# Exploring the usefulness of Twitter data for political analysis in Switzerland

## PRELIMINARY DRAFT!

Christian Mueller* Bruno Wueest† Thomas Willi†

*Paper prepared for the Annual Conference of the Swiss Political Science Association at the University of Basel, January 21-22, 2016.*

**Abstract**

Social media in general, and the Twitter micro-blogging service in particular, have recently received a lot of attention by social science scholars and political observers as potentially promising new data sources. The validity of such data, however, remains heavily disputed. One of the most serious problems that potential users of Twitter data have to face is selection bias. We propose an at least partial solution to this issue for Switzerland by compiling an original data set of all Twitter accounts affiliated with a Swiss political party. Specifically, we first use chain-referral sampling to expand an initial set of high-profile politicians and party users to a larger set of potentially relevant accounts. Then, we use text mining to filter for accounts with self-declared party affiliations. Using this method, we are able to systematically compare our sample of Twitter users against the broader population. We find significant bias with respect to ideology, gender, location, and language. Quantifying the selection bias in our sample enables potential future users of the data to mitigate some of its negative consequences. To illustrate this, we provide two proof-of-concept case studies. First, we use network analysis to exploit the relations embedded in Twitter data. This allows us to gain insights about central actors and drivers of the relationships of the communication networks. Then, we focus on political communication by using topic models to show who talks about what on Twitter and how this relates communication. Specifically, we use topic models to identify salient issues in public opinion.

**Keywords**

*Social media, Twitter, representativeness, Switzerland, topic models, network analysis*

*MSc Student, Department of Methodology, London School of Economics and Political Science. Email: chmue-ac@posteo.ch.

†Department of Political Science, University of Zurich, Switzerland. Emails: wueest@ipz.uzh.ch, thomas.willi@uzh.ch

# 1 Introduction

Twitter is a micro-blogging service with a growing user community since 2006. With some delay, the site also established a considerable user base in Switzerland. The growth of its usage motivates a closer look at how this particular social media platform plays out in terms of Swiss party politics. Moreover, not only regular citizens, but also the political elite can be expected to increasingly rely on micro-blogging. As Wallsten (2010) shows for election campaigns in the USA, bloggers and political campaigners occupy an influential position in shaping the political agenda. It does thus not come with a surprise that in Switzerland about 40% of all members of the Federal Assembly and three Federal Councillors have a Twitter account. From time to time, the Tweets they publish also reach the broader public, as it was the case with Jacqueline Badran, a representative of the Social Democratic Party in the National Council, in March 2013. After insulting a doorman on Twitter for expulsing her from a club where she illicitly lighted a cigarette, an outcry erupted in the mainstream media, which even led to demands for her resignation.

The goal of this study is twofold. The first aim is to provide systematic data on the presence and activity of Swiss parties on Twitter. There is much talk about the merits and perils of social media for politics and society. However, despite the mostly public nature of Twitter accounts and the scalability of data scraping on Twitter, it is surprisingly hard to collect data for specific purposes – such as the communication of Swiss parties as in this case – in a reliable way. This study aims to reveal the bias intrinsic to the political communication by Swiss parties on Twitter compared to offline politics. Thus, the first part of this contribution will be concerned with the comparison of the Twitter data with data on conventional politics as well as the socio-structure in Switzerland. The second part is concerned with a pilot study on how Swiss Twitter data can be used to model the public opinion and network structure during the last national election campaign in 2015.

# 2 Twitter, Politics, and the Swiss case

As "the Internet dramatically changed the communication landscape with the introduction of myriad new channels" (McCombs, 2005), many social scientists started out to explore how these new channels of political communication relate to traditional politics (Wallsten, 2010; Barberá, 2015). Micro-blogging sites like Twitter are but one of these myriad new channels on the Internet. They allow users to rapidly exchange content snippets and thus create a 'buzz' of otherwise isolated messages (Kaplan and Michael, 2011). Twitter is but one micro-blogging service, however, on a global scale, it by far ranks first with 1.3 bio. accounts and an estimated 316 mio. active users (Twitter, 2015). Also in Switzerland, Twitter is by far the most popular micro-blogging service. In 2015, already 10,7% of the total adult population in Switzerland used Twitter (NET-Metrix, 2015). For several other reasons, Switzerland is a valuable case to explore the feasibility of Twitter for studies on political communication and public opinion. First, due to the distinctive consensus-oriented character of its political system (most of all its highly federal and direct democratic institutional design (see Lijphart, 1999), Switzerland has very low access barriers to the public debate. Compared to most other countries, a larger variety of actors is thus able to engage in political communication (Höglinger, 2008). If there is a systematic relationship between Twitter and traditional politics, Switzerland therefore is a most likely case to study it more in detail. At the same time, the size of the Twitter sphere in Switzerland is comparatively moderate. Thus, it is much easier to get a comprehensive and comparable sample of Twitter accounts than in larger countries where the usership easily exceeds several millions. Finally, as in most other countries, social media are an upcoming phenomenon in the media landscape in Switzerland, which requires a closer look also from a political science perspective. Thus, a contextualization of political communication on Twitter has also practical merits for political observers and policy makers in Switzerland.

As for other social media services, a heightened debate has broken out about the merits and perils of engaging in analyses using such data in the aftermath of the 2008 presidential election in the USA, where it first was used as a campaigning tool on a massive scale. A first strand of literature gives reason to believe in the validity of Twitter analyses. Applying descriptive statistics, Tumasjan et al. (2010) found that the simple number of messages mentioning German political parties mirrored the result of the 2009 German general election surprisingly well. They maintain that the results even challenge the accuracy of traditional election polls. Similarly optimistic evaluations can be found in studies applying different methodological approaches to predict election results in the United Kingdom Lampos (2012), the Netherlands Sang and Bos (2012), Singapore Skoric et al. (2012), and the United States (Barberá, 2015). Further, O'Connor et al. (2010) showed that a simple sentiment analysis of Tweets reflects Obama job approvals during the first two year of his presidency quite accurately. Subsequently, Curini, Iacus and Porro (2014) showed that the measurement of public opinion via Twitter is feasible also in the context of Western European politics. The proponents of using Twitter as a data source stress the widespread usage of this micro-blogging service as well as the amount and richness of data available. "One distinct characteristic of this online social network is the presence of not only ordinary citizens, but also public officials, political parties, and candidates" (Barberá, 2015), which can potentially be used to infer population attitudes and the nature of political campaigns. This potential seems very promising, since Twitter allows unrestricted downloads to millions of public accounts and their messages in real-time. Furthermore, the scope for the application of methods deriving public opinion and political campaigning from Twitter data is large, since the micro-blogging platform has grown into a substantial player in the media market of most countries around the world.

In reaction to these optimistic studies, a second strand of literature advices caution against too euphoric assessments of Twitter's potential for answering social scientific research questions. Gayo-Avello (2012), to begin with, highlights his concerns with disappointing evidence from several US Senate races, which shows that electoral predictions from Twitter data using similar methods as the studies just mentioned are failing to perform better than chance (see also Barberá, 2015). Other critics maintain that at least some studies optimized the research designs in their favor. For example, if Tumasjan et al. (2010) had not restricted the sample of parties to the ones with national parliamentary representation, they would have predicted the Pirate Party to win the 2009 German general election, since it was the party with the highest number of mentions on Twitter (Jungherr, Jürgens and Schoen, 2012). Besides methodological shortcomings, these critiques prominently name two sources of selection bias that inflate the results. First, personal characteristics are likely to hamper the representativeness of Twitter data: "The average internet user is younger, more interested in politics, and comes from a higher socioeconomic background than the average citizen, which raises concerns about external validity" (Barberá, 2015). In addition, not only socio-structural differences between Twitter users and the basic population should be expected, but also ideological ones (Krueger, 2006). In line with its roots in the social movements of the 20th century, the political left in Western Europe and the United States has a preference for participatory forms of decision making (Kriesi et al., 2012). Thus, parties from the political left can be expected to have integrated interactive forms of Internet usage like micro-blogging much easier into their action repertoire. Second, the interdependence among the units of analysis (be it the Twitter accounts or the messages sent), which is sometimes exploited to generate data, is a further source of bias. Social media networks are assumed to be highly homophilic, i.e. the relationships among users are systematically clustered along ideological and/or socio-structural lines (Barberá, 2015; Lawrence, Sides and Farrell, 2010; McPherson, Smith-Lovin and Cook, 2001). Users have a higher propensity to follow other users who share their own beliefs. At the same time, they expose themselves less likely to opposing views (Barberá, 2015; Sunstein, 2001). In the end, this may lead to a non-representative distribution of the observed variables. And third, there are significant differences in the activity of Twitter users. While there is a small group of highly active users, many others have very low activity ratings in terms of the pace at which they (re-)tweet, follow other users, or favor Tweets.

Scholars who maintain that Twitter is a valid data source do not neglect these potential sources of selection bias. So far, they tried to enhance their prediction precision by externally validating their results with traditional political science data, by applying sophisticated machine learning algorithms (Pennacchiotti and Popescu, 2011), or by estimating individual traits which potentially cause selection bias such as gender and ethnicity from within the Twitter data (Barberá, 2015). This contribution, in contrast, suggests an actor based approach to systematically trace political communication in the abundance of communicative acts. More precisely, we maintain that a deliberate selection of Twitter accounts according to their self-declared affiliation to a Swiss party helps to establish a user network, which by definition represents party politics on Twitter. Thus, selection bias can be measured and subsequently amended in a reliable and transparent way. This approach is similar to the one suggested by Jungherr, Jürgens and Schoen (2012), who claim the only way to achieve an accurate prediction from Twitter data is to correctly identify likely voters and compiling an unbiased sample of such users.

In the following, we will first present our approach to identify a sample of politically relevant Swiss Twitter accounts, before we move on to the description of the bias in this sample. Lastly, we present the results of two very tentative predictions for the election campaign in 2011, a working example which allows us to evaluate the explanatory power of our Twitter data in the run up of a larger prediction of this year's federal election.

# 3 Retrieving party network data from Twitter

The first step in the data gathering process is the definition of which accounts need to be included into the analysis of Swiss party politics on Twitter. In other words, this step relates to the identification of the network boundaries. While the conception seems straightforward – theoretically, we simply would need to include the Twitter accounts of every Swiss party member into the analysis – , the empirical realization is cumbersome. The myriad of accounts and their highly unstructured descriptions actually make the definition of the network boundaries one of the most difficult challenges of this contribution. Moreover, there are no public registers of Swiss party members, which would have allowed a systematic search for Twitter accounts.

A first decision of which Twitter accounts match the definition of 'Swiss party politics' was to start with a position-based approach (see Marin and Wellman, 2011). More precisely, an initial core set of Twitter accounts was compiled by hand according to the importance of these accounts for the political system in Switzerland. This list contains members of both chambers of the national parliament (National Council and Council of States), Federal Councillors as well as the official national accounts of the eight most important parties in Switzerland.[1] Two accounts, which were not public, needed to be dropped from the analysis.[2] The final core set consists of 156 Twitter accounts (see Table 7 in the Appendix): 3 Federal Councillors, 8 party accounts, and 145 National or Councillors of State. This list is far from representative with regard to the political landscape, for example only three out of seven Federal Councillors, Alain Berset,

[1]Bürgerlich-Demokratische Partei (Conservative Democratic Party, BDP), Christlich-Demokratische Volkspartei (Christian Democratic Party, CVP), Freisinning-Demokratische Partei (The Liberals, FDP), die Grünen (Greens, GPS), Grünliberale Partei (Green-Liberal Party, GLP), Schweizerische Volkpartei (Swiss People's Party, SVP), and Sozialdemokratische Partei (Social Democratic Party, SP). During the data gathering process, additional fringe parties from the left or right pole of the ideological spectrum were coded into the category Others.

[2]By default, Twitter accounts are public. However, it is possible to block the visibility, which is why no data can be collected from the two accounts.

Johann Schneider-Amman and Simonetta Sommaruga, have a public Twitter account. However, exactly this self selection, i.e. which party affiliated people actually are participating in the political Twitter sphere, is the main method of identification also in the following network extension rounds.

The definition of the core set according to the position of accounts in the political system of Switzerland ensures that the following semi-automatic extension of the Swiss party politics network starts from the most central accounts. This means that from this core set, a chain-referral extension based on the relations of the core users was pursued. More precisely, in three extension rounds using the Stream API to tap Twitter data on a large scale, the friends and followers in the core set were collected and then coded according to their relevance. Table 1 shows an overview of the four extension rounds performed. The number of unique users surges to 61'308 in the first, 150'808 in the second, 236'251 in the third and decreases to 229'297 in the fourth extension round.

Table 1: Establishing the Swiss political party network: key figures of extension rounds.

| List | Year | N accounts | |
|---|---|---|---|
| | | New candidates | Relevant |
| Core set | | 155 | 155 |
| First extension | 2012 | 61'308 | 776 |
| Second extension | 2013 | 150'808 | 46 |
| Third extension | 2014 | 236'251 | 676 |
| Fourth extension | 2015 | 229'297 | 544 |
| Total | | 2197 | |

A friend and follower of an account already identified as relevant needs to fulfill two requirements to be included in the list of relevant accounts. First, the account is relevant if it is located in Switzerland. As a matter of fact, a large part of users actually indicated their location in the predefined entry field. With the help of the geocoding services by *Google Maps*, *Bing Maps*, and *MapQuest*, which all allow for keyword searches, the locations were systematized. This means that all instances of the same location were attributed to one single pair of geographic coordinates. For example, Zurich was indicated as 'Zürich', 'Zuerich', 'Zurich', 'ZRH' and had to be standardized to 'Zurich' before the geocodes could be retrieved.[3] For the remaining accounts, the place of residence was added as location.

Secondly, a simple keyword match with the items indicated in Table 9 in the appendix was applied to the descriptions and names of the Swiss Twitter accounts. The keyword list contains all names, abbreviations, and paraphrases of the parties as well as all official employment titles of Swiss politicians in the three official languages of Switzerland (Italian, French, and German). In addition, a general search for indications of political matters was performed in order to avoid false negatives. Since this matching filtered the number of accounts already to a considerable degree, no further automatization steps such as machine learning classifications were necessary. Consequentially, all keyword hits were manually checked for their relevance. This final step led to the inclusion of 2042 additional accounts into the sample. Besides location and party affiliation, gender and community size ($<$ 10'000, 10'000-30'000 and $>$ 30'000 inhabitants) was annotated.

As a word of caution, both the information on the party affiliation as well as the account location rely on the self declaration of the users, which of course raises the problem that the users are not representative to a larger target population, which, for example, could be defined as all Swiss citizens with a right to vote. However, in the first step, a sample reflecting as many users affiliated to Swiss parties as possible is envisaged. In the end of the day, if the source is known, such selection bias can be amended. Moreover, data collected from Twitter accounts of party members who are not indicating their political affiliation could even be worse, since these users might use their account for private and business purposes.

Moreover, it is of course the case that by far not all Twitter users indicate the location and description of their accounts. Thus, relying on the self declaration of Twitter users yields missing data. However, it is the only method to systematically extend the network without relaxing the identifying definitions for the network extension. There is no other information provided for the Twit-

[3]The efficacy of this process could be considerably enhanced by using Regular Expression Syntax and string matching via the restricted Damerau-Levenshtein distance. In addition, all non-alphanumeric characters were removed before the matching.

ter accounts besides the self declarations in the location, name, and description entry fields; and relying on searches in external data such as the official parliamentary services or Wikipedia simply is not feasible for such a large number of accounts.

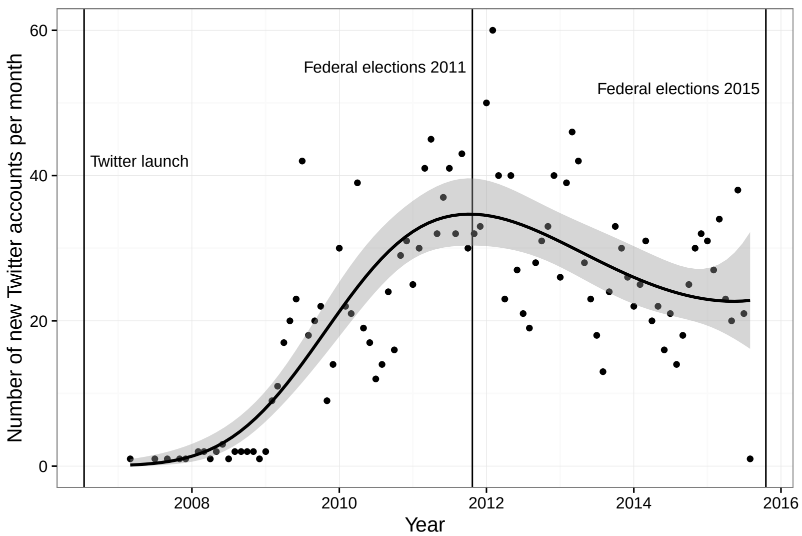
While the sample of Twitter users includes the most important party accounts in the core list which was manually compiled, the semi-automated extension also represents a systematic extension thereof. Nevertheless, it cannot be guaranteed that the second extension round already reached the network boundaries. However, the fact that the number of unique Twitter users which were newly found in the fourth extension round is decreasing compared to the third round, might be signalling that the network borders of politically affiliated Swiss Twitter users have almost been reached. Moreover, since the total amount of users in Switzerland can be estimated to about 677'000, already most of all Swiss accounts were at least once included into the chain-referral sampling.

# 4 Exploring the Swiss 'Tweetocracy'

## 4.1 Dynamic

As Figure 1 shows, the first Twitter account by a Swiss partisan actor – namely by Stefano Araujo from the Partito Comunista Ticinese (the Ticino offshoot of the Swiss Party of Labour, a small radical left party) – was created in February 2007, seven months after the introduction of the micro-blogging service. Further, the smoothed trend line shows a clear pattern. The growth of the Swiss party network on Twitter is moderate until Spring 2010. Thereafter, the pace substantially accelerates until the end of 2011. Hence, before 2010, Twitter seems to be a communication tool only at the fringes of political campaigns. As a matter of fact, however, the peak in late 2011 seems to be clearly related to the federal election in November 2011. Many politicians thus appear to have set up an account as an attempt to mobilize additional support during the election campaign. In sum, the widespread introduction of micro-blogging as an important campaigning tool in Switzerland thus lagged three years compared to the United States, where commentators see the win of Barack Obama in 2008 as being the first 'social media election' (Wallsten, 2010). The fact that the number of new accounts does level off towards 2015 possibly point to the effects of the electoral cicle in Switzerland. Many new accounts might have been set up in prospect to last year's national election.

Figure 1: Diachronic development of Twitter account setups: absolute numbers of setups and spline smoothed trend line.




## 4.2 Representativeness

The discussion of the results continues with an overview of how the 2197 party affiliated accounts compare to general population data. Table 2 shows the distribution of the number of Swiss parties' Twitter accounts in terms of gender, language and community type. Gender is major factor biasing communication on Twitter as the exceptionally high Least Squares index value of 24.5 indicates.[4] Hence, there are about three times as much party affiliated men than women on Twitter in Switzerland. Further, the geolocations of the Twitter accounts are also annotated according to three different community sizes. It becomes evident that the urban centers are heavily over-represented in the Twitter sphere. As for people living in small communities, by contrast, they are considerably under-represented. Quite intuitively, Twitter therefore does seem to be a preferred communication platform of urban, politically engaged individuals. The bias in favor of large communities results in an LSI of 21.8.

[4]Calculated as the the square root of half the sum of the squares of the difference between the quantity to be compared with and the quantity to compare, see notes in Table 2.

The distribution of the Twitter accounts by language (see the bottom rows in Table 2), finally, holds no surprise, since the differences between the socio-demographic and Twitter data are not large. Italian speaking Swiss citizens as well as the German speaking majority are slightly less well represented in Swiss party politics on Twitter. The French speaking minority and residents indicating English as their main language, by contrast, have a higher share of Twitter accounts as its size in the population. These differences, however, are modest, which is why the LSI is lowest for the language differences (9.0).

Table 2: The representativeness of the Swiss 'Tweetocracy' in terms of locations, gender and languages: Column percentages.

| | Twitter | Population[a] |
|---|---|---|
| *Gender* | | |
| Women | 26.0 | 50.5 |
| Men | 74.0 | 49.5 |
| LSI[b] | 24.5 | |
| *Community size* | | |
| Large (> 30'000) | 42.5 | 21.8 |
| Mid-sized (10'000-30'000) | 26.3 | 24.2 |
| Small (< 10'000) | 31.2 | 54.0 |
| LSI | 21.8 | |
| *Language* | | |
| German | 58.8 | 63.7 |
| French | 28.5 | 20.4 |
| English | 9.0 | 1.0 |
| Italian | 3.6 | 6.5 |
| Spanish | 0.1 | 1.1 |
| LSI | 9.0 | |

*Notes:* [a] for locations and gender: All permanent residents in Switzerland in percentages, for languages: share of self-declared main language by all permanent residents in Switzerland over the age of 15 in percentages. Source: Federal Statistical Office.; [b] LSI = Least squares index: $\sqrt{\frac{1}{2}\sum(v_i - s_i)^2}$, where $v$ is the quantity to be compared with (usually vote shares) and $s$ the quantity whose representativeness should be measured (usually parliamentary mandates).

Table 3 shows the comparison of Twitter account information for the major Swiss parties and the electoral performance at the last general election. Basically, there are three major blocks in the Swiss party system: the left parties (SPS and GPS), center-right parties (GLP, FDP, CVP, and BDP) and the right-wing party SVP. In this regard, the numbers show clear evidence in favor of a left bias on Twitter. While in the last federal elections, the left parties achieved a cumulative vote share of 25.9 percent, they managed to accumulate 33.7 percent of all accounts. Even more impressive is the achievement in terms of followers. The left has almost garnered a majority of the follower shares (47.6 percent). The political movement which clearly is underrepresented on Twitter are the right-wing parties. Depending on the indicator, the Twitter presence of the largest party in Switzerland, the Swiss People's Party is 15.6 to 17.7 percent lower than its electoral support. This corresponds to their constituency, which on average is elder and probably uses social media less intensely.

Table 3: The representativeness of the Swiss 'Tweetocracy' in terms of partisanship: Column percentages.

| | Federal election 2015 | | Twitter | | |
|---|---|---|---|---|---|
| Party | Vote share | Mandates | Followers | Accounts | Activity |
| SVP | 29.4 | 32.5 | 11.7 | 13.8 | 12.8 |
| SPS | 18.8 | 21.5 | 33.3 | 24.0 | 19.7 |
| FDP | 16.4 | 16.5 | 13.9 | 16.9 | 10.6 |
| CVP | 11.6 | 13.5 | 12.7 | 13.8 | 14.5 |
| GPS | 7.1 | 5.5 | 14.3 | 10.4 | 14.0 |
| GLP | 4.6 | 3.5 | 6.8 | 8.7 | 9.8 |
| BDP | 4.1 | 3.5 | 2.3 | 4.9 | 3.2 |
| EVP | 1.9 | 1.0 | 2.0 | 2.8 | 11.0 |
| EDU | 1.2 | 0.0 | 0.0 | 0.0 | 0.0 |
| Others | 4.9 | 2.5 | 2.9 | 4.7 | 4.4 |
| LSI | – | 4.0 | 17.3 | 12.4 | 15.5 |

*Notes:* SVP = Swiss People's Party, SPS = Social Democratic Party, FDP = Liberal Democratic Party, CVP = Christian Democratic People's Party, GPS = Green Party, GLP = Green Liberal Party, BDP = Conservative Democratic Party, EVP = Evangelical People's Party, EDU = Federal Democratic Union. See Table 11 in the Appendix for the absolute numbers of followers, accounts, and activities.

# 5 Potential use cases

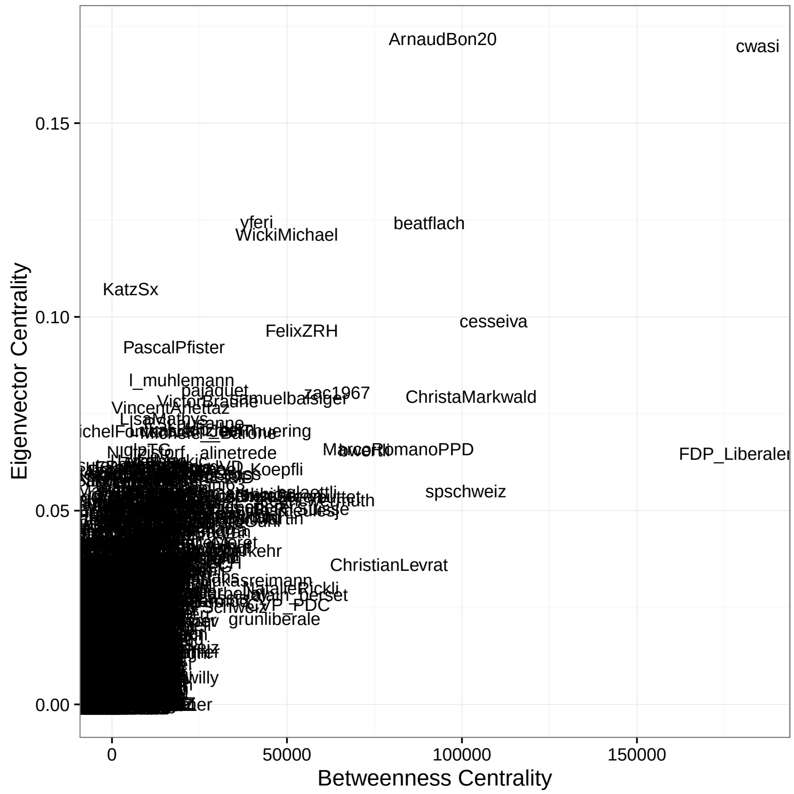
After discussing our sample's representativeness, the remainder of this paper will exemplify two potential use cases for our dataset. These illustrative anlyses rely on several aspects that make Twitter data unique.

## 5.1 Network-focused perspective

The diverse relationship information that Twitter makes available is one the most interesting types of data Twitter provides. This information is available on the level of a tweet, where users can quote or retweet other users' content and also mention them directly, as well as at the user level, where every account can subscribe to the status updates of other users ("following" in Twitter's terminology). The latter network data about Twitter users affiliated with a Swiss party will be examined in the remainder of this section. The network that is generated by the Twitter users in our sample can first be inspected graphically. Due to the size of the network, little can be gathered from inspecting the whole network. Therefore, we produce a specialized analysis to identify central actors in the network in figure 2. We use two common definitions of the centrality of an actor to measure centrality here: betweenness and eigenvector (Wasserman and Faust, 1997, 198–210). Betweenness centrality captures how likely it is that a given actor is on the shortest connection between two other actors in the network. Eigenvector centrality measures an actor's connection to other accounts, where connections to central actors are weighted higher than connections to peripheral ones.

To begin with the overall structure, it is striking that there are comparatively few accounts which are placed in the top right quadrant of the graph and thus have a high overall centrality. The majority of Twitter users with our Swiss party affiliation only follow or is friend with very few other account in the network. Accordingly, the network density, i.e. the number of actually present relations compared to all possible relationships, is only about 3%. The best connected Twitter user is Christian Wasserfallen, a young but very well-established National Councillor from the FDP. Overall, the FDP is the most highly connected party in our sample as most other accounts in the upper right part of the graph, e.g. Arnaud Bonvin ("aurnaudbon20") and Claudine Esseiva ("cesseiva") both work for the party's general secretariat, are affiliated with it.

Figure 2: Betweenness and eigenvector centrality in the Twitter network directed by friendships and followings.




Most other users with high scores on either centrality measure are members of the SPS, but there are also members of most other parties, including the GLP ("beatflach"), the Greens ("alinetrede"), and even a member of the SVP ("zac1967"). Besides the clearly identifiable ones of the FDP and the SPS, official party accounts seem to be stacked at the lower right, where it is possible to recongnise the GLP's and the CVP's accounts.

This graphical analysis reveals several things. First, that most of the best connected Twitter users are well-established, influential national politicians. Even though there are many lower-level politicians and party sympathizers in our sample, they only play a minor role when it comes to status within the network. It should be kept in mind here, that status in the sense that we use it here does not map directly to influence or anything similar. On the one hand, high centrality could be seen as the potential to reach many followers, which might or might not be realized. On the other hand, connecting with users at the fringe of the network will drive up an actors betweenness centrality but these users might not be otherwise well connected because the only use Twitter sparingly. Second, the presence of party functionaries who work as community managers from the FDP may be a hint of concerted effort by the party to link up their candidates. That might also be specifically related to the federal elections. Third, many of the very central actors are significantly younger than the average national politician. It might be that these politicians anticipate their audience to be younger and thus more receptive to campaigning on Twitter. But it is also likely that it is only age which drives that association because younger politician might simply behave in a similar way as younger people, who are on average more likely to use Twitter.

Taking the graphical analysis a step further, we use exponential random graph modelling (Lusher, Koskinen and Robins, 2012; Hunter et al., 2008) to explicitly model the likelihood of a connection between two Twitter users. We had to restrict the network to the 500 users with the highest number of tweets and favourites to make the analysis tractable. The results in table 4 were estimated by 35 iterations of MCMC-MLE where the normalizing constant was approximated by 50000 MCMC draws in each iteration. Our inspection of the MCMC chains evaluated at the maximum likelihood estimate indicates that they approximately converged. We subsequently validated the specification by comparing 250 simulated networks from our model against the observed network. Figures 3 and 4 in the appendix show that the simulations fit the degree distribution relatively well, but that there is a serious mismatch in the edgewise shared partners distribution. We are still working on a specification that adequately represents all aspects of the network and is not degenerate.

The results, which given the remarks in the preceding paragraph should be seen as transient, indicate that, in this model, party homophily is the most important factor. Specifically, the odds of forming a connection between two users are 6.2 times higher if these two users are from the same party compared to when they come from different parties. A possible explanation of this association is that, as discussed above, communication networks are often structured by shared values and beliefs and political views are most likely an important element of the belief system of self-declared politicians. Though the SPS is has

Table 4: Exponential random graph model.

| Network dynamics | Estimate | Std. error |
|---|---|---|
| *Party (reference: SPS)* | | |
| SVP | -0.06 | 0.02 |
| FDP | 0.05 | 0.01 |
| CVP | 0.25 | 0.01 |
| GPS | 0.25 | 0.01 |
| GLP | 0.20 | 0.02 |
| BDP | 0.27 | 0.02 |
| EVP | 0.23 | 0.03 |
| Others | 0.00 | 0.02 |
| *Language (reference: German)* | | |
| French | -0.15 | 0.01 |
| Italian | -0.61 | 0.03 |
| English | -0.08 | 0.01 |
| *Location (reference: large communities)* | | |
| Mid-sized | 0.11 | 0.01 |
| Small | 0.10 | 0.01 |
| Party homophily | 1.83 | 0.02 |
| log(Tweets) | -0.03 | 0.00 |
| log(Followers) | 0.26 | 0.01 |
| log(Friends) | 0.27 | 0.01 |
| Density | -24.33 | 0.32 |
| Edge-wise shared partners[a] | 5.48 | 0.13 |
| N users | 500 | |
| N relations | 34239 | |

Notes: [a] Geometrically weighted edge-wise shared partner distribution with decay parameter fixed at 1.

the highest share of accounts and tweets, changing one of the accounts in a possible connected pair to most other parties is associated with an increased likelihood for a connection between the two accounts. The two only exceptions to this rule are the SVP or radical and splinter groups that we subsumed in the "other" category. The negative sign of the coefficient for log(Tweets) is a bit suprising. Even though the effect size is quite small, also practically no effect is noteworthy because that means that higher numbers of status updates are associated with a very slightly lower likelihood that a connection between two users is formed.

The two analyses presented in this section are only meant to serve as starting points for more in-depth research that makes use of the rich network data Twitter provides. Further analysis, for example of sub-networks spanned by intra-party relationships, could reveal whether there is indeed evidence for different social media strategies, and if so, whether such strategies have proven to be successful. If success is defined as reaching as many potentially interested Twitter users as possible, high overall centrality as we defined and measured it might not be the goal after all.

The model-based analysis we present is equally simple and could be expanded in various ways. First, additional effects, for example separate homophily effects for each party or language group, could be added to the model. Then, additional terms capturing network dynamics could be introduced. While our model contains the number of friends, followers, and tweets as approximations for popularity and sociality effects, it does not contain terms for other network dynamics such as reciprocity or transitivity. Last, our specification does not make much use of the directed nature of the network. Many of the terms in the model could be further split by considering separate sender and receiver effects. Furthermore, the whole analysis would benefit a lot from cross-validating the findings with other network modelling approaches, as inference in this context is wholly model-dependent.

## 5.2 Thematic analysis

In a second preliminary analysis, we identify the substance of Twitter communication is by estimating a topic model (Roberts et al., 2014), which allows us to estimate probabilities for every Tweet how it relates to latent variables, called topics. More precisely, we use the Latent Dirichlet Allocation, a hierarchical mixed-membership model in which the document-topic and word-topic probabilities have a common prior drawn from a Dirichlet distribution (Blei, Ng and Jordan, 2003). A crucial decision in every application of a topic model pertains to the granularity, i.e. the number of topics. A topic model with too few topics will produce overly broad, diffuse topics, while a model with too many topics will results in many small, hardly distinguishable topics (Greene and Cross, 2015). An increasingly popular strategy to resolve this problem is to compare the coherence of different topic models. To this purpose, we use word2vec Mikolov and

Dean (2013), which learns and aggregates term similarities through a shallow neural network process. By comparing the coherence within and between the vectors of most probable words for each topic model, word2vec suggests a granularity of 14 for our corpus and a candidate range of 3 to 15 topics (see Figure 5).

In this analysis, we only cosider the 165'385 German Tweets available for the campaign period of four months before the election date. Previous to the estimation, all tweets are prepared using a set of preprocessing methods, including text segmentation into sentences, tokenizing, removal of punctuation, as well as stemming and converting all words to lowercase. After the estimations, the topics were attributed to political issues by interpreting the most closely associated words in terms of their high probability ranking (see Table 5, in which the 15 most relevant words are listed for each topic).[5] Note that the inclusion of prevalence and content covariates, i.e. document or word level indicators to model similarities among words and documents conditional on their topic membership, failed for this anaylsis, possibly due to the problems of the data as discussed below.

Most notably, only nine of the 14 estimated topics seem to be at least somewhat related to a substantive political issue. There is one topic each related to economics, environment and fugitives. Further, it seems that the traditional media's coverage of bilateral relations receives attention in the Tweets. Hence, we can find references to the most important issues of the electoral campaign 2015 on Twitter. Further, table 6 gives a tentative evaluation of two of these topics by indicating the correlations of the account holders' party affiliation, community size, the number of tweets and followers as well as a daily trend with the topic prevalences, i.e. the proportion of a topic in a Tweet. Intuitively, it makes sense that the right-wing parties (SVP and EDU) champion the topic centering on fugitives, while the liberal parties (FDP and GLP) emphasize the topic on economics. Nevertheless, in general, there seems to be not much expression of the public opinion in substantial terms. The remaining topics and thus the lion's share of communication on Twitter centers on either the election campaign itself, election contests, votes or no clear political concept such as the weather. Hence, at least for this preliminary analysis, the results cast some doubts that Twitter could be a meaningful complement or substitute to public

[5] The words are indicated as stems, which constitute the basis of the LDA data generation process.

Table 5: 15 most closely associated words for 14 topic model

| economics | environment | media and bilateral relations | media and votes | unclear | electoral contest | unclear |
|---|---|---|---|---|---|---|
| mehr | neu | svp | nein | via | polit | schon |
| wenig | sich | blickch | sagt | nzz | bern | gut |
| braucht | statt | medi | richtig | macht | zurich | jahr |
| wichtig | geht | isch | alt | frau | cvp | gibt |
| geld | energiew | selb | srfarena | tagesanzeig | kanton | link |
| stark | bleib | spiel | srfnews | warum | uns | wurd |
| staat | akw | bilateral | abst | min | partei | erst |
| grund | entscheid | petition | bleibt | kind | bdp | seit |
| gross | gleich | mull | sag | watsonnews | burg | recht |
| platz | bglaettli | asylchaos | abstimm | arbeit | stadt | lang |
| unternehm | unterschreib | nimmt | rtvg | weit | cvppdc | beim |
| wirtschaft | kampf | journalist | zwei | darf | losung | zeit |
| demokrati | ford | bund | antwort | falsch | evppev | mal |
| hoh | facebook | deutlich | darum | bess | sitz | bild |

| unclear | fugitives | unclear | unclear | mobilization | weather | electoral contest |
|---|---|---|---|---|---|---|
| wer | schweiz | war | frag | dank | heut | wahlch |
| bitt | fluchtling | mal | srf | tag | wett | wahl |
| weiss | mensch | einfach | klar | viel | morg | grun |
| eigent | europa | wohl | steht | toll | uhr | list |
| gern | muss | genau | lass | herzlich | grad | fdp |
| leut | land | halt | eig | unterstutz | kommt | grunliberal |
| twitt | leb | seh | person | freu | deutschland | jung |
| huberf | welt | komm | artikel | best | gewitt | nationalrat |
| imm | brauch | hatt | interessant | nach | htt | stimm |
| jemand | griechenland | leid | stell | lieb | gest | glp |
| wirklich | weg | ganz | off | spannend | teil | wahlkampf |
| zeitungsj | zahl | lieb | wort | wunsch | klein | fdpliberal |
| wieso | deutsch | find | gemacht | uns | aktuell | gruenech |
| nohillsid | red | dafur | heisst | woch | abend | spschweiz |

opinion polls in terms of the content of Tweets. This lack in content most likely is aggravated by serious problems with the data basis, which can be exemplified by the following two Tweets from our sample:

- *"Es geht los - Duume drugge!! Vamos Basilea carajo. #rotblaulive #fcbase l#BSLTOT"*
- *"@srfvirus seeehr guuuut"*

Table 6: OLS regressions on selected topic prevalences.

| | Fugitives | | | Economics | | |
|---|---|---|---|---|---|---|
| | Estimate | Std.Err. | Pr(> $\|t\|$) | Estimate | Std.Err. | Pr(> $\|t\|$) |
| (Intercept) | 0.075 | 0.001 | *** | 0.074 | 0.001 | *** |
| Party (ref. BDP) | | | | | | |
| CVP | 0.001 | 0.001 | | -0.004 | 0.001 | *** |
| EDU | 0.025 | 0.014 | . | 0.001 | 0.012 | |
| EVP | 0.002 | 0.001 | . | -0.003 | 0.001 | *** |
| FDP | 0.005 | 0.001 | *** | 0.005 | 0.001 | *** |
| GLP | 0.001 | 0.001 | | 0.006 | 0.001 | *** |
| GPS | 0.009 | 0.001 | *** | -0.004 | 0.001 | *** |
| Others | 0.018 | 0.001 | *** | 0.002 | 0.001 | |
| SPS | 0.012 | 0.001 | *** | 0.002 | 0.001 | * |
| SVP | 0.029 | 0.001 | *** | 0.000 | 0.001 | |
| Community size (ref. = small) | | | | | | |
| mid-sized | 0.003 | 0.000 | *** | 0.002 | 0.000 | *** |
| large | 0.002 | 0.001 | *** | 0.003 | 0.000 | *** |
| none | 0.003 | 0.001 | ** | -0.003 | 0.001 | *** |
| number of Tweets | 0.000 | 0.000 | *** | 0.000 | 0.000 | *** |
| number of followers | 0.000 | 0.000 | | 0.000 | 0.000 | *** |
| daily trend | 0.000 | 0.000 | *** | 0.000 | 0.000 | *** |
| N | | | 165'059 | | | 165'059 |
| Res. st. error: | | | 0.063 (df=165'043) | | | 0.079 (df=165'043) |
| Adj. R-squared: | | | 0.015 | | | 0.034 |
| F-stat.: | | | 174.6 *** (df=15/165'043) | | | 393.5 *** (df=15/165'043) |

Signif. codes: *** <= 0.001, ** <= 0.01, * <= 0.05

These are certainly two extreme cases, but references to dialect and foreign languages, word concatenations ('rotblaulive' actually are three words), typos (e.g. 'fcbase l' should have no whitespace in between) and emphasis through a repetitive entering of vocals (e.g. 'guuuut' instead of gut) are clearly occuring more often on Twitter. This makes the task for bag of words models such as LDA even more demanding than it is anyways.

# 6 Conclusion

This contribution put forward an exploration of the Swiss party network on Twitter. Starting from an actor based approach, a sample of Twitter users was established on the basis of their self-declared affiliation to Swiss party politics. More precisely, an initial set of pre-defined accounts could be extended to a comprehensive network of party affiliated Twitter users in Switzerland. On the bases of this sample of Swiss Twitter users, several aspects of the selection bias inherent to the Swiss 'Tweetocracy' could be shown. First, the analysis of the dynamic of communication over time has revealed that for the political left, the maturation of the Swiss party network on Twitter took place in 2011, so it lags well behind the general introduction of the micro-blogging service as well as its widespread application by parties in the United States.

Second, the political left, men and the urban areas are substantially over-represented on this micro-blogging platform. In turn, this also means that studies on the political right, women and/or the urban areas will likely underestimate their findings. Third, as for the network structure of the Swiss 'Tweetocracy', it was found that it is split into a large crowd of rather passive consumers of Twitter messages and a core of highly active users, to which party leaders and promising young politicians belong. This bias towards the political elite might pose severe problems for studies claiming to measure public opinion on Twitter. In sum, the presence of Swiss parties on Twitter is not even rudimentarily representative of traditional offline politics. Ignoring these sources of bias most likely leads to misguided conclusions and unsound predictions.

Our preliminary graphical analysis suggests that different parties might have different social media strategies. Some of them seem to use a top-down approach, where the general secretariat bundles and connects the party members' social network profiles, while others rely more on the initiative of individual politicians. The thematic analyses revealed that much of the communication by the party affiliated accounts lacks a clear relationship to political issues and even to politics in general. Tweets might be a useful data basis for some explorations of campaigning, but the non-structured language Twitter users frequently use poses serious challenges to large-scale quantitative analyses.

## References

Barberá, Pablo. 2015. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23(1):76–91.

Blei, D. M., A. Y. Ng and M. I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning* 3:993–1022.

Curini, L., S. M. Iacus and G. Porro. 2014. "Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France." *New Media & Society* 16(2):340–358.

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.

Greene, J. and D. Cross. 2015. *Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach.* Paper presented at the 5th annual conference of the European Political Science Association in Vienna, June 25-27, 2015.

Höglinger, D. 2008. "Verschafft die direkte Demokratie den Benachteiligten mehr Gehör? Der Einfluss institutioneller Rahmenbedingungen auf die mediale Präsenz politischer Akteure." *Swiss Political Science Review* 14(2):207–243.

Hunter, David R., Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, Morris and Martina. 2008. "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks." *Journal of Statistical Software* 24(3):1–29.

Jungherr, A., P. Jürgens and H. Schoen. 2012. "Why the pirate party won the german election of 2009 or the trouble with predictions: A response to 'predicting elections with twitter: What 140 characters reveal about political sentiment'." *Social Science Computer Review* 30(2):229–234.

Kaplan, A. M. and H. Michael. 2011. "The early bird catches the news: Nine things you should know about micro-blogging." *Business Horizons* 54(2):105–113.

Kriesi, H., E. Grande, M. Dolezal, M. Helbling, S. Hutter, D. Hoeglinger and B. Wueest. 2012. *Political Conflict in Western Europe.* Cambridge, MA: Cambridge University Press.

Krueger, B. 2006. "A Comparison of Conventional and Internet Political Mobilization." *American Politics Research* 34:759–776.

Lampos, V. 2012. *On voting intentions inference from Twitter content: a case study on the UK 2010 General Election.* Arxiv preprint arXiv:1204.0423.

Lawrence, E., J. Sides and H. Farrell. 2010. "Self-Segregation or Deliberation? Blog Readership, Participation, and Polarization in American Politics." *Perspectives on Politics* 8:141–157.

Lijphart, A. 1999. *Patterns of democracy: government forms and performance* *in thirty-six countries*. New Haven, CO: Yale University Press.

Lusher, Dean, Johan Koskinen and Garry Robins. 2012. *Exponential Random Graph Models for Social Networks*. New York, NY: Cambridge University Press.

Marin, A. and B. Wellman. 2011. Social Network Analysis: An Introduction. In *The SAGE Handbook of Social Network Analysis*, ed. J. Scott and P. J. Carrington. London, UK: Sage Publications pp. 11–25.

McCombs, M. 2005. "A Look at Agenda-setting: past, present and future." *Journalism Studies* 6(4):543–557.

McPherson, M., L. Smith-Lovin and J. Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual Review of Sociology* 27:415–444.

Mikolov, T., K. Chen G. Corrado and J. Dean. 2013. *Efficient Estimation of Word Representations in Vector Space*. CoRR, abs/1301.3781.

NET-Metrix. 2015. *Base 2015-01. Internetnutzungsverhalten der User in der Schweiz*. Zurich: .

O'Connor, B., R. Balasubramanyan, B. Routledge and N. Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. pp. 122–129.

Pennacchiotti, M. and A. Popescu. 2011. A machine learning approach to twitter user classification. In *Fifth International AAAI Conference on Weblogs and Social Media*. p. 281.

Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. Gadarian, B. Albertson and D. Rand. 2014. "Structural topic models for open-ended survey responses." *American Journal of Political Science* 58:1064–1082.

Sang, E. and J. Bos. 2012. Predicting the 2011 Dutch Senate Election Results with Twitter. In *Proceedings of SASN 2012, the EACL 2012 Workshop on Semantic Analysis in Social Networks*. Avignon, France: p. 53.

Skoric, M., N. Poor, P. Achananuparp, E. Lim and J. Jiang. 2012. Tweets and Votes: A Study of the 2011 Singapore General Election. In *45th Hawaii International Conference on System Sciences*. Maui, HI: pp. 2583–2591.

Sunstein, C. 2001. *Republic.com*. Princeton, NJ: Princeton University Press.

Tumasjan, A., T. Sprenger, P. Sandner and I. Welpe. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. pp. 178–185.

Twitter. 2015. *Twitter usage / company facts*. https://about.twitter.com/company, San Francisco, California.

Wallsten, K. 2010. "Yes We Can: How Online Viewership, Blog Discussion, Campaign Statements, and Mainstream Media Coverage Produced a Viral Video Phenomenon." *Journal of Information Technology & Politics* 7:163–181.

Wasserman, Stanley and Katherine Faust. 1997. *Social Network Analysis: Methods and Applications.* New York, NY: Cambridge University Press.

# Appendix

Table 7: Core set of Swiss political Twitter users.

| Name | Party affiliation | Name | Party Affiliation |
|---|---|---|---|
| Bernhard Guhl | BDP | Roberta Pantani | Others |
| BDP Schweiz | BDP | Mauro Poggia | Others |
| Martin Landolt | BDP | Ricardo Lumengo | SPS |
| Lorenz Hess | BDP | Cédric Wermuth | SPS |
| Rosmarie Quadranti | BDP | SP Schweiz | SPS |
| Schmid Federer | CVP | Christian Levrat | SPS |
| Marco Romano | CVP | Jacqueline Badran | SPS |
| Viola Amherd | CVP | Bea Heim | SPS |
| Kathy Riklin | CVP | Philipp Hadorn | SPS |
| Brigitte Haeberli | CVP | Roberto Zanetti | SPS |
| Jacques Neirynck | CVP | Carlo Sommaruga | SPS |
| Luc Barthassat | CVP | Jacqueline Fehr | SPS |
| CVP PDC PPD PCD | CVP | Jean-Francois Steiert | SPS |
| Pirmin Bischof | CVP | Alain Berset | SPS |
| Brigitte Häberli | CVP | Evi Allemann | SPS |
| Christophe Darbellay | CVP | Yvonne Feri | SPS |
| Elisabeth Schneider | CVP | Pascale Bruderer | SPS |
| Yannick Buttet | CVP | Jean Chr. Schwaab | SPS |
| Paul-Andre Roux | CVP | Edith Graf-Litscher | SPS |
| Jean-René Fournier | CVP | Didier Berberat | SPS |
| Christian Lohr | CVP | Matthias Aebischer | SPS |
| Alois Gmür | CVP | Silvia Schenker | SPS |
| Ida Glanzmann | CVP | Leutenegger Oberholz | SPS |
| Stefan Müller | CVP | Mathias Reynard | SPS |
| Graber Konrad | CVP | Geraldine Savary | SPS |
| Dominique de Buman | CVP | Manuel Tornare | SPS |
| Filippo Lombardi | CVP | Cesla Amarelle | SPS |
| Candinas Martin | CVP | Paul Rechsteiner | SPS |
| Ruth Humbel | CVP | Marra Ada | SPS |
| Stefan Engler | CVP | Simonetta Sommaruga | SPS |
| EVP Schweiz | EVP | Valérie Piller | SPS |
| Fabio Regazzi | EVP | Roger Nordmann | SPS |
| Marianne Streiff | EVP | Martina Munz | SPS |
| Maja Ingold | EVP | Maire Jacques-André | SPS |
| Felix Gutzwiller | FDP | Marina Carobbio | SPS |
| Hugues Hiltpold | FDP | Rebecca Ruiz | SPS |
| FDP.Die Liberalen | FDP | Eric Nussbaumer | SPS |
| Christian Wasserfallen | FDP | Andy Tschümperlin | SPS |
| Hans-Peter Portmann | FDP | Claude Janiak | SPS |
| Isabelle Moret | FDP | Nadine Masshardt | SPS |
| Schilliger Peter | FDP | Sylvie Perrin-Jaquet | SPS |
| Theiler Georges | FDP | Rossini Stéphane | SPS |
| Christa Markwalder | FDP | Barbara Gysi | SPS |
| Ignazio Cassis | FDP | Chantal Galladé | SPS |
| Ruedi Noser | FDP | Claudia Friedl | SPS |
| Filippo Leutenegger | FDP | Maria Bernasconi | SPS |
| Petra Gössi | FDP | Ursula Schneider Sch | SPS |
| Andrea Caroni | FDP | Thomas Hardegger | SPS |
| Pierre-André Monnard | FDP | Daniel Jositsch | SPS |
| Fathi Derder | FDP | Lukas Reimann | SVP |
| Daniel Stolz | FDP | Toni Brunner | SVP |

*(continued on next page)*

Table 8: Core set of Swiss political Twitter users (cont.).

| **Name** | **Party affiliation** | **Name** | **Party Affiliation** |
|---|---|---|---|
| Hutter Markus | FDP | SVP Schweiz | SVP |
| Johann Schneider-Ammann | FDP | Natalie Rickli | SVP |
| Doris Fiala | FDP | Andreas Aebi | SVP |
| Grünliberale Schweiz | GLP | Grin Jean-Pierre | SVP |
| Beat Flach | GLP | Luzi Stamm | SVP |
| Isabelle Chevalley | GLP | Liliane Maury-Pasquier | SVP |
| Jürg Grossen | GLP | Thomas de Courten | SVP |
| Roland Fischer | GLP | Pierre Rusconi | SVP |
| Martin Bäumle | GLP | Marianne Binder | SVP |
| Thomas Maier | GLP | Christoph Mörgeli | SVP |
| Bastien Girod | GPS | Thomas Hurter | SVP |
| Balthasar Glättli | GPS | Oskar Freysinger | SVP |
| Yvonne Gilli | GPS | Verena Herzog | SVP |
| Antonio Hodgers | GPS | Ulrich Giezendanner | SVP |
| Grüne Schweiz | GPS | Jean-Pierre Graber | SVP |
| aline trede | GPS | Adrian Amstutz | SVP |
| Ueli Leuenberger | GPS | Werner Hösli | SVP |
| Peter Haag | GPS | Alfred Heer | SVP |
| Daniel Vischer | GPS | Claudio Miotti | SVP |
| Regula Rytz | GPS | Hansjörg Knecht | SVP |
| Robert Cramer | GPS | Yves Nidegger | SVP |
| Adèle Thorens | GPS | Florin Schütz | SVP |
| Francine John | GPS | Heinz Brand | SVP |
| Anne Mahrer | GPS | Maximilian Reimann | SVP |
| van Singer Christian | GPS | Céline Amaudruz | SVP |
| Maya Graf | GPS | Andrea Geissbühler | SVP |
| Lorenzo Quadri | Others | | |

Table 9: Keyword gazetteer for the party recognition.

| Regular expression | Party label | Language |
|---|---|---|
| BDP | BDP | de |
| buergerlich.demokratisch | BDP | de |
| bourgeois.democratique | BDP | fr |
| PBD | BDP | fr |
| borghese.democratico | BDP | it |
| christlich.sozial | CSP | de |
| CSP | CSP | de |
| chretien.social | CSP | fr |
| PCS | CSP | fr |
| christdemokrat | CVP | de |
| christlich.demokrat | CVP | de |
| christlichdemokratisch | CVP | de |
| CVP | CVP | de |
| démocrate.*?chretien | CVP | fr |
| PDC | CVP | fr |

Table 9: Keyword gazetteer for the party recognition (continued).

| Regular expression | Party label | Language |
|---|---|---|
| democratico.cristiano | CVP | it |
| PPD | CVP | it |
| evangelische.volkspartei | EVP | de |
| EVP | EVP | de |
| parti.evangelique | EVP | fr |
| PEV | EVP | fr |
| partito.evangelico | EVP | it |
| FDP | FDP | de |
| freisinn | FDP | de |
| liberale.partei | FDP | de |
| liberalen | FDP | de |
| LPS | FDP | de |
| liberaux.radicaux | FDP | fr |
| parti.libéral | FDP | fr |
| PLR | FDP | fr |
| liberali.radicali | FDP | it |
| popolare.democratico | FDP | it |
| GLP | GLP | de |
| gruen.liberal | GLP | de |
| grünliberal | GLP | de |
| PVL | GLP | fr |
| vert.libéral | GLP | fr |
| vert.liberaux | GLP | fr |
| verdi.liberali | GLP | it |
| GP | GPS | de |
| gruene | GPS | de |
| grüne | GPS | de |
| ökoliberal | GPS | de |
| écologiste | GPS | fr |
| parti.ecologiste | GPS | fr |
| verts | GPS | fr |
| ecologista | GPS | it |
| AL | RL | de |
| alternative | RL | de |
| julia | RL | de |
| partei.der.arbeit | RL | de |
| PDA | RL | de |
| alliance.de.gauche | RL | fr |
| la.gauche | RL | fr |
| parti.suisse.du.travail | RL | fr |
| la.sinistra | RL | it |
| partito.operaio.e.popolare | RL | it |
| solidarit | RL | |

Table 9: Keyword gazetteer for the party recognition (continued).

| Regular expression | Party label | Language |
|---|---|---|
| EDU | RR | de |
| eidgenoessisch.demokratisch | RR | de |
| FPS | RR | de |
| freiheitspartei | RR | de |
| nationale.aktion | RR | de |
| schweizer.demokraten | RR | de |
| SD | RR | de |
| démocrats.suisses | RR | fr |
| MCG | RR | fr |
| MCR | RR | fr |
| mouvement.citoyens.genevois | RR | fr |
| mouvement.citoyens.romands | RR | fr |
| suisse.de.la.liberte | RR | fr |
| democratici.svizzeri | RR | it |
| Lega | RR | it |
| lega.ticinesi | RR | it |
| lega.ticino | RR | it |
| svizzero.della.liberta | RR | it |
| unione.democratica.federale | RR | it |
| PSL | RR | |
| JS | SP | de |
| jungsozialist | SP | de |
| juso | SP | de |
| sozialdemokrat | SP | de |
| SP | SP | de |
| PS | SP | fr |
| socialiste | SP | fr |
| socialista | SP | it |
| giso | SP | |
| second@ | SP | |
| seconda | SP | |
| schweizerische.volkspartei | SVP | de |
| SVP | SVP | de |
| union.democratique | SVP | fr |
| PDP | SVP | it |
| unione.democratica.del.centro | SVP | it |
| UDC | SVP | |
| politi | General | |
| gemeinderat | Title | de |
| kantonsrat | Title | de |
| landamman | Title | de |
| landrat | Title | de |
| nationalrat | Title | de |

Table 9: Keyword gazetteer for the party recognition (continued).

| Regular expression | Party label | Language |
|---|---|---|
| regierungspräsident | Title | de |
| regierungsrat | Title | de |
| schultheiss | Title | de |
| staatsrat | Title | de |
| ständerat | Title | de |
| standeskommission | Title | de |
| conseildesétats | Title | fr |
| conseildétat | Title | fr |
| conseilexécutif | Title | fr |
| conseillerauxétats | Title | fr |
| conseillermunicipal | Title | fr |
| conseilmunicipal | Title | fr |
| conseilnational | Title | fr |
| grandconseil | Title | fr |
| consigliocomunale | Title | it |
| consigliodeglistati | Title | it |
| consigliodistato | Title | it |
| granconsiglio | Title | it |

Table 10: Number of accounts by socio-structural characteristics.

| | |
|---|---|
| *Gender* | |
| Women | 433 |
| Men | 1241 |
| *Community size* | |
| Large | 863 |
| Mid-sized | 535 |
| Small | 638 |
| *Language* | |
| German | 1290 |
| French | 628 |
| English | 198 |
| Italian | 80 |
| Spanish | 1 |

Table 11: Number of followers and accounts as well as cumulative activity index.

| Party | Followers | Accounts | Activity |
|---|---|---|---|
| SVP | 142'295 | 304 | 379'556 |
| SPS | 404'700 | 530 | 582'646 |
| FDP | 168'436 | 374 | 315'216 |
| CVP | 154'441 | 304 | 428'451 |
| GPS | 173'247 | 230 | 415'189 |
| GLP | 83'151 | 191 | 290'226 |
| BDP | 28'290 | 108 | 95'007 |
| EVP | 24'037 | 62 | 326'958 |
| EDU | 269 | 1 | 252 |
| Others | 35'486 | 103 | 129'968 |

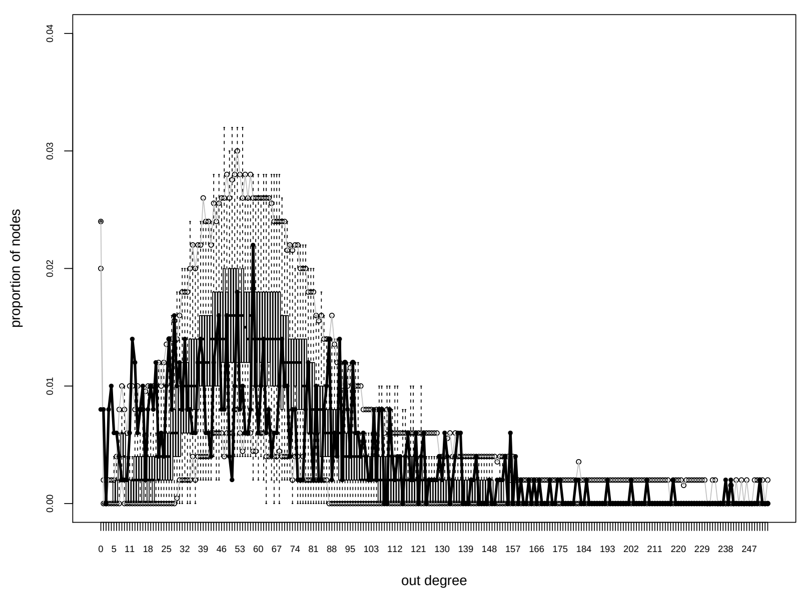



Figure 3: ERGM goodness-of-fit: Outdegree distribution.

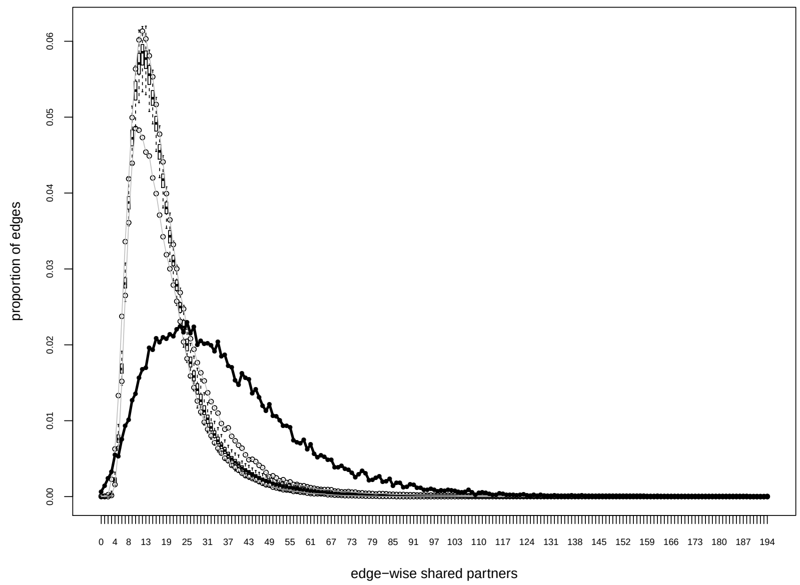



Figure 4: ERGM goodness-of-fit: Edge-wise shared partners distribution.

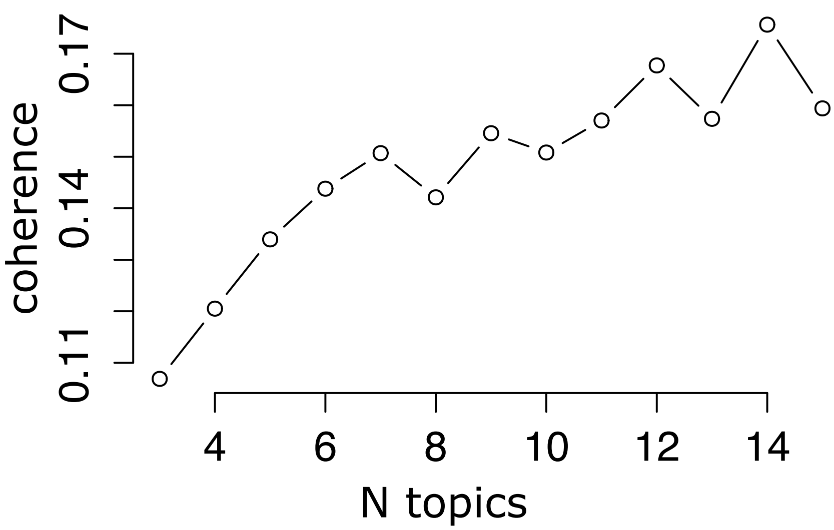



Figure 5: Word2vec topic model coherence.

Table 12: Negative binomial models on number of party mentions

| | SVP | SPS | FDP | CVP |
|---|---|---|---|---|
| (Intercept) | 1.24 | $1.68^{***}$ | $0.92^{*}$ | $2.42^{***}$ |
| | (0.73) | (0.42) | (0.44) | (0.53) |
| Gender | $-0.56$ | $-1.09^{*}$ | $-0.00$ | $-1.03$ |
| (women=1) | (0.63) | (0.51) | (0.49) | (0.59) |
| CVP | $-2.04^{*}$ | $-2.08^{***}$ | $-1.78^{**}$ | |
| | (0.87) | (0.60) | (0.62) | |
| SPS | $-2.54^{**}$ | | $-2.61^{***}$ | $-3.64^{***}$ |
| | (0.86) | | (0.63) | (0.69) |
| FDP | $-1.70$ | $-2.62^{***}$ | | $-3.17^{***}$ |
| | (0.88) | (0.63) | | (0.71) |
| Others | $-2.92^{**}$ | $-1.94^{**}$ | $-1.02$ | $-2.82^{***}$ |
| | (0.94) | (0.60) | (0.60) | (0.71) |
| GPS | $-0.75$ | $-1.23$ | $-1.62^{*}$ | $-2.96^{***}$ |
| | (0.95) | (0.68) | (0.70) | (0.83) |
| SVP | | $-2.13^{**}$ | $-0.84$ | $-2.49^{**}$ |
| | | (0.72) | (0.72) | (0.81) |
| Mid-sized | $-0.04$ | $-1.02$ | | $-1.17$ |
| | (0.75) | (0.70) | | (0.78) |
| Small | $-1.27$ | $-1.52$ | $-1.51$ | $-3.14^{**}$ |
| | (1.09) | (0.82) | (0.84) | (1.18) |
| Language | $-0.99$ | $-0.49$ | $-0.39$ | $-0.60$ |
| (French = 1) | (0.60) | (0.48) | (0.48) | (0.55) |
| AIC | 355.00 | 528.72 | 449.31 | 455.04 |
| BIC | 394.74 | 568.46 | 484.24 | 494.78 |
| Log Likelihood | -166.50 | -253.36 | -214.65 | -216.52 |
| Deviance | 87.10 | 129.71 | 117.12 | 106.06 |
| Num. obs. | 274 | 274 | 243 | 274 |

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 13: Negative binomial models on number of party mentions with inverse ceontrality weights

| | SVP | SPS | FDP | CVP |
|---|---|---|---|---|
| (Intercept) | 1.23 | 1.53* | 0.75 | 2.32** |
| | (1.08) | (0.61) | (0.66) | (0.78) |
| Gender | −0.89 | −1.20 | −0.03 | −1.12 |
| (women=1) | (1.02) | (0.78) | (0.78) | (0.95) |
| CVP | −2.12 | −2.17* | −1.75 | |
| | (1.28) | (0.87) | (0.91) | |
| SPS | −2.61* | | −2.73** | −3.96*** |
| | (1.29) | | (0.98) | (1.08) |
| FDP | −1.88 | −2.65** | | −3.38** |
| | (1.30) | (0.91) | | (1.06) |
| Others | −3.01* | −1.90* | −0.97 | −2.90** |
| | (1.38) | (0.86) | (0.89) | (1.05) |
| GPS | −0.39 | −1.24 | −1.74 | −3.27* |
| | (1.48) | (1.06) | (1.14) | (1.36) |
| SVP | | −2.29* | −0.84 | −2.54* |
| | | (1.05) | (1.06) | (1.19) |
| Mid-sized | −0.18 | −0.86 | | −0.89 |
| | (1.12) | (0.99) | | (1.14) |
| Small | −1.33 | −1.40 | −1.42 | −3.03 |
| | (1.66) | (1.19) | (1.23) | (1.72) |
| Language | −0.93 | −0.34 | −0.20 | −0.46 |
| (French = 1) | (0.87) | (0.67) | (0.69) | (0.81) |
| AIC | 170.70 | 249.28 | 213.41 | 214.46 |
| BIC | 210.36 | 288.95 | 248.26 | 254.13 |
| Log Likelihood | -74.35 | -113.64 | -96.71 | -96.23 |
| Deviance | 38.84 | 60.15 | 52.50 | 46.74 |
| Num. obs. | 272 | 272 | 241 | 272 |

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$