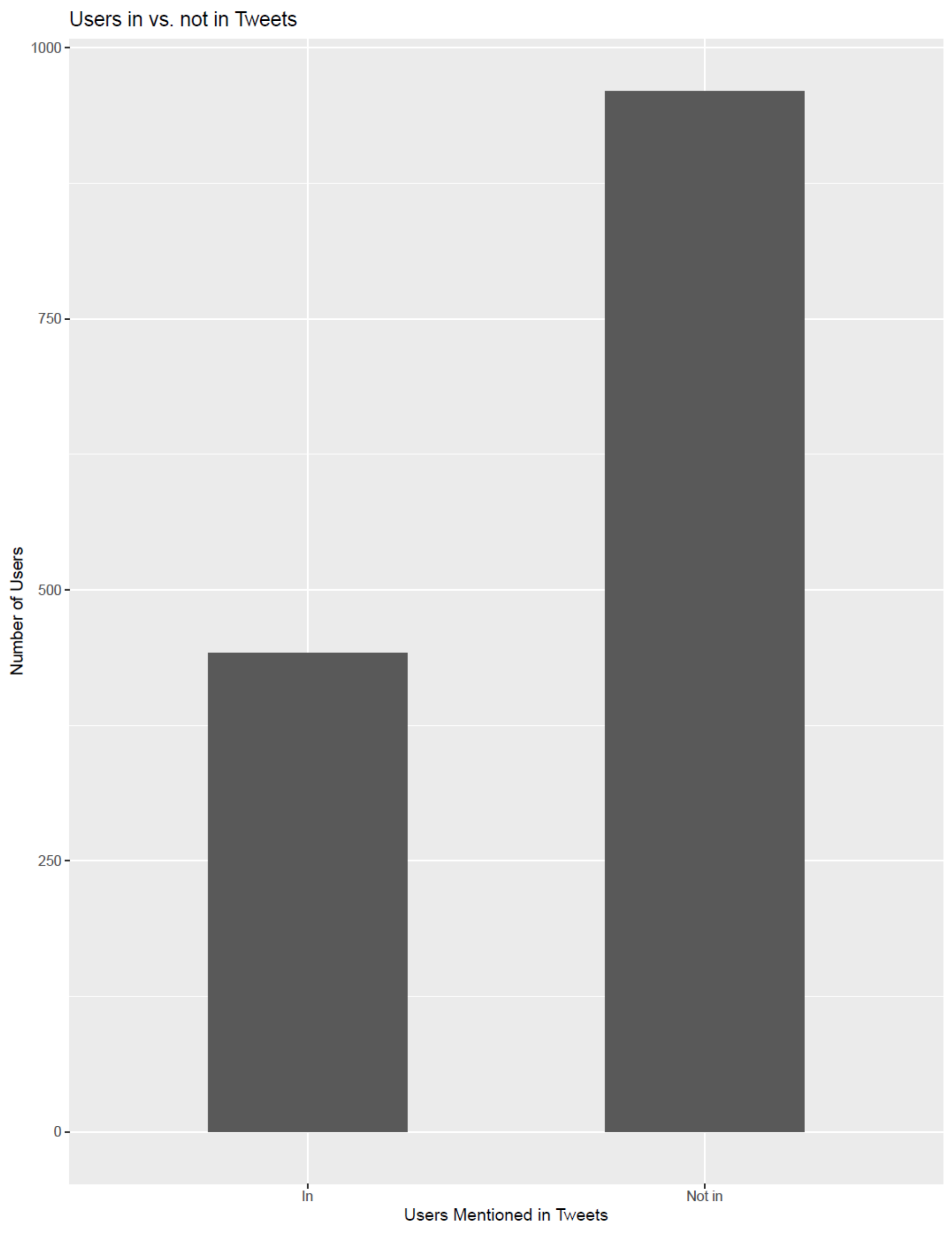
Tweets

Unique Tweets	Total Tweets
100,725.00	107,614.00



Users

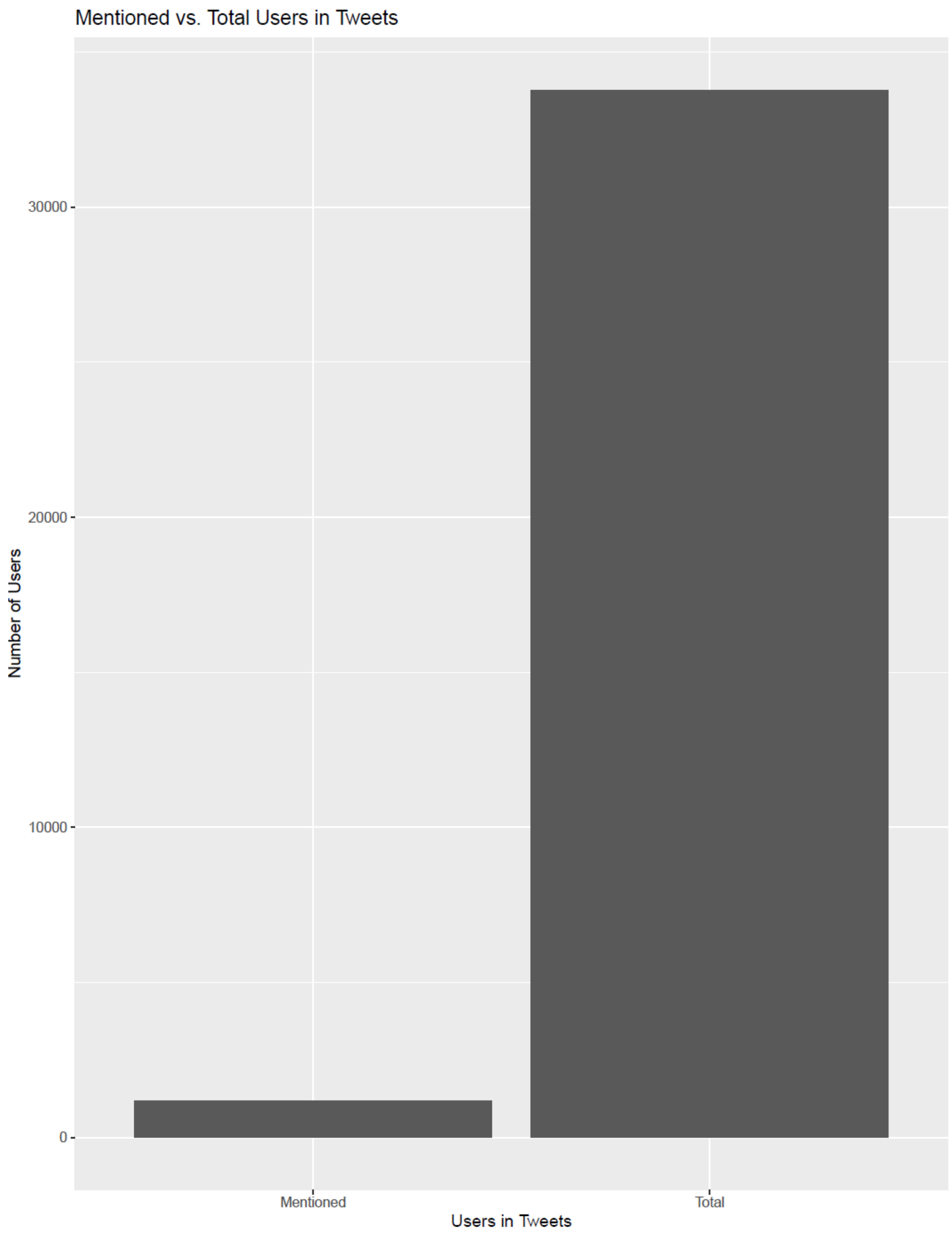
Unique Username	Total Usernames
34,147.00	107,614.00



Mentioned Users

Top 5 Mentioned Users

mentioned_users	volume
meduzaproject	304
navalny	181
SobolLubov	167
MosSobyanin	116
CIKRussia	84



Data Manipulation and Production

- Import CSV file into R
  - Cyrillic script
    - Text mining and sentiment analysis unachievable
- Identify users discussing political parties
- Create rows and columns for party affiliation
- Pivot table to process data by reorganizing it
- 10 parties
  - United Russia, Communist Party (CPRF), Yabloko, A Just Russia, LDPR, Rodina, Communists of Russia, The Greens, Civilian power, Party of Growth
- 3 forms of Twitter activity (replies, retweets, favorites)

Data Preview

username	Count.of.text	Sum.of.Twit	Sum.of.party
105	1	0	0
115446	3	2	0
185790	1	6	0
430014	4	17	0
580947	1	0	0
600427	6	3	0

Side note: Two datasets used for multiple analyses

Model Dispersion

- Residual Deviance is greater than the degrees of freedom in the Poisson regression models, which means that over-dispersion exists
  - Variance is greater than mean
- Estimates made by the models are correct
  - standard errors and standard deviation are wrong and unaccounted for by the model
- Consider Quasi-Poisson model

Analysis: Sum.of.Twit ~ Sum.of.party + Count.of.text

Poisson Regression

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.6893	0.0008	4428.76	0.0000
Sum.of.party	0.7908	0.0005	1601.83	0.0000
Count.of.text	0.0108	0.0000	1435.42	0.0000

Quasi-Poisson Regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.6893	0.0682	54.06	0.0000
Sum.of.party	0.7908	0.0404	19.55	0.0000
Count.of.text	0.0108	0.0006	17.52	0.0000

Analysis: Sum of Twitter Interactions

Poisson Regression

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.4300	0.0009	3615.66	0.0000
Count.of.text	0.0119	0.0000	1667.27	0.0000
party_A_Just	-1.1152	0.0069	-161.61	0.0000
party_Comm	-3.0009	0.0131	-228.80	0.0000
party_CPRF	2.1728	0.0019	1116.88	0.0000
party_Civ_Power	-1.0184	0.0728	-13.99	0.0000
party_Greens	-3.9482	0.0589	-67.07	0.0000
party_LDPR	-1.1651	0.0041	-284.33	0.0000
party_Growth	0.2440	0.0141	17.34	0.0000
party_Rodina	-0.0177	0.0070	-2.51	0.0121
party_United	2.4484	0.0021	1194.07	0.0000
party_Yabloko	0.8955	0.0039	230.56	0.0000

Quasi-Poisson Regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.4300	0.0721	47.56	0.0000
Count.of.text	0.0119	0.0005	21.93	0.0000
party_A_Just	-1.1152	0.5246	-2.13	0.0335
party_Comm	-3.0009	0.9971	-3.01	0.0026
party_CPRF	2.1728	0.1479	14.69	0.0000
party_Civ_Power	-1.0184	5.5332	-0.18	0.8540
party_Greens	-3.9482	4.4753	-0.88	0.3777
party_LDPR	-1.1651	0.3115	-3.74	0.0002
party_Growth	0.2440	1.0702	0.23	0.8196
party_Rodina	-0.0177	0.5350	-0.03	0.9737
party_United	2.4484	0.1559	15.71	0.0000
party_Yabloko	0.8955	0.2953	3.03	0.0024

Analysis: Sum of replies

Poisson Regression

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.7010	0.0037	187.76	0.0000
Count.of.text	0.0120	0.0000	446.43	0.0000
party_A_Just	-0.5892	0.0221	-26.71	0.0000
party_Comm	-2.1682	0.0367	-59.15	0.0000
party_CPRF	2.3009	0.0077	298.99	0.0000
party_Civ_Power	-0.1445	0.1693	-0.85	0.3936
party_Greens	-3.7280	0.2589	-14.40	0.0000
party_LDPR	-0.5407	0.0144	-37.65	0.0000
party_Growth	-0.3841	0.0659	-5.82	0.0000
party_Rodina	-0.3077	0.0263	-11.69	0.0000
party_United	1.9918	0.0088	225.36	0.0000
party_Yabloko	0.6840	0.0152	44.91	0.0000

Quasi-Poisson Regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.7010	0.0542	12.93	0.0000
Count.of.text	0.0120	0.0004	30.74	0.0000
party_A_Just	-0.5892	0.3204	-1.84	0.0659
party_Comm	-2.1682	0.5324	-4.07	0.0000
party_CPRF	2.3009	0.1118	20.59	0.0000
party_Civ_Power	-0.1445	2.4595	-0.06	0.9532
party_Greens	-3.7280	3.7601	-0.99	0.3215
party_LDPR	-0.5407	0.2086	-2.59	0.0095
party_Growth	-0.3841	0.9577	-0.40	0.6884
party_Rodina	-0.3077	0.3822	-0.81	0.4207
party_United	1.9918	0.1284	15.52	0.0000
party_Yabloko	0.6840	0.2212	3.09	0.0020

Analysis: Sum of retweets

Poisson Regression

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.8196	0.0021	855.42	0.0000
Count.of.text	0.0120	0.0000	764.98	0.0000
party_A_Just	-0.9983	0.0150	-66.49	0.0000
party_Comm	-2.7815	0.0276	-100.68	0.0000
party_CPRF	2.2281	0.0044	509.77	0.0000
party_Civ_Power	-1.9213	0.2427	-7.92	0.0000
party_Greens	-3.1040	0.1000	-31.05	0.0000
party_LDPR	-0.9972	0.0090	-110.60	0.0000
party_Growth	0.0735	0.0344	2.14	0.0325
party_Rodina	0.1504	0.0147	10.25	0.0000
party_United	2.2983	0.0048	483.66	0.0000
party_Yabloko	0.6934	0.0091	75.85	0.0000

Quasi-Poisson Regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.8196	0.0696	26.14	0.0000
Count.of.text	0.0120	0.0005	23.38	0.0000
party_A_Just	-0.9983	0.4914	-2.03	0.0422
party_Comm	-2.7815	0.9041	-3.08	0.0021
party_CPRF	2.2281	0.1430	15.58	0.0000
party_Civ_Power	-1.9213	7.9407	-0.24	0.8088
party_Greens	-3.1040	3.2710	-0.95	0.3427
party_LDPR	-0.9972	0.2950	-3.38	0.0007
party_Growth	0.0735	1.1247	0.07	0.9479
party_Rodina	0.1504	0.4803	0.31	0.7542
party_United	2.2983	0.1555	14.78	0.0000
party_Yabloko	0.6934	0.2991	2.32	0.0205

Analysis: Sum of favorites

Poisson Regression

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.1211	0.0011	2822.01	0.0000
Count.of.text	0.0119	0.0000	1412.47	0.0000
party_A_Just	-1.2050	0.0083	-144.82	0.0000
party_Comm	-3.1733	0.0164	-193.85	0.0000
party_CPRF	2.1467	0.0023	948.62	0.0000
party_Civ_Power	-0.9910	0.0855	-11.59	0.0000
party_Greens	-4.2608	0.0760	-56.10	0.0000
party_LDPR	-1.2703	0.0049	-261.35	0.0000
party_Growth	0.3382	0.0159	21.29	0.0000
party_Rodina	-0.0419	0.0084	-4.97	0.0000
party_United	2.5229	0.0024	1072.77	0.0000
party_Yabloko	0.9661	0.0045	216.38	0.0000

Quasi-Poisson Regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.1211	0.0773	40.40	0.0000
Count.of.text	0.0119	0.0006	20.22	0.0000
party_A_Just	-1.2050	0.5813	-2.07	0.0382
party_Comm	-3.1733	1.1436	-2.77	0.0055
party_CPRF	2.1467	0.1581	13.58	0.0000
party_Civ_Power	-0.9910	5.9711	-0.17	0.8682
party_Greens	-4.2608	5.3060	-0.80	0.4220
party_LDPR	-1.2703	0.3396	-3.74	0.0002
party_Growth	0.3382	1.1097	0.30	0.7606
party_Rodina	-0.0419	0.5884	-0.07	0.9433
party_United	2.5229	0.1643	15.36	0.0000
party_Yabloko	0.9661	0.3119	3.10	0.0020

Conclusion

- Predictor variables are significant:
  - Number of unique Tweets
  - Number of different political parties discussed
  - Type of political party discussed
    - Always – Communists of Russia, CPRF, LDPR, Unite Russia, Yabloko
    - Sometimes – A Just Russia
- Response variable requires further study
  - Number of Twitter interactions (replies, retweets, favorites)
  - Number of replies
  - Number of retweets
  - Number of favorites

- Explore impact of different political parties onto different Twitter interactions
- Consider removing count of text from analysis
- Consider cross validation and predictive analytics
  - Sample of data for training set

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