Facebook likes and public opinion: Predicting the 2015 Finnish parliamentary elections

Tapio Vepsäläinen*, Hongxiu Li, Reima Suomi

Information Systems Science, Department of Management and Entrepreneurship, Turku School of Economics, University of Turku, FIN-20014, Finland

ARTICLE INFO

ABSTRACT

Predicting an electoral outcome using 'big' social media data is a new research theme that has emerged due to the exponential growth of social media. Mainstream research in this field focuses on Twitter and Facebook. This study examines the extent to which Facebook Likes can be used to predict electoral outcomes. The study was conducted in Finland, where 2146 candidates across the country competed for the 200 seats in the Finnish Parliament in 2015. The primary objective of this study was to analyze the degree to which Facebook Likes could be used to predict the outcome of the 2015 Finnish parliamentary election. A complete candidate-level dataset was compiled from Facebook and other sources to analyze the relationship between Facebook Likes and the vote shares of candidates. The initial sample contained 2.7 million Facebook Likes, which were gathered directly from candidates' official Facebook pages. The prediction based on Facebook Likes was less accurate than using incumbency and traditional polling on whether a candidate would be elected. However, Facebook Likes and votes were found to have a significant positive relationship. Hence, social media data, like Facebook data, might be a significant but weak indicator of electoral success.

1. Introduction

The diffusion of social media has enabled new opportunities for measuring public opinion. Predicting the outcome of elections using 'big' data from online social networks is a new form of political forecasting and is mainly used for making predictions about election outcomes by collecting relevant data from social media.

Some authors have suggested that analyzing data collected from social media during an election campaign could be a useful supplement to traditional polling methods (Sang & Bos, 2012; Schober, Pasek, Guggenheim, Lampe, & Conrad, 2016). The main advantages of measuring public opinion through social media are affordability and velocity because large amounts of social media data can be collected by algorithms at short intervals with few additional costs (Schober et al., 2016). In addition, compared to traditional polling, social media may enable the continuous tracking of public opinion in real time (Sang & Bos, 2012). In traditional polling, public opinion is measured in comprehensive random sample surveys. However, gathering high-quality survey data is expensive, time-consuming and requires significant effort (Schober et al., 2016). Traditional polling is also influenced by errors and biases, such as researcher bias and increasing non-response rates (De Leeuw & De Heer, 2002). Thus, a 'data-driven' approach could offer a solution to the limitations of traditional polling.

People have increasingly integrated social media into their lives. Prior studies show that the majority of political candidates have adopted social media to help with their campaigning (Gulati & Williams, 2013; Hong & Nadler, 2012; Lilleker et al., 2011; Strandberg, 2013). Social media has also been found to exert a significant impact on the behavior of people regarding making friends, purchasing habits, opinions and preferences (Asur & Huberman, 2010). Some researchers have also attempted to use social media data to find the causalities between social media data and elections; prior studies have made plausible findings in predicting elections in different countries (Schoen et al., 2013).

In the recent years, social media has become popular in Finland and people are increasingly using social media in their lives. According to the recent report from a major tele-operator in Finland (DNA, 2015), 68% of Finns aged between 16 and 74 use Facebook. More and more Finnish politicians are also increasingly using social media. As Strandberg (2013) found: in the 2011 Finnish parliamentary elections 88% of the candidates had a Facebook presence. In addition, the Finnish parliamentary system has several positive characteristics that make the study of this topic in Finland feasible. Due to the open list proportional representation system, parliamentary elections are very personalized and the role of the candidate is central (Strandberg, 2013). This makes the simultaneous analysis of numerous candidates (2146)

---

possible. However, in the prior literature, few studies have attempted to examine the impact of social media on political elections in Finland and little is known about whether social media data, like Facebook data, can be used to predict the outcome of Finnish political contests.

In this study, the main research question is: 'Was it possible to forecast the results of Finland's 2015 parliamentary elections using candidates' Facebook page Likes?'. The aim of this research is to study whether Facebook data can be efficiently used to forecast political elections and to examine what Facebook Likes reveal about Finnish citizen's political opinions during elections. To analyze the relationship between a politician's Facebook Likes and vote share, we compiled a candidate-level dataset with data taken from Facebook and the Finnish Ministry of Justice. In line with the prior research, the volume of Facebook Likes was found to predict election results only to a limited extent. This paper helps to explain the limitations of forecasting elections by using social media data and what can be done to help overcome those limitations. It also helps to build some understanding of citizens' and politicians' online political behavior.

The remainder of the article is organized as follows. First, we present a literature review on related research about forecasting political elections by using social media data. Next, we describe the research methods, including data collection and data analysis. Then we present and discuss the results of the analysis. Finally, conclusions are drawn followed by identifying avenues for future research.

2. Related research

The revolutionary rise of social media has attracted considerable attention from academics. Some research has attempted to use social media to predict real world outcomes. Prior research evidence suggests that social media is an important indicator for guiding user behavior and steering their preferences in such contexts as brand promotion, sales prediction, website promotion and political elections (Cameron, Barrett, & Stewardson, 2015; Gerodimos & Justinussen, 2015; Kudeshia, Sikdar, & Mittal, 2016; Petrocchi, Asnaani, Martinez, Nadkarni, & Hofmann, 2015).

Recently, the importance of social media for political election campaigning has received a lot of attention. The two key topics related to forecasting elections with social media are the social media usage of politicians and the social media usage of citizens. Prior research suggests that politicians have adopted social media mainly as a campaigning tool and that there are different factors that influence their usage patterns (Strandberg, 2013). Meanwhile, the general public have widely adopted social media and are increasingly engaging in political discussions within social media. As Bond et al. (2012) suggest, social media can mirror the political behavior of citizens, and social media content may also influence their political decision-making, but the overall dynamics of this interaction are very complex (Bond et al., 2012).

Politicians are increasingly using social media in campaigns. Online campaigning became the norm in the early 2000s, and social media has enabled new ways for political organizations to connect with citizens (Chadwick, 2007). Social media first appeared as a significant campaign tool in the 2006 US midterm elections, although it was uncertain how candidates should integrate social media, like Facebook, into their campaigns (Gulati & Williams, 2007). Regardless of the early prudence, some campaigns had promising results with YouTube and MySpace. The breakthrough came with the victory of Barack Obama in the 2008 presidential elections – a result which has been widely praised for using successful social media strategies. Since then, the diffusion of social media among politicians has been unstoppable (Gulati & Williams, 2013; Hong & Nadler, 2012; Lilleker et al., 2011; Strandberg, 2013).

Prior research found that multiple variables influence the adoption of social media for political campaigning, such as party orientation (Vergeer, Hermans, & Sams, 2013), culture (Lilleker et al., 2011), the size of the party (Aragon, Kappler, Kaltenbrunner, Laniado, & Volkovich, 2013), education and incumbency (Strandberg, 2013). However, Politicians tend to use social media primarily for one-way information distribution (Aragon et al., 2013; Klinger, 2013), not for two-way communication between politicians and citizens. Offline campaigns are still more popular than social media campaigns (Aldrich, Gibson, Cantijoch, & Konitzer, 2015).

Nevertheless, social media may influence an electorate's decision-making. Prior studies have investigated how social media influences the political activities of citizens and have produced unclear research findings. On one hand, some researchers suggest that social media may influence voting decisions (Aldrich et al., 2015; Bimber, Cunill, Copeland, & Gibson, 2015; Bond et al., 2012; David, Zhitomirsky-Geffet, Koppel, & Uzan, 2016). On the other hand, some researchers have observed situations where social media has a negative impact on political activity (Baumgartner & Morris, 2009; Lu, Heatherly, & Lee, 2016; Theocharis & Lowe, 2015). For users who tend to seek political information online, or who would otherwise lack the resources to participate, social media functions as an enabler (Carlisle & Patton, 2013; Gil de Zúñiga, Molyneux, & Zheng, 2014). For users that are exposed to political views they disagree with, it has an adverse impact on political engagement (Lu et al., 2016). Social media's influence on citizen participation in politics seems to be dependent on the context and individual. After Donald Trump's election as US President in 2016, many people and organizations have speculated on the influence and role of social media on election campaigns.

In elections, traditional polling has been the dominant tool for forecasting the results of political contests. In traditional polling the prediction is made through a scientifically optimized sample. According to the experts in this field, the most important aspect for successful or correct polling is to select an unbiased sample representing the likely voters (Metaxas, Mustafaraj, & Gayo-Avello, 2011). Clearly, social media data is different from data collected in polling, and the selection of an unbiased sample is beyond what social media can do today. As Gayo-Avello (2013) argued, researchers have two options: *"acknowledging the bias in the user base but arguing the method is bias tolerant, or assuming that bias-tolerant methods are not feasible and trying to weigh each user and tweet according to the demographic composition of the voting population."* (p. 662).

Prior research using social media data mainly employ statistical forecasting models to predict election outcomes. Some research combines social media data with other relevant variables to predict election outcomes, such as economic performance, government performance, incumbency and various political indexes. MacWilliams (2015) combined a statistical forecasting model with Facebook data to predict US senate votes by including incumbency, the estimated advantages generated from Facebook for candidates and the partisan vote index, which measures past election results. Lewis-Beck (2005) suggests that when measuring the quality of a statistical model in elections, one should take the following four factors into account: lead (how many days before election day was the forecast), reproducibility (how easy is it to repeat the experiment), accuracy (mean absolute error, Standard error of estimate, etc.) and parsimonious (variable quality over quantity) (Lewis-Beck, 2005).

Some studies on social media election forecasts are presented in Table 1. The table shows that the main research stream has focused on Twitter and Facebook. Twitter's high popularity as a data source can be explained by the convenience of retrieving data with the simple API offered by Twitter. In contrast, collecting data from Facebook and other social media channels requires a much higher commitment and a broader skill set, e.g. programing skills.

In prior studies, sentiment analysis and algorithm analysis are the main research methods for investigating the predictive power of social media in elections. The literature reveals conflicting findings on the predictive power of social media data, whether Twitter or Facebook, for making electoral predictions. As presented in Table 1, there is plenty of evidence that Twitter might be a strong tool for making electoral