# Measuring the impact of candidates' tweets on their electoral results

Dimitra Papaxanthi, Evangelia Kartsounidou and Ioannis Andreadis

Aristotle University of Thessaloniki

## Abstract

As candidates are increasingly using Twitter in their political communication, a question about its effectiveness as a political marketing tool is raised. Of course, along Twitter usage, there are more "traditional" political campaigning tools such as door-knocking, visiting businesses and social organizations etc. The aim of this paper is to study the factors that could influence the electoral performance of the candidate MPs and whether Twitter usage is one of them. To this aim we use information about candidates' previous political activity, electoral campaign tools, money spent on campaign etc acquired through survey data. We combine this dataset with data produced by candidates' Twitter activity, to study their effectiveness on the 2019 Greek Parliamentary election results. We are studying which are the factors with the greater influence on preferential votes each candidate gets in his constituency. The first data source is the Hellenic Candidate Study 2019. This dataset provides useful information about candidates' political campaigns such as campaign spending, campaign means etc. The other source is Twitter activity data. We also measure candidates' popularity using two different approaches: i) Wikipedia (number of hits on each candidate's Wikipedia page), ii) experts' evaluation. We are particularly interested in candidates' Twitter audience. We apply an innovative method of analysis of candidate's network starting from retweeters for each candidates' post. We also take into account the followers of each retweeter and the candidate's followers. In this way we explore whether the size of the Twitter audience of a candidate is related to the electoral success, or in other words, whether larger network sizes are associated with more successful candidates relatively to the party share in each electoral constituency. First indications suggest that the size of candidates' Twitter audience affect the preferential votes that a candidate receives in his constituency.

## Introduction

Social media are now widely used by politicians during their electoral campaigns. A large part of innovative Obama' first presidential campaign was held through social media. The campaign established the use of social media as political marketing tools (Tumasjan, Sprenger, Sandner, &

Welpe, 2010). They are also considered as great contributions to electoral success of the candidates. Twitter has gained a lot of attention as a big part of Trump's presidential campaign. Twitter's usage in Trump's presidential campaign triggered the interest of scholars either to discuss about Twitter's predictive power for real world events (Kassraie, Modirshanechi, & Aghajan, 2017) or to discuss twitterbots influence (Rizoiu et al., 2018). In spite of the attention that Trump's presidential campaign brought on Twitter, scholars have been studying Twitter's role in politics before 2016. However, they are mostly focusing on the content of the text of the tweets and apply frequency or sentiment analysis to the words contained in the tweets.

The aim of this paper is to study the factors that could influence the electoral performance of the candidate MPs and whether Twitter usage is one of them. We examine the impact of Twitter on the electoral performance of candidate MPs in July 7th Greek parliamentary elections by applying an innovative method of analysis of candidate's network starting from retweeters. We study the network of Greek candidate MPs on Twitter, linking Twitter data (the size and structure of the network of the followers, retweeters, and retweeters' followers of candidates) with the preferential votes they get. We explore whether the size of the Twitter network of a candidate is related to the electoral performance, or in other words, whether larger network sizes are associated with more successful candidates taking into account the party share in each electoral constituency. Candidates' tweets were mined from May 27th 2019 to July 7th 2019. For the analysis we combine datasets from Greek Candidate Survey of 2019 and Twitter. The datasets contain information about candidates' electoral campaign, electoral performance, and twitter activity and network (retweeters of each tweet, retweeter's followers, candidate's followers).

Combining information about different means of political campaign, both traditional and new (i.e. social Media) to study the impact they have on electoral performance of candidate MPs this paper offers an innovative perspective in the field of political science, in terms of political campaigning. One of the innovations of this paper lies on the fact that it studies the overall appeal of the candidates on Twitter. A simple measure that counts the total number of followers (or likes) of a candidate would not be an adequate measure of the overall appeal of the candidate on Twitter, because having multiple followers is just one of the many dimensions of the overall appeal of a candidate. To compute the overall impact of each tweet, we rely on the idea that each retweet does not have the same efficiency: a retweet from someone who has no followers does not have the same impact as a retweet by a Twitter user who has thousands of followers. For this reason, in addition to the number of retweets, we use information about the number and the network structure of the followers of users who have retweeted a tweet posted by a candidate.

In the following units we describe the theoretical background focusing on preferential voting and social media and we built our research hypothesis. Data and methodology follow. Then we present some descriptives and the findings of our model. Finally, the paper ends with some conclusions and a discussion regarding the further steps of this study.

# Theoretical background

## Preferential voting

Preferential voting is a democratic voting system which gives the opportunity to the voters to vote not only for a political party but also to choose among individual candidates on a party list that includes multiple candidates (Karvonen, 2004). Preferential voting is widespread especially in western-european democracies and there are different types mainly related to the type of the list and the party influence (Marsh, 2006). In Greece, for instance, preferential voting is described as a vote for a candidate in addition to party vote. Since 1974, Greece has had a reinforced proportional election system. Voters are only allowed to cast one vote, but they can support several candidates of the same party in a specific constituency. The number of preferential votes depends on the size of the constituency.

There are different factors that influence preferential voting, and as a result the electoral success of candidates. Many of these factors are related to the overall status of the candidate. Scholars argue that the electoral success of a candidate is highly related to the fact of being a member of parliament in previous elections (Ashworth & Bueno de Mesquita, 2008; Hirano & Snyder, 2009; Lee, 2001; Moore, McGregor, & Stephenson, 2017). Apart from incumbency, other important factors that have been pointed out by scholars are the occupation of the candidate (Mechtel, 2014), the support from lobbies (Lutz, Mach, & Primavesi, 2018) and the position of the candidate on the ballot (Lutz, 2010). Therefore, we formulate the following research hypotheses:

*H1.a: Incumbency is a significant factor for candidates' electoral performance.*

*H1.b: Candidates who were members of associations and organizations and candidates with political experience i.e. elected in regional/local level will perform better on the elections.*

Campaign activities also play a significant role in preferential voting. According to Krebs (1998), another important factor that has an impact on the electoral success of a candidate, is the money he/she spends on the electoral campaign along with the party's support for the candidate. In addition media exposure also increases the candidate's salience. The impact of campaign activities differs from country to country. In general, the level of individualized campainining differs between

and within the countries of Western Europe (Karlsen and Skogerbø, 2015). In Greece the electoral campaign is mainly party-centered, and does not focus on specific candidates.Individual candidates can run a personal campaign, receiving very limited (or not at all) funding from their party, using flyers, posters, personal contact and personal websites and other web-based platforms. The Internet and especially the use of social media, such as Facebook and Twitter, provides candidates with the opportunity to personalize their electoral campaign and make contact with voters more independently of the party leadership (Karlsen, 2011; Zittel, 2009).

*H2. The larger the amount of money spent on an electoral campaign the better the electoral performance of the candidate.*

*H3: Engaging in electoral campaign activities affects candidates' electoral performance.*

*H3a:Individualization of the electoral campaign affects candidates' electoral performance.*

## Social media and preferential voting

Social media can serve as a low-cost tool of political campaign. More and more candidates use social media in their personal campaigns to reach a broader audience and to interact with their potential voters. Lately, Twitter has been used as a predictive tool for election outcome. Many researchers investigating the predictive power of Twitter by counting the tweets mentioning a party or a candidate and then apply sentiment analysis to the tweets to check whether they are expressing intentions or emotions (Coletto, Lucchese, Orlando, & Perego, 2016; Sang & Bos, 2012; Shi, Agarwal, Agarwal, Garg, & Spoelstra, 2012; Tumasjan et al., 2010).

Apart from social media posts content analysis, there are only a few studies considering using other social media variables than text. According to Zhang (2018), likes on candidates' Facebook pages and posts are positively related to election results. Between other variables (gender, incumbency etc) the number of candidates' followers on Twitter has been used to explain the election outcome (Spierings & Jacobs, 2014). Tsakalidis, Aletras, Cristea and Liakata (2018) used information about users' retweeters network and found out that network information is better at predicting voting intentions than the content of their posts. However, some scholars are concerned about social media predictive power and suggest that some common rules should be established for this kind of analysis (Metaxas, Mustafaraj, & Gayo-Avello, 2011). They highlight social media sampling issues and identification problems of the political or not, character of tweets (Gayo-Avello, 2012)

Based on the aforementioned theoretical background we formulate the following hypothesis:

*H4: The bigger the size of the network on Twitter direct audience (followers, retweeters) and indirect audience (retweeters' followers), the better the electoral performance of the candidate.*

# Data

We use data both from the Greek Candidate Survey of 2019 and Twitter. The target population of our analysis consists of candidate MPs running for the Parliamentary Election of 2019, who satisfy the following criteria:

i) To be eligible to receive preferential votes.

ii) To have responded to the Greek Candidate Survey

The target population of the analysis are the candidate MPs who are eligible to receive preferential votes. Therefore, we exclude the Party Leaders and the ex-Prime Ministers. Party leaders or ex-Prime Ministers get all the party votes in the electoral constituencies they are running for and there is no way to measure their personal votes. Voting for them is the same as voting for their party. Therefore, including them in our analysis could create a problem in the model because of the different levels of analysis: candidate level and party level.

In addition, we exclude the Candidate State MPs. State Deputies are not elected in a specified constituency but rather throughout the country at large, in proportion to the overall electoral strength of the party of their affiliation. State MPs cannot be more than 1/20 of the total number of MPs in the Hellenic Parliament (currently, no more than 15). Each party provides an ordinal fixed list with Candidate State MPs, where preferential voting is not available for them.

The Greek Candidate Survey of 2019 is part of the Comparative Candidate Survey (CCS). CCS is a joint multi-national project with the goal of collecting data on candidates running for national parliamentary elections in different countries using a common core questionnaire. Among other issues, campaigning is a major topic in this core questionnaire. The Greek candidate survey of 2019 was conducted as a web survey after the Greek national elections of July 2019. After the parliamentary election held on 7th of July 2019, the Greek parliament consisted of six parties: i) New Democracy (ND), ii) Coalition of Radical Left (SYRIZA), iii) Movement of Change (KINAL), iv) Communist Party of Greece (KKE), v) Greek Solution (EL), amd vi) MERA 25. Out of the six parties elected, we excluded the communist party (KKE) of our analysis because the Candidate MPs of this party refuse to participate in the survey.

In addition, we searched for the Twitter accounts of the candidate MPs of the

aforementioned parties and mined their tweets from March 27 to July 7 2019. We chose to start from March 27 because the European Parliament election was held on 26th of March and by the night of 26th it was clear that the country was heading to parliamentary election within approximately a month. We found that 541 out of 1982 candidates have a Twitter account; From them 510 candidates have an active and publicly accessible Twitter account, which they have used at least once to post a tweet. However, we are interested in the impact that a tweet may have only during the electoral campaign (until July 7th) when there is still time for a voter to decide which candidate should vote for. Hence, we do not count the Twitter accounts of the candidates who did not tweet something during the period we are interested in. We also have excluded the Twitter accounts which have been deactivated afterwards, due to ethical reasons; assuming that a user who deactivates his/her account does not consent for his/her post or other information retrieved from Twitter to be publicly available.With those restrictions, we use Twitter data of 285 candidates, which results to 10,880 tweets. On average, a candidate posts approximately 38 tweets (mean=38,2) and the maximum number of tweets per candidate is 723, while the minimum number of tweets is 1.

Table 1 displays all the Candidate MPs that we use in the analysis. There are 227 Candidate MPs who are eligible to receive preferential votes and they have answered the Greek Candidate Survey of 2019. Most of the Twitter accounts we use in the analysis are accounts of the candidate MPs of New Democracy (145) that have posted 6101 tweets, in total. The candidates of SYRIZA who had a Twitter account and fitted the criteria set, are less than the candidates of ND (75 candidates). However, they seem to use Twitter slightly more frequently during the electoral campaign (3397 tweets. Although the candidates of MERA25 are under-represented on Twitter (only 9 candidates), they seem to tweet quite often (287 tweets). The candidates of KINAL (42 candidates, 944 tweets) follow. Finally, we have 161 tweets from 14 candidates of Elliniki Lisi (EL) who seem to use less frequently during their electoral campaign (11.50 tweets on average per candidate).

Finally, we use a sub-group of candidates who have both answered the Greek Candidate Survey of 2019 and they have a publicly available and active, during the political campaign, personal account on Twitter. At this point of analysis only 36 candidates have both answered the survey and they have a Twitter account which satisfied the aforementioned criteria of the analysis. Most of them (16) are candidate MPs of ND, 7 are candidates of SYRIZA and 9 are candidates of KINAL. Then EL follows with 3 candidates and MERA25 with only one candidate MP in common in the two datasets of the survey and Twitter. When it comes to their Twitter activity, the sub-group of the candidates of SYRIZA seems to tweet more frequently (815), then the candidates on ND and KINAL follow, with 317 and 154 tweets during the electoral campaign, respectively. Although

there is only one candidate of MERA25 in this sub-group, we observe an increased activity on Twitter (130 tweets). Finally, we observe 39 tweets from the sub-group of candidates of Elliniki Lisi (EL) who seem to use Twitter less frequently during their electoral campaign, as we observe in the total number of EL Twitter accounts of our analysis, as well.

**Table 1. The Candidate MPs of the analysis per party and their Twitter Activity**

| | Candidate MPs (n) | Twitter accounts (n) | Tweets (n) | Sub-group Twitter accounts (n) | Sub-group Tweets (n) |
|---|---|---|---|---|---|
| ND | 55 | 145 | 6101 | 16 | 317 |
| SYRIZA | 46 | 75 | 3397 | 7 | 815 |
| KINAL | 71 | 42 | 944 | 9 | 154 |
| EL | 23 | 14 | 161 | 3 | 39 |
| MERA25 | 28 | 9 | 287 | 1 | 130 |
| Total | 223 | 285 | 10880 | 36 | 1455 |

**Table 2. Frequencies of retweets**

| Retweets (n) | Tweets (n) | Sub-group Tweets (n) |
|---|---|---|
| 1-10 times | 7905 | 949 |
| 11-100 times | 2730 | 457 |
| 101-611 times | 255 | 13 |

Apart from the original tweets, we are also interested in the interaction of the tweets, in terms of retweeting. Retweeting usually is related to a positive reaction to the original post, even endorsement, but could also have a negative interpretation or disapproval. In any case, through retweeting, an original tweet could reach a broader audience. The tweets of the analysis have been retweeted at least once, while the maximum number of retweets is 611. On average, a tweet has been retweeted 14 times. As Table 2 demonstrates, most of the tweets (7,905) have been retweeted 1-10 times. A smaller amount of tweets (2,730) have been retweeted 11-100 times, while only 255 tweets have been retweeted more than 100 times. A similar pattern also applies to the 1455 tweets which correspond to the 36 candidate MPs who have both answered the survey and we have collected information from Twitter during the electoral campaign.

We aim to count all the Twitter accounts who were able to see each tweet of the dataset. For this reason, apart from counting the total number of followers that a candidate has, we compute the overall audience for each tweet. Audience of a tweet is considered to be the number of Twitter accounts that were able to see the tweet. Two "types" of audience are defined for each tweet. We use the term **direct audience** for the set that consists of the followers of the candidate who posted the tweet and **indirect audience** for the $n$ sets that consist of the followers of the $n$ people who have retweeted the tweet. To measure the **overall unique audience** of tweets (Twitter users that could see this tweet), we take the number of elements in the union of the $n$+1 aforementioned sets. Thus, the overall unique audience refers to the different Twitter users who were able to see the original

tweet.

Candidates' followers ids (direct audience) were also collected. Each tweet of the same candidate has (almost) the same amount of direct audience. Table 3 shows the number of direct audience the candidates have on Twitter. Most of the candidates (132) have more than 1000 direct followers. There are 86 candidates who have more than 100 until 1000 direct followers. Some candidates (46) have more than 10000, and only 21 candidates have less than 100 direct followers. There is a similar observation, also for the sub-group of candidates that we examine. Most of the candidates (18) have more than 1000 to 10000 direct followers or between 101 to 1000 (14).

**Table 3. Frequencies of Direct Audience**

| Direct followers (n) | Candidates (n) | Sub-group Candidates (n) |
|---|---|---|
| 5-100 | 21 | 2 |
| 101-1000 | 86 | 14 |
| 1001-10000 | 132 | 18 |
| 10001-216974 | 46 | 2 |

As mentioned above, in addition to the direct audience, a tweet can reach a broader indirect audience through retweets. But, each retweet does not have the same impact: a retweet from someone who has no followers does not have the same impact as a retweet by a Twitter user who has thousands of followers. Thus, we use information about the number and the network structure of the followers of users who have retweeted a tweet posted by a candidate. For example, tweet A and tweet B have been retweeted 42 times. However, the indirect audience for tweet A is 30,937 Twitter accounts and the indirect audience for tweet B is 135,923 Twitter accounts. That is because the 42 retweeters of tweet B have more followers than the 42 retweeters of tweet A.

**Table 4. Frequencies of Estimated Indirect Audience**

| Estimated_indirect_audience | Candidate MPs (n) | Sub-group Candidate MPs (n) |
|---|---|---|
| 0-100 | 44 | 8 |
| 101-1000 | 66 | 9 |
| 1001-10000 | 101 | 18 |
| 10001-149623 | 34 | 1 |

For each one of 10,880 tweets of the analysis, we mined the retweeters ids which are always up to 100 because this is the upper bound according to the standard Twitter API. When all retweeters ids have been collected, we initiated the process of collecting the followers ids for every retweeter (indirect audience). As one can observe in Table 4 candidate MPs majority is followed by 1 to 10000 followers, both in the total group and in the sub-group of candidate MPs. Most of the candidate MPs have (101 in the total group and 18 in the sub-group) have an estimated indirect

audience consisted of 1001 to 10000 indirect followers. Only 34 candidates in the total group and 1 candidate in the sub-group have more than 100001 followers.

# Methodology

The aim of the analysis used in this paper is to examine the factors that play a significant role in the electoral performance of Greek candidate MPs and to identify the effect of Twitter network on candidates' electoral performance.

The electoral performance of a candidate is measured by a percentage calculated as the ratio of preferential votes for the candidate divided by the number of votes collected by the candidate's party in the same constituency (or district). However, the number of potential votes a candidate can get is highly related to the intra-party competition in a specific constituency. The intra-party competition in a specific constituency depends on the total number of candidates per party and the maximum number of preferential votes, which are both determined by the size of the constituency. If voters choose randomly, all the candidates of a party in a specific constituency will have the same chances to be elected. In this case, we can calculate an "**expected percentage**" taking into account the probabilities of a candidate to be selected. This means that if Candidate A is running in a constituency where only one preferential vote, and a maximum of four candidates per party are permitted, a voter may either vote for the party only (without any preferential votes) or may choose a candidate. Each of these outcomes have a probability equal to 0.5. If the voter chooses a candidate, each candidate will have 0,25 probability to be selected. As a result, taking into account also the probability that someone could only vote for the specific party without voting for a candidate the expected percentage for each candidate is 0.5*0.25=0.125 or 12.5%. In addition, if Candidate B is running in a constituency where two preferential votes and a maximum of six candidates per party are permitted, with a similar way we can calculate that the expected percentage of candidate B is 16.7 %, and so on and so forth.

We calculate the "**observed percentage**" dividing the number of preferential votes of each candidate by the aggregate number of votes the party has in the constituency. Hence, candidate A who received 9782 preferential votes and his party received 17018 votes in total in his constituency, has an observed percentage of 57.5%. On the other hand, candidate B who received 9054 preferential votes and his party received 35775 votes in total in his constituency, has 25.3% observed percentage.

Finally, for the preferential votes of the candidates to be comparable, regardless of the party size in a constituency, we estimate a "**performance index**". To create the "performance index", we

divide the observed percentage by the expected. If the value of this ratio is higher than one means that the candidate is more successful than it was expected if voters selected candidates randomly; if the value is lower than one means that the candidate is less successful than it was expected. In our first example, the performance index of candidate A is 4.6 (0.575/0.125); while the performance index of candidate B is 1.5 (0.253/0.167).In the model of our analysis the "performance index" is the dependent variable.

Under the assumption that electoral performance is dependent on many factors we examined the relationship between the performance index and a number of variables that could have an effect on the index. Some of these variables are retrieved from the survey data and some others from Twitter. These variables can be grouped in four major groups each one related to our research hypotheses.

## Previous political activity and popularity

The first group contains variables with information about the past political activity of the candidate. The electoral performance of a candidate is highly related to the candidate's previous involvement with politics in terms of previous election in any level, membership in associations/organizations/unions etc. To verify this assumption we use variables containing information about the candidate's previous candidacy and previous election. We also take into account information about his/her participation in organizations or associations such as membership in trade union/business association/ religious association/environmental association/ human and civil rights association/ sports club.

Information about the candidate's previous experience in politics is also considered. Therefore, we include variables such as work in campaigns in the past (volunteer or/and paid employee, membership in regional or/and national party leadership, being major and/or vicemajor and/or regional governor and/or member of the regional government. and/or member of the European Parliament.

We also take into account the incumbency of the candidates. Assuming that the electoral performance of a candidate is highly related to his/her success in the previous elections, we have created a binary variable for those who have been elected before and for those who have not. This assumption relies on the fact that candidates who were elected during previous elections are already known at least to a part of the electorate. In addition, they have more resources to dedicate to their campaign, they have easier access to media, and increasing their network on Twitter, might not have the same significance, compared to candidates who have not been elected before.

To a certain extent, this group of variables is related to the popularity of the candidates. To

identify the popularity of the candidates we use two different variables. The first one relies more on a quantitative approach, counting the times that a candidate has been searched on wikipedia. After, we searched and gathered candidates' wikipedia handles, we were able to mine the hits that each candidates' wikipedia page received per day during the timeframe in interest from the 27th of May 2019 up to 7th of July 2019 i.e. the pre-electoral period. The other approach relies on experts' evaluations. We have asked 169 students of political science to answer whether they know a list of 200 different names of candidate MPs and they should choose one out of three choices: "I do not know him/her", or "I know him/her" or "I know him/her. He/she is a frontline politician" . These lists have been randomly created (the lists included randomly selected names retrieved from the list with all parliamentary candidates from the five parties in interest) and assigned to each student. Furthermore, for the candidate MPs that are evaluated as known by the students there was a follow-up question, asking them why they are known: due to their political activity, their professional activity or their personal life.

## Money spent on the campaign

The second group of variables contain information about the financial aspect of the candidate's campaign. The electoral performance of a candidate is highly related to the amount of money spent on the campaign. To test this assumption we employ a variable containing information about the amount of money spent on campaign and a second variable concerning the employment of a professional consultant.

## Means of electoral campaign and individualization

The third group of variables contain information about the candidate's campaign i.e the engagement to certain campaign activities and the emphasis placed on certain issues. We use a variable derived from the attention that the candidate paid on raising local issues and a second scale variable measuring if the candidate aimed to draw attention to his/her party or to him/herself. The values of the scale that stand for attracting attention for themselves as candidates could indicate individualized campaigning. There is already evidence from previous research that frontline candidates tend to focus more on themselves in the campaign (Karlsen and Skogerbø, 2015). In addition, candidates of some parties have a higher tendency to focus on themselves than candidates of other parties (Karlsen and Enjolras, 2016). Therefore, we examine their political party affiliation. Assuming that candidates of well-established parties have to compete with already popular candidates differ considerably from candidate MPs of new parties, we use a binary variable which equals 1 if the candidate runs with one of the three established parties (ND, KINAL, SYRIZA) and 0 if the candidate runs with one of the two new parties participating at the election for the first time

(MERA25, Elliniki Lisi).

Moreover, we employ a number of variables describing the significance of a variety of campaign activities for the candidate, both traditional and modern, such as door knocking, distribution of party materials, calling up voters, press releases, personal ads, personal flyers, sms (including WhatsApp, VIber etc), website, use of web platforms etc. These variables will be treated as one variable. That is because the results of mokken analysis suggest that candidates' answers in the question items' scale were not able to not group the candidates. Therefore, we calculated a new variable defined as the mean of the group of campaign activities variables per case.

## Demographic characteristics

Another group of variables that we examine are related to the demographic characteristics of the candidate MPs. Among other variables we pay attention to the gender of the candidates MPs, their age and their economic status.

## Twitter activity and network

According to our hypothesis, another factor that could influence the performance index of a candidate is the visibility he/she has on Twitter. A simple way to measure this is by counting the number of followers a candidate has on Twitter. Followers count is an independent variable of the model. However, assuming that the visibility someone's tweets have on Twitter is more complicated, we also take into account the audience someone can reach through retweets and the followers of the retweeters. Hence, the median number of retweeters of all the tweets of a candidate (median retweet count) is another independent variable that could affect the performance index. However, the number of followers and the number of retweeters are not enough to describe the overall audience of a candidate on Twitter. Therefore, we also include in our analysis what we define as the indirect audience of a candidate's tweets i.e. the followers of the retweeters.

Given that we can not get more than 100 retweeters per tweet, we had to estimate the indirect audience for each tweet. For example, there is substantial change to the audience when a tweet has been retweeted 100 times and when a tweet has been retweeted 200 times.Thus, we create a new variable based on weighted indirect audience for tweets with more that 100 retweeters. Then we use the median estimated indirect audience of tweets posted per candidate. This is another independent variable of our model that may explain the variation of the performance index under the assumption that the higher the values of estimated indirect audience the better the performance of the candidate.

Finally, another independent variable is the number of tweets that a candidate has posted during the electoral campaign, as a factor that describes the activity of a candidate on Twitter. More tweets could be related to higher visibility on Twitter and thus higher performance index. Descriptives statistics for the variables can be found in the following table.

**Table 5. Descriptives**

| Variables | N | Min | Max | Mean | SD |
|---|---|---|---|---|---|
| money_spent_oncampaign | 223 | 0 | 100000 | 5268.8 | 8933.782 |
| candidate_in_elections_before | 223 | 0 | 1 | 0.39 | 0.49 |
| previous_election | 223 | 0 | 1 | 0.12 | 0.32 |
| member_central_party_leadership | 223 | 0 | 1 | 0.17 | 0.38 |
| governor_or_vice_governor | 223 | 0 | 1 | 0.04 | 0.2 |
| constituency_councilor | 223 | 0 | 1 | 0.12 | 0.32 |
| trade_union_member | 223 | 0 | 1 | 0.16 | 0.36 |
| em_attention_candidate_to_party | 223 | 0 | 1 | 6.45 | 3.32 |
| all_campaign_staff | 223 | 0 | 62 | 9.02 | 11.79 |
| male | 223 | 0 | 1 | 0.52 | 0.5 |
| established | 223 | 0 | 1 | 0.77 | 0.42 |
| mean_views | 223 | 0 | 184.07 | 3.08 | 16.27 |
| followers | 223 | 0 | 34525 | 453.6 | 2578.85 |
| median_estimated_indirect_audience | 223 | 0 | 14543.5 | 323.8 | 1392.71 |

As the data we use have a relatively small number of observations and a large number of variables, we chose to use Lasso regression (least absolute shrinkage and selection operator) instead of OLS. Lasso includes the OLS objective function for minimization of the sum of squares of differences between the observed and the predicted values and a penalty term which pushes all coefficients all the way to 0 (Boehmke and Greenwell, 2019).

As regression coefficients are shrinked variance is decreasing, offering better prediction accuracy. Furthermore, not significant variables for the model are removed as their coefficients are reduced to 0 (Friedman Hastie, & Tibshirani. 2001).

# Findings

Considering the performance index as the dependent variable we added independent variables to the model. From a variety of independent variables, Lasso regression allows only the significant variables to enter the model. At first we used all the variables except the ones that come from Twitter data (followers, median_retweet_count, median_estimated_indirect_audience. The first group of variables related to the previous political activity of the candidates and their popularity, we observe that incumbency contributes significantly to the electoral performance of a candidate (b=0.869). This variable seems to be a key factor for candidates' electoral performance, confirming

that if a candidate has been elected before is a significant factor to his/her electoral success (H1a). In addition, in line with H1b, having been a member of the central party leadership (b=0.045), or a regional governor or vice-governor (b=0.329) or a member of the regional council (b=0.324) seem to also affect candidates' electoral performance. Finally, regarding the popularity of the candidate MPs we observe that the mean number of visits on their page on wikipedia has also a positive coefficient ( 0.005) on their electoral performance.

**Table 6. Regression model**

| Variables | Coefficients |
|---|---|
| (Intercept) | 1.3 |
| money_spent_oncampaign | 4.177*10^-6 |
| previous_election | 0.87 |
| member_central_party_leadership | 0.05 |
| governor_or_vice_governor | 0.33 |
| regional_councilor | 0.32 |
| trade_union_member | -0.02 |
| em_attention_candidate_to_party | -0.01 |
| male | 0.07 |
| established | -0.3 |
| mean_views | 0.01 |

As for the second group of variables, we observe that the money a candidate spends on his/her electoral campaign plays a significant role to the electoral performance (b= 4.178*10^-6), as it was expected (H2).

In terms of electoral campaign activities, we could not observe a significant factor that has an impact on the electoral performance of candidates. However, there is evidence that attracting attention for the party has a negative coefficient in the model (b= -0.012). Slightly inverting this finding, this could indicate that a more individualized campaign, where the primary aim of a candidate is to attract attention for himself/herself, could affect his/her electoral performance.

As for the party affiliation of the candidate, being a member of one of the established parties of the parliament (ND, SYRIZA, KINAL) displays a negative coefficient in the model (0.299). This means that a candidate MP of a new party (EL or MERA 25) could have a higher electoral performance compared to a candidate of one of the established parties that has to compete among already well-known politicians in the same ballot. Finally, regarding the demographic characteristics, we find that being a man is more related to a higher electoral performance of a candidate.

In a second step of the analysis, we use the same model including all variables but this time we add to the model the group of variables that comes from Twitter data i.e.number of tweets, followers, retweet count and estimated indirect audience (H4).

**Table 7. Regression model with Twitter variables**

| Variables | Coefficients |
|---|---|
| (Intercept) | 1.3 |
| money_spent_oncampaign | 3.99*10^-6 |
| previous_election | 0.87 |
| member_central_party_leadership | 0.04 |
| governor_or_vice_governor | 0.34 |
| regional_councilor | 0.32 |
| trade_union_member | -0.03 |
| em_attention_candidate_to_party | -0.01 |
| male | 0.07 |
| established | -0.3 |
| followers | 5.86*10^-6 |
| median_estimated_indirect_audience | 1.11*10^-5 |

The second regression model shows that the number of followers that candidate MPs have on Twitter have a significant and positive beta coefficient. Therefore, the model suggests that the performance of a candidate is increasing as he/she is gaining more followers on Twitter (H4a). That is the first indication that the size of Twitter network affects the electoral performance. To check H4b, we added in the model the estimated indirect audience variable which also has a positive and significant effect on candidate's electoral performance. The two new variables enter the model caused slight changes to the rest of the variables entering also the first model e.g. previous election coefficient shifted from 0.869 to 0.873

# Discussion

To sum up, candidates' electoral performance is affected by a series of factors such as demographics, financial factors, political experience, incumbency, wikipedia account, and Twitter posts' audience. Previous political activity in terms of incumbency and previous involvement with politics as a member of central party leadership or in the regional council either as governor or as a member seems to affect considerably the electoral performance of a candidate (H1.a, H1.b). Another factor that has a significant impact on electoral performance is the money spent on the campaign. The individualization of the electoral campaign also seems important to electoral performance of the candidates. Another interesting finding is that party affiliation also plays a significant role. Candidates of established parties experience the consequences of the higher intra-party competition, which could result in a lower electoral performance compared to candidates of new parties. Finally, sex also appears to have a significant impact on the electoral performance, concluding that men are in a or favorable position when it comes too the electoral results.

Along with these observations another important conclusion that we can draw from this paper is that candidates' Twitter activities and presence alter their electoral performance, adding to the discussion about the role of Twitter in political campaigns and electoral performance. The bigger the audience that is able to see candidates' posts, the better the candidates' performance. That means that Twitter is offering new rings for politics.

However, our data is not completely harmonized yet. Out of the 223 respondents to the survey, we had Twitter data for only 36 of them. It is our primary goal to increase the number of survey respondents with special focus to the ones who have Twitter accounts and we could mine their data. Since we have already gathered all candidates Twitter accounts and mined the relevant data, we can not have Twitter data for the 187 candidates because they do not have Twitter accounts.

There is also more work to be done in order to enhance the accuracy of the variables generated from Twitter activity. We aim in performing further analysis to the tweets mined. Content analysis could provide some useful insights. Followers and followers of retweeters are clear indicators of the audience of a tweet but there are some other factors that may increase the number of people who were able to see one tweet. For example, content analysis could provide valuable information about the number of tweets containing hashtags because these tweets could have been seen from the Twitter accounts clicked on the hashtag.

A closer look to Twitter data indicates that some candidates' Twitter accounts included in the data that they may be the aspiration for further research. For example, a candidate is detected who appear to have many thousands of followers, but yet his tweets are retweeted very few times in comparison with the number of their followers. A quick look to their followers suggest that many of them are from other countries so it can be an issue for them to understand what the tweets are talking about. In addition, many of this candidate's followers seem to be inactive or protected Twitter accounts. If this is the case, then the audience that actually were able to read that tweets is substantially lower than the followers that Twitter account that initially posted the tweets has. Of course, it is possible that one can buy followers as this service is provided lately.

It is also noticed that some tweets have been retweeted more than once by the same person using multiple accounts. It is possible that multiple Twitter accounts can be controlled by a Twitter bot which performs different actions on Twitter such as retweets. Assuming that multiple Twitter accounts controlled by the same person have real Twitter accounts as followers, the high number of

retweets and the large indirect audience may affect negatively the indirect tweets' audience if they could realize that the retweeter is spamming.

## Acknowledgements

This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning 2014- 2020» in the context of the project "Measuring the impact of candidates tweets on their electoral results" (5047884).[1]




## References

Ashworth, S., & Bueno de Mesquita, E. (2008). Electoral Selection, Strategic Challenger Entry, and the Incumbency Advantage. *The Journal of Politics*, *70*(4), 1006–1025. https://doi.org/10.1017/S0022381608081024

Boehmke, B., & Greenwell, B. M. (2019). *Hands-on machine learning with R*. CRC Press).

Coletto, M., Lucchese, C., Orlando, S., & Perego, R. (2016). Electoral predictions with Twitter: a machine-learning approach. In *6th Italian Information Retrieval Workshop*. Retrieved from http://eprints.imtlucca.it/3489/

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1, No. 10). New York: Springer series in statistics).

Gayo-Avello, D. (2012). "I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper"; -- A Balanced Survey on Election Prediction using Twitter Data. *ArXiv Preprint ArXiv:1204.6441*. Retrieved from http://arxiv.org/abs/1204.6441

Hirano, S., & Snyder, J. J. M. (2009). Using Multimember District Elections to Estimate the Sources of the Incumbency Advantage. *American Journal of Political Science*, *53*(2), 292–306. https://doi.org/10.1111/j.1540-5907.2009.00371.x

Karlsen, R., & Skogerbø, E. (2015). Candidate campaigning in parliamentary systems: Individualized vs. localized campaigning. *Party Politics*, *21*(3), 428-439.

Karlsen, R., & Enjolras, B. (2016). Styles of social media campaigning and influence in a hybrid

1 Some of the data of the Greek Candidate Study used in this paper have been collected as part of DataPoPEU project, a research work supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: 3572).

political communication system: Linking candidate survey data with Twitter data. *The International Journal of Press/Politics*, *21*(3), 338-357.

Karlsen, R. (2011). A platform for individualized campaigning? Social media and parliamentary candidates in the 2009 Norwegian election campaign. *Policy & Internet*, *3*(4), 1-25.

Karvonen, L. (2004). Preferential Voting: Incidence and Effects. *International Political Science Review*, *25*(2), 203–226. https://doi.org/10.1177/0192512104041283

Kassraie, P., Modirshanechi, A., & Aghajan, H. K. (2017). Election Vote Share Prediction using a Sentiment-based Fusion of Twitter Data with Google Trends and Online Polls. In *DATA* (pp. 363–370). Retrieved from https://www.scitepress.org/papers/2017/64843/64843.pdf

Krebs, T. B. (1998). The Determinants of Candidates' Vote Share and the Advantages of Incumbency in City Council Elections. *American Journal of Political Science*, *42*(3), 921–935. https://doi.org/10.2307/2991735

Lee, D. S. (2001). The Electoral Advantage to Incumbency and Voters' Valuation of Politicians' Experience: A Regression Discontinuity Analysis of Elections to the US. *National Bureau of Economic Research*, *No.w8441*. Retrieved from http://www.nber.org/papers/w8441

Lutz, G. (2010). First come, first served: The effect of ballot position on electoral success in open ballot PR elections. *Representation*, *46*(2), 167–181. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/00344893.2010.485808

Lutz, G., Mach, A., & Primavesi, R. (2018). Interest Group Support and Electoral Success in the Swiss Elections of 2015. A Candidate Survey Analysis. *Swiss Political Science Review*, *24*(4), 487–509. https://doi.org/10.1111/spsr.12337

Marsh, M. (2006). The Voters Decide?: Preferential Voting in European List Systems*. *European Journal of Political Research*, *13*(4), 365–378. https://doi.org/10.1111/j.1475-6765.1985.tb00132.x

Mechtel, M. (2014). It's the occupation, stupid! Explaining candidates' success in low-information elections. *European Journal of Political Economy*, *33*, 53–70. Retrieved from https://www.sciencedirect.com/science/article/pii/S0176268013000980

Metaxas, P. T., Mustafaraj, E., & Gayo-Avello, D. (2011). How (Not) to Predict Elections. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and In 2011 IEEE Third International Conference on Social Computing* (pp. 165–171). IEEE. https://doi.org/10.1109/PASSAT/SocialCom.2011.98

Moore, A. A., McGregor, R. M., & Stephenson, L. B. (2017). Paying attention and the incumbency effect: Voting behavior in the 2014 Toronto Municipal Election. *International Political Science Review*, *38*(1), 85–98. https://doi.org/10.1177/0192512115616268

Rizoiu, M. A., Graham, T., Zhang, R., Zhang, Y., Ackland, R., & Xie, L. (2018). # DebateNight: The Role and Influence of Socialbots on Twitter During the 1st 2016 US Presidential Debate. In *Twelfth International AAAI Conference on Web and Social Media* (pp. 300–309). California. Retrieved from

https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/viewPaper/17886

Sang, E. T. K., & Bos, J. (2012). Predicting the 2011 dutch senate election results with Twitter. *Proceedings of the Workshop on Semantic Analysis in Social Media*. Association for Computational Linguistics. Retrieved from https://dl.acm.org/citation.cfm?id=2389976

Shi, L., Agarwal, N., Agarwal, A., Garg, R., & Spoelstra, J. (2012). Predicting US Primary Elections with Twitter. In *workshop social network and social media analysis: methods, models and applications* . (NIPS).

Spierings, N., & Jacobs, K. (2014). Getting Personal? The Impact of Social Media on Preferential Voting. *Political Behavior*, *36*(1), 215–234. https://doi.org/10.1007/s11109-013-9228-2

Tsakalidis, A., Aletras, N., Cristea, A. I., & Liakata, M. (2018). Nowcasting the Stance of Social Media Users in a Sudden Vote. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management - CIKM '18* (pp. 367–376). New York, New York, USA: ACM Press. https://doi.org/10.1145/3269206.3271783

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Icwsm*, *10*(1), 178–185. Retrieved from www.aaai.org

Zhang, X. (2018). Social media popularity and election results: A study of the 2016 Taiwanese general election. *PLOS ONE*, *13*(11), e0208190. https://doi.org/10.1371/journal.pone.0208190

Zittel, T. (2009). Lost in technology? Political parties and the online campaigns of constituency candidates in Germany's mixed member electoral system. *Journal of Information Technology & Politics*, *6*(3-4), 298-311.