Predicting Swedish Elections with Twitter: A Case for Stochastic Link Structure Analysis

Nima Dokooehaki  
Swedish Institute of Computer Science (SICS)  
SE-16429 Kista, Sweden  
Email: nima@sics.se

Filippia Zikou  
Royal Institute of Technology (KTH)  
SE-16440 Kista, Sweden  
Email: filippia@kth.se

Daniel Gillblad  
Swedish Institute of Computer Science (SICS)  
SE-16429 Kista, Sweden  
Email: dgi@sics.se

Mihhail Matskin  
Royal Institute of Technology (KTH)  
SE-16440 Kista, Sweden  
Email: misha@kth.se

Abstract—The question that whether Twitter data can be leveraged to forecast outcome of the elections has always been of great anticipation in the research community. Existing research focuses on leveraging content analysis for positivity or negativity analysis of the sentiments of opinions expressed. This is while, analysis of link structure features of social networks underlying the conversation involving politicians has been less looked. The intuition behind such study comes from the fact that density of conversations about parties along with their respective members, whether explicit or implicit, should reflect on their popularity. On the other hand, dynamism of interactions, can capture the inherent shift in popularity of accounts of politicians. Within this manuscript we present evidence of how a well-known link prediction algorithm, can reveal an authoritative structural link formation within which the popularity of the political accounts along with their neighbourhoods, shows strong correlation with the standing of electoral outcomes. As an evidence, the public time-lines of two electoral events from 2014 elections of Sweden on Twitter have been studied. By distinguishing between member and official party accounts, we report that even using a focus-crawled public dataset, structural link popularities bear strong statistical similarities with vote outcomes. In addition we report strong ranked dependence between standings of selected politicians and general election outcome, as well as for official party accounts and European election outcome.

I. INTRODUCTION

Twitter as a micro-blogging platform has become a widely anticipated platform within politics domain. Such anticipation is seen world-wide from both politicians and voters, as the real-time and democratic nature of information dissemination on Twitter allows express and follow opinions freely. Such phenomena is visible in Sweden, as an increasing number of politicians leverage social media to communicate their daily activities to masses. Twitter became the centerfold medium for observing the course of public political discussions during and in the aftermath of the Iranian presidential elections back in 2011. Ever since, public and private bodies have been eager to search, browse and track the politicians, their respective parties and to gain an understanding of their respective reputation on Twitter. Fascination with aggregation and mining Tweets in order to predict possible outcome of real world events has led to a large number of case-studies [1].

Obviously, political domain was not an exception either. What is fascinating is mix of positive [2] and negative attitudes [3] in political data mining from Twitter to forecast vote outcomes. On the other hand, an overview of literature shows how much existing methodologies for mining democratic data on Twitter are mostly focused on analyzing the content of Tweets. For instance, a number of researchers have proposed common practices for classifying polarity of tweets [4] and sentiment analysis for the task of outcome prediction [2]. Comparatively speaking, studies mining topological features of underlying tweet networks are less popular. This has been visible in several works that try to combine social network and content analysis techniques to shed light on both interaction and polarization characteristics of networks involved [5], [6], [7]. Link prediction [8] techniques have gained strong momentum in recent years due to their adoption along with their probabilistic and scalable capabilities for analyzing and predicting dynamism of link-structures in the social networks [9]. We highlight two advantages of link prediction techniques: first, link prediction could capture density of conversations about parties and their respective members, which in turn should reflect on their popularity within and across the network. Second, dynamism of interactions which reflects on inherent shift in popularity of accounts of politicians, can in turn be captured by a stochastic link mining technique. Finally, link prediction algorithms are topic-sensitive thus valuable to political opinion mining.

Thus within this manuscript we make a case for studying evolution of link structures surrounding political tweets to capture the popularity of parties and their respective members. Our implementation entices an exploratory analysis comprised of basic and advanced social network mining studies aimed mainly at understanding the dynamics of parties interaction within and without their neighborhood. We present an in-depth study of dynamics of party and individual members link-structure based popularities along the time-line of two European and general elections during 2014 on Sweden. We validate the resulting popularity estimates using the official statistics of voting outcomes. We consider a combined selection of mentions and retweets as the source of interactions. This choice combined with the link mining algorithm centralize the clusters of network vertices around main topics of discussions rather than individuals [10]. This also helps with selection of relevant topics thus avoiding noisy content, as it is vital to sanity of social media studies. The data presented
throughout the paper is the result of a continuous crawl of public Twitter streams for a period of eight months grossing to 7 million tweets from Sweden.

Specific contributions of this paper, as follows:

- Social network analysis including fitness studies for conversation graphs of European and General election timelines,
- Popularity modeling using personalized PageRank of political parties and their respective members using a stochastic link-structure mining approach,
- Statistical correlation modeling between the vote outcome and inferred top follower (out-degree) of party and individual accounts

II. RELATED WORK

A. Political Conversations on Twitter

Twitter as a conversation medium [11], provides several means for individuals to communicate with one another. Mentions, or simply replies, are the most common way to interact with another user while mentioning their id (using @ sign) in the content. Honeycutt and Herring observed that tweets with mentions showed higher variance for content [12]. Retweets are another conversation means for sharing other user’s tweets, usually as a whole. Hashtags seem to be the main method used to report an event or communication, while at the same time making similar themed conversations findable. As a result, hashtags can add semantics to the stream of tweets by giving them semantic structure [13]. A number of researchers study the specific characteristics of political interactions on Twitter. Conover et al. [5], analyze the retweet and mention networks one and half month leading to 2010 US midterm election and they differentiate between retweet and mention networks based on limited correlation between right and left oriented users in retweet network. Following Conover’s study we also have chosen to take into account mentions and retweets. Within our study chosen tweets are a combination of both retweets and mentions.

B. Predicting Elections on Twitter

With increasing popularity of social media, analysts became interested in understanding the role of Twitter in elections. Study published by Livne et al.[6], focuses on the use of Twitter by 2010 midterm elections in the US and through correlating between network structure, content and election results, they can accurately predict results of the election. Tomasjan et al.[2], analyze political sentiment from text of tweets of Germany’s multi-party election in 2009 and show that Twitter can complement traditional methods of political forecasting. On the other hand, Jungherr [14] focuses on reported studies on hashtag usage and state that such political predictions are not possible. Gayo-Avello [15] summarized existing work for predicting elections, and proposed a scheme to characterize Twitter prediction methods. The scheme covers every aspect from data collection to performance evaluation, through data processing and vote inference. We will highlight in the next sections that advanced link analysis techniques are recent.

C. Political Link Mining and Recommendations on Twitter

Existing works which focus on analyzing characteristics of interactions with politicians or their parties have been recent [7]. This is also the case also with evidence on analyzing structure of interactions with politicians on Twitter. This has been highlighted on a recent survey of election prediction on Twitter [16]. In 2012 during a primary in California, Nielsen [17] reported that in three out of four races, most frequently mentioned candidate won. Gaurav et al [18] propose for prediction model based on the number of times the name of a candidate is mentioned in tweets. They develop several methods to augment the counts by counting not only the presence of candidate’s official names but also their aliases and commonly appearing names. They report success in predicting the winner of three presidential elections in Latin America in 2013. Cameron et al.[19], studies data from the 2011 New Zealand General Election and candidates social networks on Facebook and Twitter. They report a statistically significant relationship between the structural size of social networks and election voting and election results. Makazhanov and Rafiei [20] focus on Alberta 2012 general election and show that political taste of users can be predicted according to their interaction with political parties. They propose prediction models using contextual and behavioral features specially interaction factors (retweets, following size, etc). Given the increasing literature on interaction analytics, within this manuscript we make a case for link-structure mining and analysis. Getoor and Diehl [21], generalize the notion of link mining as data mining techniques that explicitly consider links, when building predictive or descriptive models of the linked data. One of the most anticipated of link mining approaches, is the link prediction problem, which was first introduced by Liben-Nowell and Kleinberg [8]. Tied directly to network evolution, link prediction asks how much evolution of a social network can be modeled using features local and salient to the network itself. There has been increasing attention in link prediction on Twitter, that use network proximity for network popularity estimation. Zou and Fekri [22] propose two approaches to exploiting both popularity and similarity for link recommendation. The first approach employs the rank aggregation technique to combine rankings generated by popularity-based and similarity-based recommendation algorithms. Yuan et al.[23], note that sentiment analysis studies miss out on the value that incorporation of social relationships. Thus they exploit sentiment proximity for link prediction on a Twitter dataset in one month during U.S. 2012 political campaign along the follows relationship between users. Similar to Yuan et al. [23], we focus on studying how existing link-structure of the network can be used to build a party-centric popularity, although we don’t leverage on sentiment features in this work. As a result, the novelty of this work lies within showing how popularity estimates using an authoritative and stochastic link-structure mining could reveal popularity rankings that could
closely correlate with voting outcome, and this is done without taking sentiment features into account.

III. SWEDISH ELECTIONS ON TWITTER

A. Previous Swedish Election Analysis on Twitter

Swedish elections have been subject to data analysis in the past. Larsson and Moe [24] focused on the 2010 Swedish election campaign and the related discussions in Twitter in that period. They aimed at identifying the various types of users based on their usage of Twitter. They describe the political network and how users relate to each other by assigning the users to nodes and designing their interconnections he measured the density, centralization and their position at the social graph. This work focused on crawling all the tweets with #val2010 hashtag which was the prominent hashtag for the 2010 elections. Evidently, the day of the elections showed most traffic of political tweets and the most active users were party members or journalists. We also observed the same impact by party members, as presented later on. Our advanced link mining experiments help exploiting inherent value of such interactions as presented later on.

B. 2014 Swedish Elections

We outline two time-lines in 2014 where we crawled tweets from: European parliamentary elections (EU) and Swedish General elections (General). Table I summarizes the Swedish names and their title in English along with their results as reported by Swedish election authority 1.

European Parliamentary elections took place on 25 May 2014. At the election, twenty members of the European Parliament (MEPs) were from the Swedish political landscape. The outcome of the vote, saw the rise of left-wing parties in popularity, specially Social Democrats which took the highest proportion of votes. Whilst the right-leaning parties took the lower vote percentage, specially Moderate Party, who were at the time in power. General elections were held in Sweden on 14 September 2014. During the elections, The Center-right Alliance coalition (comprised of Moderate Party, Liberal People’s Party, Center Party and Christian Democrats) competed for a third term in office. In contrast to the 2010 election, three dominant left-leaning parties (Social Democrats, Green Party and Left Party) took the lead and won, while Alliance came second, and Sweden Democrats, a nationalist party took the third place. Finally, the Feminist Initiative and a left leaning party, did not secure the 4% needed threshold to join the government.

IV. DATA ANALYSIS FRAMEWORK

Within this section we present the data analysis framework along with experiments that we have devised to analyze the Twitter data at hand. We begin by presenting the characteristics of the data itself. In the subsequent section we present community modeling and analysis of interactions, focusing on basic and advanced features of the network. In the later section we leverage these networks for popularity analysis.

A. Data Aggregation

We focus crawled public Twitter streams 2 from the beginning of February 2014 until the end of September 2014, a while after the general election. We gathered a total of 7,000,000 tweets within Sweden. More than 70% are written in Swedish language. From these tweets almost the 2,000,000 had political content, referring to either the elections, debate, a specific party or a politician. From these data we identified 21,000 users, from which we were able to relate to 130,000 profiles associated with a specific party. Crawler used three kind of filters, one for the coordinates of the country, one for political and election related hashtags (e.g. #svpol, #val2014) and one for the political profile accounts (e.g. @Folkpartiet, @Feministerna). A selection of most frequent used hashtags per each party was presented in table I.

Figure 1 visualizes the frequency of tweets during the timeline of the crawl. As seen from the image, we can spot two bursts on the plot. Two visualized bursts are attributed to European parliamentary and then Swedish general elections. It is interesting to observe higher density of tweets during the European elections as compared to national elections. This is further motivation to distinguish between two election timelines, specially for the European parliamentary election.

B. European Parliamentary Election Time-Line

In this section we focus on the time-line during mainly the European elections. Tweets analyzed for this time-line corresponds to the May month. Choice of studying this time-line comes from the fact of observing hight density of tweets, compared to general election. We focus on the community of political parties on Twitter. To model such community formation we study the graph shaped at the heart of the interactions taking place in between political parties. While researchers differentiate between analyzing Retweet and Mention graphs, for our study we combine both of these graphs. Such community analysis will help us understand how coherent the parties are in terms of their internal formation, meaning how much they interact with their own members as well as how much other parties interact with them.

There are practical issues with respect to producing such plot. For instance, we have realized that members usually

1 http://www.val.se/

2 https://dev.twitter.com/streaming/public
### Swedish Political Parties

<table>
<thead>
<tr>
<th>Party Acronym</th>
<th>Title</th>
<th>European Election</th>
<th>General Election</th>
<th>#Posts</th>
<th>#Mentions</th>
<th>#Accounts</th>
<th>Hashtag Valence</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>Moderates</td>
<td>13.65%</td>
<td>23.33%</td>
<td>558</td>
<td>12124</td>
<td>2459</td>
<td>#jobbvalet #svrigenomte</td>
</tr>
<tr>
<td>C</td>
<td>Center</td>
<td>6.49%</td>
<td>6.11%</td>
<td>889</td>
<td>6742</td>
<td>1459</td>
<td>#gbgftw #klimat</td>
</tr>
<tr>
<td>FP</td>
<td>Liberals</td>
<td>9.91%</td>
<td>5.42%</td>
<td>624</td>
<td>7049</td>
<td>1400</td>
<td>#almedalen #kfgbg</td>
</tr>
<tr>
<td>KD</td>
<td>Christian Democrats</td>
<td>5.93%</td>
<td>4.57%</td>
<td>373</td>
<td>29</td>
<td>601</td>
<td>#aliansen #svpola</td>
</tr>
<tr>
<td>S</td>
<td>Socials</td>
<td>24.19%</td>
<td>31.01%</td>
<td>839</td>
<td>20811</td>
<td>3853</td>
<td>#mpkongress #redof #miljo</td>
</tr>
<tr>
<td>V</td>
<td>Left</td>
<td>6.3%</td>
<td>5.72%</td>
<td>267</td>
<td>9799</td>
<td>1425</td>
<td>#08pol #intetillsalu #kollektivavtal</td>
</tr>
<tr>
<td>MP</td>
<td>Greens</td>
<td>15.41%</td>
<td>6.89%</td>
<td>915</td>
<td>3551</td>
<td>697</td>
<td>#sd</td>
</tr>
<tr>
<td>SD</td>
<td>Sweden Democrats</td>
<td>9.07%</td>
<td>12.36%</td>
<td>203</td>
<td>32585</td>
<td>4951</td>
<td>#feminism #ukraine #taplats</td>
</tr>
<tr>
<td>FI</td>
<td>Feminists</td>
<td>5.49%</td>
<td>5.12%</td>
<td>2503</td>
<td>6953</td>
<td>1215</td>
<td>#sd #val2014 dinrosta</td>
</tr>
<tr>
<td>P</td>
<td>Pirates</td>
<td>2.23%</td>
<td>0.43%</td>
<td>450</td>
<td>1459</td>
<td>2459</td>
<td>#jobbvalet #svrigenomte</td>
</tr>
</tbody>
</table>

Table I

**Swedish Political Parties:** Table summarizes the party names, their respective acronyms, count of posts, count of mentions, count of users, and most frequently used hashtags.

---

**Fig. 2.** Labeled visualization of the party interactions during European parliamentary time-line. Visualization is Force-directed and uses Kamada-Kawai layout. Nodes have been labeled with respect to their parties.

prefer to tag the alliance instead of the individual parties. Figure 2 shows the force directed visualization of the graph. Being a multi-partisan political scene, nodes have been colored to reflect on this fact. Such community formation depicts how users associated with a political party communicate with each other. Each node in the graph represents a political user and each edge indicates that these users mentioned each other at least in one of their tweets. The graph is result of the analysis result of 20,000 nodes and containing almost 40,000 edges. We filtered out only discussion involving the politicians. To produce such dense population we put a threshold on number of tweets selecting only very active users. Each party comprises official accounts of each parties and their respective members.

The community formation depicts cohesion in interaction between members of each parties. This is depicted by how dense nodes are clustered on the graph. As observed, Folk Party (FP), Center Party (CP), Sweden Democrats (SD) and in turn Feminists (F) show high cohesion by keeping their words within their own inner circle. Visualization also shows how disperse each party’s interaction has been through engaging in discussion with members of other parties. This is depicted by how much scattered nodes of the parties are with respect to other party nodes. For instance we can observe high dispersion in discussion between Moderates (M) and Social Democrats (S) and their neighboring parties.

While the visualization reveals a lot of interesting facts about the nodes of the graph, we won’t know much about the formation of edges of network. Figure 5 shows cumulative distribution for total degrees of edges in European election network. Plots show these distributions for the interactions taking place during the month leading to European election month and also afterwards. As seen, graph seems be proper fit for a power law distribution. We compare this degree distribution to degree distribution of general election in next section.

**C. Swedish General Election Time-Line**

In this section we focus on the time-line during mainly the general elections. General election time-line corresponds to the September month. According to data gathered, volume of discussion during general election peaks during the weeks leading to election and will continue for couple of weeks after the elections. The graph of general elections contains over 50,000 nodes and 440,000 edges. This graph contains the discussion between public and politicians, explaining such
larger data. This means we have to make up for the impact of having larger number of tweets. Visualization of graph in figure 4 reveals overall conversations same community shape as of compared to IV-B. Visualization is not labeled due to size of graph, although source and destination of edges are colored differently to reveal source and destination of interactions. Even though the graph is not labeled, almost the same community formation is revealed.

Figure 5 shows cumulative distribution for total degrees of edges in general election network.

Plot shows these distributions for the interactions taking place on the month of general election month. Edge distributions for general elections also seem to be a proper fit for power law distribution. Visually speaking, degree distributions for latter network seems to be a better representation of a scale-free network [25], as one might expects in a social network setting.

D. Link Prediction with Electoral Networks

Following the introductory analysis of characteristics of networks at hand, we explain how we have used link prediction to analyze the evolution of the networks at hand to measure a model of popularity for the parties. Given that a bipartite graph can be extracted from networks introduced, one could estimate a notion of popularity based upon the density of interactions they receive or provide. What makes link prediction of interest are two folds: primarily, link structure semantics could explain the density of conversations about parties and their respective members, which in turn should reflect on their popularity within the network at hand. Secondly, dynamics of interactions with the accounts of politicians, reflects on any shift in popularity that a stochastic approach link mining approach can capture.

1) Stochastic Link-Structure Mining: We leverage an invariant of SALSA (Stochastic Link-Structure Analysis)[26] for estimating the scores of nodes in link prediction. similar to HITS [27], SALSA uses the random walks to constructs a bipartite graph from a collection of nodes, which eventually converge onto two sets of nodes, collectively referred to as Hubs and Authorities. Random walks are bi-directional due to bi-directional edges of graph which also guarantee convergence of the algorithm. This guarantees that we always devise a bipartite graph from the source edge list. Each walk leverages a set of Markov chains, and each chain is dependent on the stochastic properties of random walks performed on the graph.

2) Personalized PageRank: Having the bipartite graph at hand, we would like to rank the nodes. To address the problem of ranking the vertices in a bipartite graph, we have leveraged an estimation of Personalized PageRank (PPR). One of the most fundamental graph computation problem, personalized PageRank (PPR) has proven to be effective in link prediction and friend suggestions in on-line social networks [28], [29]. PageRank is computed by executing a set of random walks that at each iteration, with a probability $P$, picks a random node and with probability $1 - P$ follows a random outgoing link from the node at hand. Personalized PageRank is the same as PageRank [30], although all the random picks of nodes eventually end at the source node, for which we are individualizing the PageRank for. Given the graph $G(V,E)$, personalized PageRank of node $v$, from the source node $u$, referred to as $\pi_u(v)$ is computed as follows [31]:

$$\pi_u(v) = \epsilon \delta_u(v) + (1 - \epsilon) \sum_{w \leq (w,v) \in E} \pi_u(w) \alpha_{u,v}$$

Leveraging Personalized PageRank in turn generates a collection of egocentric (personalized) random walks. Egocentricity of the walks are implemented by not visiting adjacent nodes during each active walk. One might consider visiting adjacent nodes if interested in modeling opinion influence [32].
3) Scaling Massive Random Walks: With increasing size of graphs, a single random walk is not enough to satisfy time and space for computation in social networks. For overall computations, we chose GraphChi [33] that implements a resource efficient framework for large scale graph processing. For simulating the random walks we worked with DrunkardMob [34] component of GraphChi, which allows parallelization of large random walks to be executed. We chose this component to address increasingly important issue of executing large quantities of random walks in a resource efficient manner. To summarize, each iteration of our link analysis approach repeat the following steps:

1) Simulate an egocentric random walk and pick $f$ most frequently visited vertices.
2) Build a bipartite graph by placing $f$ vertices on the Hub matrix, and a subset of their respective followers on the Authority matrix.
3) Execute SALSA on the graph and take the top $k$-scored vertices on the Authority matrix as the Top-followers of the node.

Such as most applications, we work with the top-$k$ values (and their corresponding nodes) in each PageRank vector, given the suitable $k$ of course.

E. Popularity Estimates: Parties, Members and Coalitions

We experimented with proposed link analysis approach using our two sets of graphs. We took in around 50000 source nodes, in case of European election network, to 100000 source nodes for general election network in two separate experiments. For each iteration of bi-directional SALSA execution, we executed at least 50 million walks. To analyze the results, we devised an analysis focused only on party accounts, and another analysis focused on a selection of member (individual) accounts. The identifiers were chosen from a popular portal that lists Swedish political presence on Twitter. This decision is due the fact that a political party has its own account, while politicians often have separate personal accounts. The accounts chosen for the study have been taken from this listing for the sake of validation.

Figure 6 respectively show estimated size of top-k (out-degree) scores for each parties according to personalized PageRank values. In each experiment, we took the identifier of the source node of the egocentric walk and estimated the personalized follower score accordingly. Figure 6 show the results of computation for official party accounts on Twitter. Compared to election results the popularity rankings of the parties for European elections match the outcome to a large extent, as seen Social Democrats and their respective coalitions stand above the Moderates and their coalitions. This is even more visible for Swedish Democrats as the European elections reflected on their rise to popularity. By far this is the best popularity ranking estimated. This is while comparatively looking at the popularity standings of general elections, we see almost the similar results which are comparable to outcome results by small difference. The difference between scores are interesting to observe.

Figure 7 visualises the comparison of dynamics of group rankings. The aim is to see if group-wise we can estimate the right standing for two popular coalitions (Allianz and Red-Greens), followed by third runner-up Sweden Democrats. In the comparison, both European and general election aggregates seem to reflect on the outcome of rankings by significantly distinguishing between the place of all three groups.

We computed a median of scores for each group of individual members. Since member scores are combined in the previous figure, we have plotted the individual standings in figure 8. While Moderates and Social Democrats comprise the largest parties, small gains in the popularity estimates for Moderates compared to very large gain in popularity estimates for Social Democrats reflect the standings of first and second places in both elections. Even more interesting is large gain

---

in popularity of leader of the third party Social Democrats that is visible in the plot. The result of general election is similar to general election vote outcome, which makes a case for studying interactions with member accounts, not only party accounts. Given the observed relationship between results of measure popularity rankings within the next section we study the statistical correlation between our results and the official outcomes corresponding to respective electoral time-lines.

F. Correlating Network Dynamism to Vote Outcomes

We observed that results presented have revealed interesting commonalities between link-structure driven popularity standings for member and party accounts on Twitter. But the question remain how strong these commonalities are.

To measure any possible dependence between two results, we take the popularity standings as one statistical variable and vote outcome as another variable. We devised two sets of tests to see if we can evaluate both unranked and most favorably, ranked dependence. Ranked correlation is most favorable as we are dealing election standing list. First, we used Pearson’s coefficient (denoted as Pearson’s $r$), which is the covariance of the variables divided by product of their standard deviations, estimated as follows:

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2\Sigma(y_i - \bar{y})^2}}$$

Where $x_i$ and $y_i$ are the statistical variables. Pearson is more effective in fitting a linear dependence between values. Most favorable for the ranked correlation, Spearman’s coefficient (denoted as Spearman’s $\rho$) is a more general metric of statistical correlation between two variables. Spearman’s coefficient can show how well the dependence between two variables can be explained using a monotonic function, and is estimated as follows:

$$\rho = 1 - \frac{6\Sigma d_i^2}{n(n^2 - 1)}$$

Where $n$ is the size of population and $d_i$ is the difference between the ranks. Table II show how positive the results of the statistical correlation between our results and the official outcomes are for most of the comparisons.

Pearson coefficients show that there is a strong linear correlation between the popularity rankings of both party and member accounts for European election time-line. This is while linear dependence for general election time-line is not as high as general election. What is interesting for us is the result of Spearman ranks, which show quite strong dependence for both European and general election time-lines. The most interesting result is how the choice of ranking of member accounts shows strong dependence with outcome of general election, while the choice of ranking of party accounts shows strong dependence with outcome of European elections.

To justify the latter results, we explain possible behavioral traits that might contribute to these observations. First of all, the density of interactions with party accounts during European parliamentary time-line as we saw in figure 1 is rather high. This is while authority reported low voter turn over (reported 51,07 % of eligible population)$^4$, which would explain how a proportion of voters followed the election event on-line, instead of going to ballots. Comparatively, the gain in increased interaction with member accounts during the general elections could explain increased participation from voters (reported 85,81% of eligible population)$^5$. Second, result shows how activity of member accounts could have significant impact on explaining result of elections, as compared to solely focusing on the party accounts. This is also verifiable by the fact that the number of Swedish politicians using Twitter, are increasing thus attracting public attention and interaction.

V. CONCLUSION AND FUTURE WORK

This paper aimed at adding to further evidence that Twitter data can actually be used to explain and perhaps predict the outcome of the elections. We studied a link mining approach that leverages the structural features of the interaction network underlying the conversation with politicians during the timeline of two elections. We presented evidence of how our approach reveals an authoritative structural link formation within which the popularity of the political accounts along with their neighbourhoods, shows strong correlation with the vote outcome. The public time-lines of two electoral events from 2014 elections of Sweden on Twitter were studied. By distinguishing between individual and official party accounts, we report that estimated popularities reveal strong statistical similarities with vote outcomes. We also revealed strong ranked dependence between standings of selected politicians and general election outcome, along with official party accounts and European election outcome. For future work, first we want to analyze other elections using link prediction to see if we can generalize this concept. We will focus on textual content of tweets using sentiment analysis and topic modelling.


We are also planning to study the diversification and language evolution in the linguistic features of political tweets [35].

VI. ACKNOWLEDGEMENT

This is to acknowledge that the work was carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship.

REFERENCES


### TABLE II

<table>
<thead>
<tr>
<th>Member accounts</th>
<th>Party accounts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Timeline</strong></td>
<td><strong>Pearson’s r</strong></td>
</tr>
<tr>
<td>European Election</td>
<td>0.792188828</td>
</tr>
<tr>
<td>General Election</td>
<td>0.649179744</td>
</tr>
</tbody>
</table>