

Comparison of Twitter Posts and Survey Responses of the Greek Parliamentary Candidates

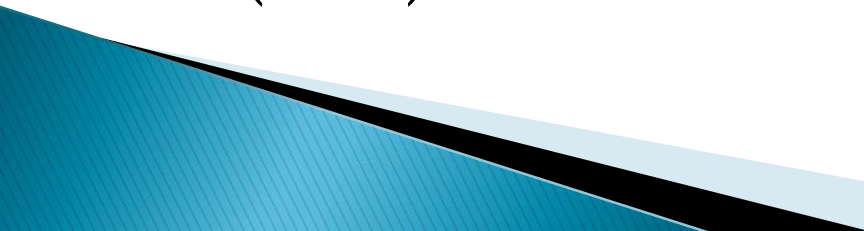
Ioannis Andreadis

Evangelgia Kartsounidou

Dimitra Papaxanthi

ESRA2019 Zagreb 17/07/2019

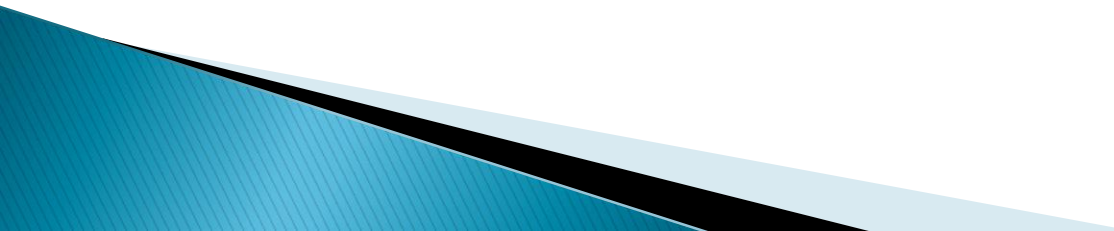
Introduction

- The increasing use of social media and internet produces immense volume data, known as big data
 - Big data are used in political campaigns (e.g. Trump's campaign in 2016)
 - Big data are also utilized by researchers in social sciences in the context of salient issues e.g.
 - Combination of other sources of data with Google big data to measure the salience of climate change (see for instance Bromley-Trujillo, R., Leising, J., & Poe, J. (2014).
 - Combination of MIP question with Google big data for identifying the relationship (see for instance Mellon, J. (2014).
- 

Our aim

- ▶ **Combine and compare social media big data and survey data in the context of salient issue, using human and computer assisted text analysis**

Research questions

- ▶ Is there high convergence between human and automated text coding?
 - ▶ Is there high convergence between survey data and Twitter data in terms of salient issues
- 

Data and Methodology

- One subgroup consisting of candidates in Greek parliamentary election, January 2015.
 - 50 candidates
- Two datasets
 - Hellenic Candidate Study 2015. The survey was conducted from February to July 2015
 - Question about the three most important problems that Greece is facing (MIP question)
 - 145 answers because some of the candidates did not fill the third problem row

Data and Methodology

- Twitter. Tweets were mined from January to July 2015
 - 11,297 tweets

Condition: Candidates should have answered the MIP question and have an active and publicly open account in Twitter

Data and Methodology

- ▶ Candidates' answers to MIP question were coded
 - by two experts and in case of disagreement the final coding was given by a referee
 - by machine
- ▶ Tweets were only automated coded because of the high cost of expert coding

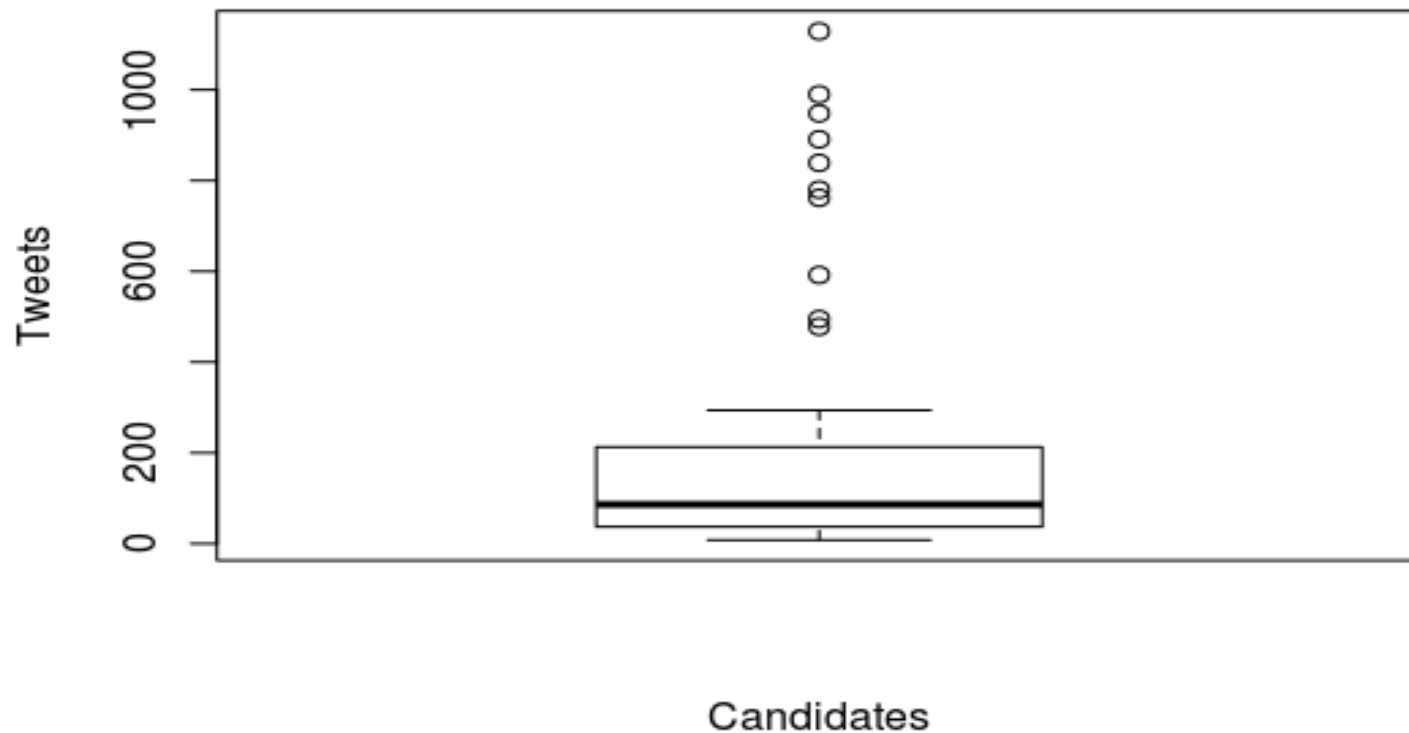
Dictionary

- Human and machine coding were realized based on the dictionary of Laver, M., & Garry, J. (2000). Estimating policy positions from political texts. *American Journal of Political Science*, 619–634
 - 1. Culture
 - 2. Economy (government's intervention)
 - 3. Environment
 - 4. Groups
 - 5. Institutions
 - 6. Law and Order
 - 7. Values

Tweets and codings per party

Party	Candidates	Tweets	Tweets/Candidates	Tweets coded	Codings
SYRIZA	14	3490	249.28	2133	3648
ND	17	3326	195.64	2091	3240
POTAMI	12	2831	235.91	1910	2874
PASOK	2	35	17.5	24	37
ANEL	5	1615	323	816	1194
TOTAL	50	11297	-	6974	10993

Tweets per candidate



Results

Human vs automated coding for MIP

- ▶ Human coding returned 145 codings– one coding for each answer
- ▶ In some cases machine coding returned more than one coding per answer (e.g. word “produce” is in the 2nd category while stem word “produc” is in the 3rd category)

Human vs automated coding for MIP

- Using as denominator the total numbers of human coding per category we found
 - First answer: 94% match (47 / 50)
 - Second answer: 82% match (41 / 50)
 - Third answer: 84.4% match (38 / 45)
- If there was no match, there were two cases
 - Hidden texture in the answer (e.g. emotion) that can not be detected by machine coding
 - Some answers could not be coded because the words were not contained in the dictionary

Tweets per category

Categories	Twitter codings	Percentage
Culture	1100	10.1
Economy	3258	29.9
Environment	249	2.28
Groups	345	3.16
Institutions	3522	32.3
Law and Order	802	7.35
Values	1634	15.0
Total	10910	100

MIP answers per category

Categories	1 st ranked issue in Hellenic Candidate Study	2 nd ranked issue in Hellenic Candidate Study	3 rd ranked issue in Hellenic Candidate Study
Culture	-	1	-
Economy	43	39	23
Environment	-	-	1
Groups	-	-	-
Institutions	6	9	15
Law and Order	1	-	-
Values	-	1	6
Total	50	50	45

Match in each category

Ranked issues in Hellenic Candidate Study 2015	Candidates matched tweets frequency with the ranked issues	Total answers in Hellenic Candidate Study 2015	Percentage
1 st	17	50	34
2 nd	20	50	40
3 rd	4	45	8.8

MIP answers for those with no match

Categories	1 st ranked issue in Hellenic Candidate Study	2 nd ranked issue in Hellenic Candidate Study	3 rd ranked issue in Hellenic Candidate Study
Culture	-	1	-
Economy	28	22	22
Environment	-	-	1
Groups	-	-	-
Institutions	4	6	14
Law and Order	1	-	-
Values	-	1	4
Total	33	30	41

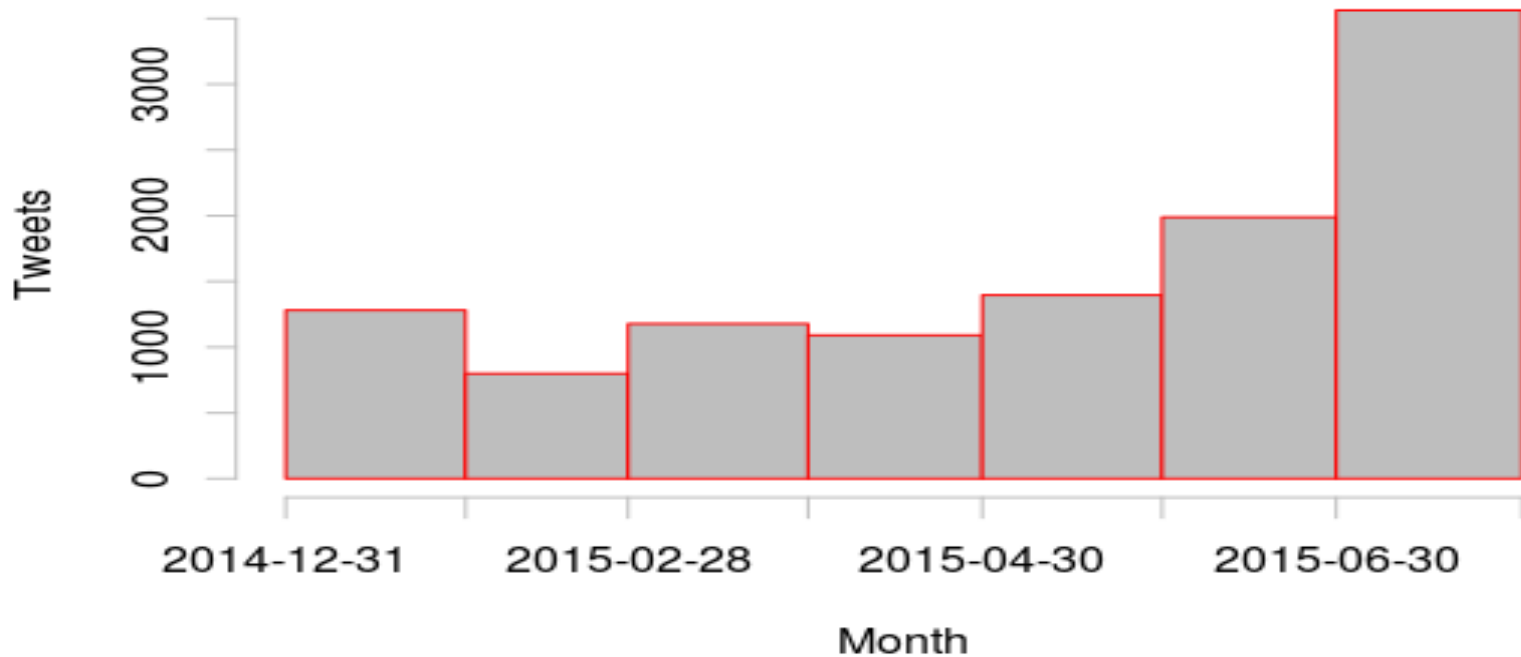
Twitter issue frequency

- ▶ For those that there was not a match and they mentioned Economy as MIP in the survey
 - Most of them were tweeted frequently about Institutions and Values
- ▶ However, for all of them Economy was one of the three most frequent issues in Twitter

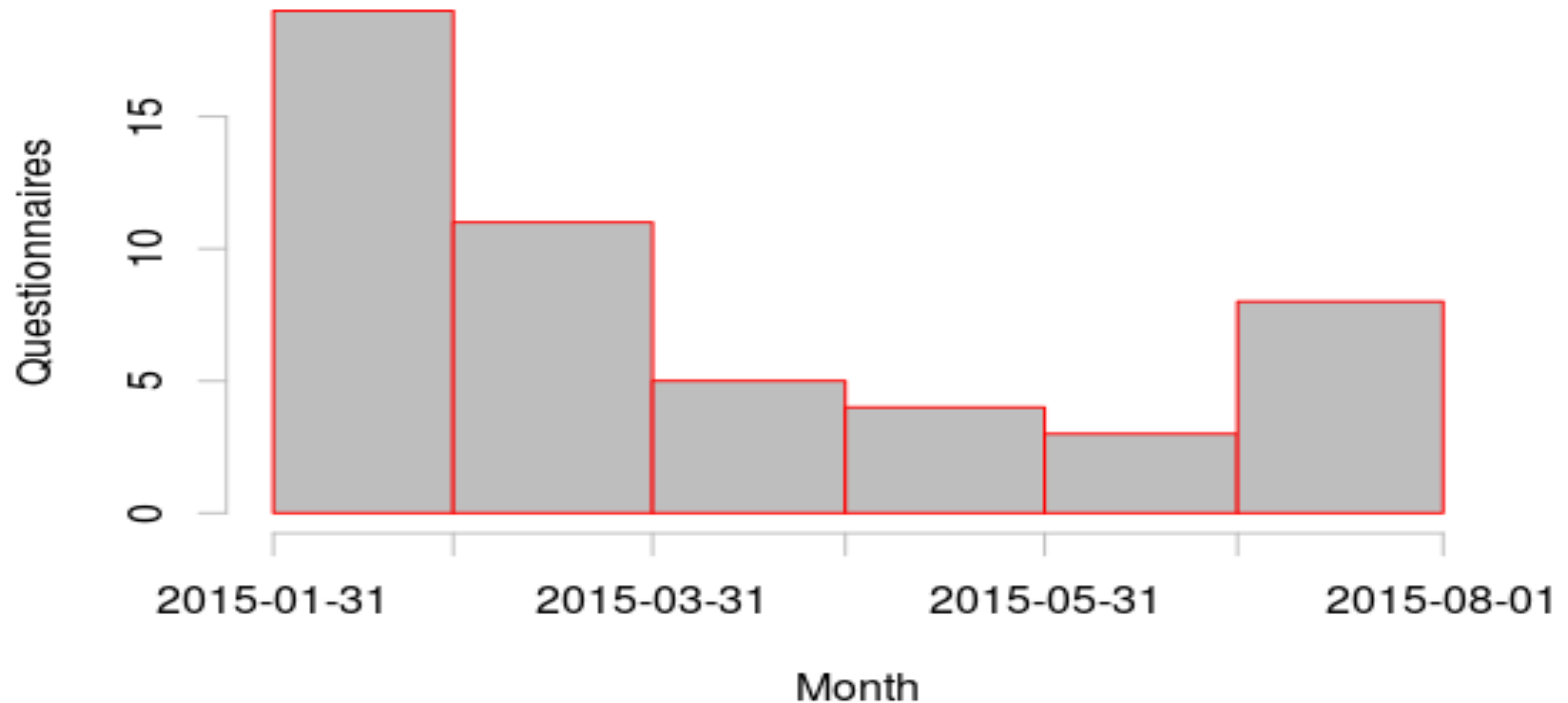
Without ranking

- ▶ When the issue ranking is ignored
 - We keep the three most frequent issues in Twitter per candidate
 - We keep unique codings per candidate answer in MIP question which is 82 unique answers
 - Then there is a 92.68 match (76/82)

Tweets per month




Questionnaires answered per month



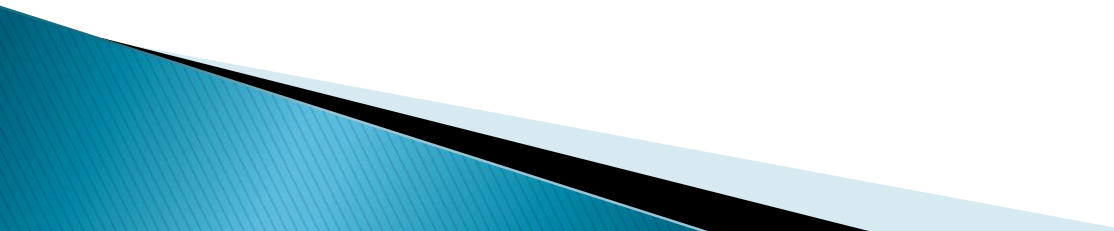
Twitter most frequent issues excluding June and July

Categories	Number of codings	Percentage
Culture	673	0.13
Economy	1473	0.28
Environment	115	0.02
Group	159	0.03
Institutions	1519	0.29
Law and Order	364	0.07
Values	826	0.16
Total	5129	100

Summarize and discussion

- ▶ Is there high convergence between human and automated text coding?
 - We found a high convergence between human and machine coding. The computer assisted coding is validated
 - ▶ Is there high convergence between survey data and Twitter data in terms of salient issues
 - The MIP answers showed that Economy and then Institutions were the most important problems in Greece 2015
- 

Summarize and discussion

- Twitter showed that candidates tweeted most about Institutions and Economy
 - Economy had gained a lot of attention in Greece in 2015
 - The difference between MIP question and Twitter can be partially explained from the time difference
 - If we don't pay attention to the ranking, there is a great convergence between tweets' frequency and MIP question
- 

Summarize and discussion

- ▶ Can Twitter big data substitute or supplement survey data?

Thank you!
Any comments,
feedback?

