Divided They Tweet: The Network Structure of Political Microbloggers and Discussion Topics

Albert Feller, Matthias Kuhnert, Timm O. Sprenger, Isabell M. Welpe
Technische Universität München
Lehrstuhl für Betriebswirtschaftslehre, Strategie und Organisation
Leopoldstraße 139, 80804 Munich, Germany

Abstract

In the context of a national election, this study explores more than 69,000 Twitter messages containing mentions of political parties and about 2,500 related user profiles to investigate the network structure of political microbloggers with respect to, first, their party preference and, second, the topics they discuss. We find that political microbloggers tend to follow like-minded peers. Microbloggers in a cohesive group tend to have the same political preferences. In addition, we conduct a content analysis of the political debate on Twitter to explore which topics and politicians are discussed and whether this debate reflects an ideological divide among participating users. While there are some discussion topics that are dominated by politically like-minded microbloggers, the majority of topics is discussed by a diverse group of microbloggers with various political preferences.

Introduction

Both scholars (Asur and Huberman 2010; O'Connor et al. 2010) and practitioners (Brustein 2010) have come to use microblogging content as an indicator of public opinion. Political commentators have identified sentiment analysis as a major trend during the 2010 U.S. mid-term election (Brustein 2010) and national TV channels relied on Twitter content for election coverage. However the focus of previous research has been on the investigation of the message content and largely ignored the network structure as a reflection of the offline world. This paper therefore analyzes the network structure of microbloggers by party preference.

The popularity of social media, and especially Twitter, has attracted researchers from all kinds of fields. The scientific work ranges from Twitter's role as a news media (Kwak et al. 2010) to research on user intentions (Java et al. 2007). In the political context, social media has been used to attract potential voters and to foster political discussion. In an investigation of traditional blogs, Adamic and Glance (2005) find that blogs from one end of the political spectrum more often link to other like-minded blogs than to political adversaries. In contrast, politicians are cited more often by bloggers leaning to the opposite party suggesting that bloggers tend to criticize others. This illustrates the scholarly focus on blogs as opposed to the followership of single party members. Tumasjan et al. (2010) have investigated microblogs and show that Twitter messages can reflect the political landscape offline including political preferences of individual users. However, the authors did not explore the network structure among those users. The present study aims to fill this research gap by investigating the network structure of political microbloggers and therefore, the first research question evaluates whether they follow like-minded peers or seek contrarian viewpoints.

With respect to the topics discussed by political microbloggers, Shamma, Kennedy, and Churchill (2009) studied the stream of Twitter messages during the 2008 U.S. presidential debates. Even though the vocabulary used in microblogging messages differs from the rhetoric of the debate, the posts on Twitter still reflect the topics of the debate in real-time and can therefore, be used to analyze the content and semantic structure of media events. The study suggests that Twitter is mainly used for reaction and less for summarizing the topics of the debate.

In sum, related research covers the network structure of traditional blogs and the content discussed in political messages on Twitter. To complement these findings, this paper focuses on the network structure of those discussion topics. Our research objective is to understand which topics are discussed among political microbloggers and how the users, discussing these topics, are linked. The second research question we pursue is whether there is an ideological divide among political microbloggers with respect to the topics they discuss.

This paper contributes to the research on social media in the political context in two ways. First, it explores whether the network structure of users on Twitter reflects political proximity. Second, it confirms the results of Adamic and Glance (2005) in the context of microblogs and shows that groups of ideologically similar microbloggers discuss different political topics more intensely. Nevertheless, the finding that bloggers more often cite political figures with a different opinion could not be verified for microbloggers.

Method and Data

From August 13th to September 9th, 2009 69,318 German tweets containing mentions of one of the six major German

parties (CDU, CSU, FDP, GRUENE, LINKE) identified by a hash tag were crawled from the Twitter network. In order to derive a network of users as party followers, their tweets were used to identify the users' possible party preferences, which resulted in 759 matched profiles. Users preferring more than one party were skipped. Political microbloggers have adopted the convention to add a plus or a minus sign to the parties mentioned in their tweets to express their sentiment towards that party (i.e. "#CDU+"). For each message a user has posted, we evaluated the possible party preference by evaluating "+" or "-" signs added to a party mention.

Next to the messages, we crawled the Twitter network of all users having posted messages in our sample, which allows us to identify connections between these microbloggers. The resulting network of 2,428 users was in a direct manner, reflecting the Twitter network as a network where microbloggers follow each other.

To analyze how the ideological background of users influences their discussion topics, the possible party a user might be in favor of had to be derived from the text of their messages as described above. Even though the negative mentions outweigh the number of positive mentions, this study focuses only on the positive mentions, since they are the strongest indicator for party preference. Naturally, some users composed messages which included more than one positive party mention or favored different parties in distinct messages. As Tumasjan et al. (2010) have already shown, party accounts dedicate 80% of their mentions to their own party. Thus, only users who voted exclusively in favor of one party were used in the later steps to reduce noise. Table 1 shows how many of those 1215 users were in favor of which party.

Table 1: Amount of users exclusively in favor of a party

	CDU	CSU	SPD	FDP	LINKE	GRUENE
Exclusive Users	212	6	376	313	132	176
Share	17,45 %	0,49 %	30,95 %	25,76 %	10,86 %	14,49 %

The resulting data was clustered in parties by an algorithm developed by Wakita and Tsurumi (2007). It clusters the nodes of a graph a way that maximizes the connections within clusters. To visualize the results, we used a directed graph generated by the force-directed Fruchtermann-Reingold algorithm. As a result, the visual proximity of two nodes reflects their relatedness.

To extract the relevant political topics from the data set, a list of all 30,660 words used in the Twitter messages was generated and sorted by frequency. The top 2,000 words of the list (all having at least 50 mentions) were manually classified for belonging to a topic of the political debate or mentioning a high-profile politician of the election. All in all, 98 were labeled as relevant keywords and 24 political persons were extracted. Since the list of topics contained many words describing similar political topics, six clusters of words were built as shown in Table 2. The topic clusters and mentions of politicians were both used to derive a network of Twitter users mentioning the same clusters/mentions.

Table 2: Topic clusters

Topic	Words (examples)
telco data retention	censorship, Zensurla, internet policy, barrier, data retention, internet regulation, surveillance state
Afghanistan	Afghanistan, war, taliban, pullout, pullout plan, troop withdrawal
education	school, university, tuition fees, scholastics, educational policy
nuclear power	nuclear power, Gorleben, AKW, nuclear phase-out, nuclear waste, Asse, permanent repository, nuclear power plant, mileage
financial crisis	crisis, finance, Euro, bank, recession
minimum wages	wage agreement, labor relations, standard wages

Results

This section first shows the network structure of political microbloggers. Then, the structure of the network regarding the discussed topics is described.

Network of microbloggers by political preference

Figure 1 shows the network graph of political microbloggers mentioning the same party more than two times. With this measure as many microbloggers as possible are taken into account while reducing noise by ignoring voters with only one or two party mentions.

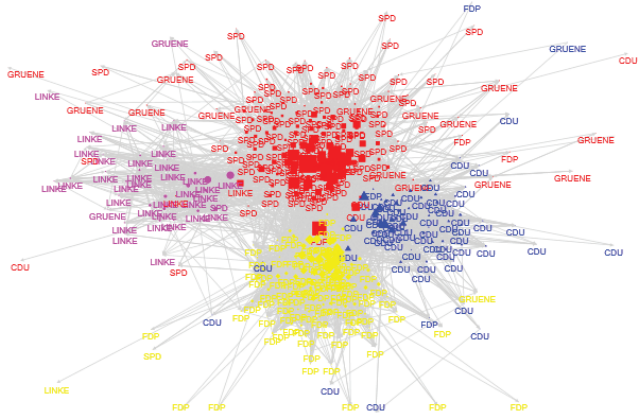

Figure 1: Network of microbloggers

Note: Different colors depict different parties.

Generally, we note a fragmentation between the individual parties. Furthermore the political landscape is reflected in this graph as the combination of the social democratic SPD and the socialist LINKE and the right of center parties CDU and FDP, which find themselves on the same end of ideological spectrum, are more closely connected to each other than to other parties thus forming larger "meta clusters" in line with potential coalitions. Followers of the Green party are assigned to other clusters. Many of them are assigned to the group of SPD followers. This might be a result of the ideologically closeness of these two parties. Besides many connections within single parties there are many connections between parties, especially between SPD and CDU. Table 3 quantifies the results by displaying the network structure between parties in total edges and share. Given that the graph is directed, the matrix is not symmetric.

Table 3: Network structure of microbloggers

	SPD		CDU		FDP		GRUENE		LINKE	
	Total	Share	Total	Share	Total	Share	Total	Share	Total	Share
SPD	1176	**76,26%**	160	10,38%	150	9,73%	21	1,36%	35	2,27%
CDU	123	20,78%	366	**61,82%**	97	16,39%	3	0,51%	3	0,51%
FDP	152	11,55%	171	12,99%	975	**74,09%**	13	0,99%	5	0,38%
GRUENE	47	40,17%	15	12,82%	14	11,97%	36	**30,77%**	5	4,27%
LINKE	72	24,00%	8	2,67%	28	9,33%	3	1,00%	189	**63,00%**

It shows the average number of links between party followers. The party SPD has most internal edges reflecting the big and close cluster of SPD supporters in the graph. Further, SPD followers (76,26%) as well as FDP followers (74,09%) form tight party networks. CDU and LINKE just for about 62%. The Green party is more connected to possible SPD followers (40,17%) than internally explaining the fact that the clustering algorithm found no cluster for this party. In conclusion, we can see that there are many more connections between users in favor of a particular party than between users favoring opposing parties.

The network structure of political discussion topics

The previous analysis has shown that political microbloggers tend to connect to like-minded peers. In this section, we will investigate whether they also tend to debate the same topics as other users supporting the same parties.

Figure 2 shows the network structure of main political topic clusters discussed on Twitter. While some topics are mainly discussed by users in favor of a certain party (e.g. the "data retention" cluster has a high share of users preferring the free-market FDP), many topics are discussed among users of different parties.

Table 4: Share of ideological similar users per topic cluster

Party/Topics	Data Retention		Afghanistan		Minimum Wages		Exclusive Users
	Share	ABS	Share	ABS	Share	ABS	Share
CDU	10,68%	*6,77%*	10,58%	*6,87%*	7,50%	*9,95%*	17,45%
CSU	0,00%	0,49%	0,00%	0,49%	0,00%	0,49%	0,49%
SPD	23,30%	*7,65%*	25,00%	*5,95%*	48,75%	**17,80%**	30,95%
FDP	37,38%	**11,62%**	18,27%	*7,49%*	16,25%	*9,51%*	25,76%
LINKE	12,62%	1,76%	33,65%	**22,79%**	16,25%	*5,39%*	10,86%
GRUENE	16,02%	1,53%	12,50%	1,99%	11,25%	3,24%	14,49%

Note: Deviations (ABS) of the overall share are marked in bold above 10%.

Table 4 allows a more quantitative view on those topics. It displays the network structure of users around the topics they mention. The users are summarized by the party they support. The FDP, which has a total share of 25,76% of exclusive users in the data set, has 37,38% of all users discussing the "data retention" topic. The deviation shows, that this liberal topic was especially relevant for users in favor of the FDP during the election. Another topic which was discussed more frequently by users of one particular party is "Afghanistan". Even though only 10,86% of users are in favor of LINKE, 33,65% of all users who participated in this discussion preferred this party. This is noteworthy, since the LINKE was the only party that favored the withdrawal of German troops from the war in Afghanistan. A third notable fact is the number of users in favor of SPD (social democrats) who discussed the social topic "minimum wages". Nearly half of the members of this discussion preferred the SPD while only a third of all users in the data set have the same political preference.

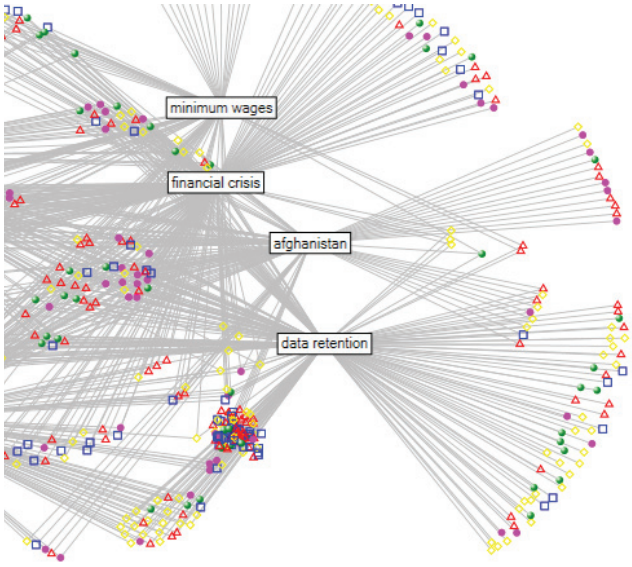

Figure 2: Political microbloggers by discussion topic

Note: Colors and shapes depict different party preferences.

Repeating the above mentioned analysis for the politicians mentioned in the dataset, the mentioned politicians show a higher divide of the Twitter users. Table 5 shows the politicians mentioned most often in the dataset. The share of ideological similar users per politician has in general more deviations than the discussion topics had. For example Guido Westerwelle, Gregor Gysi and Jürgen Trittin, the leaders of FDP, LINKE and GRUENE at that time, all have more mentions from users who prefer their party, respectively. The results also show that acting chancellor Angela Merkel (CDU), who was running for re-election, has relatively many mentions by users who prefer the opposing SPD. Moreover, Horst Seehofer (CSU) is mentioned half of the time by users in favor of the FDP, which he criticized quite a lot during the election campaign.

Discussion and Further Research

As the results show, users who support the same party tend to follow each others' posts more frequently and, thus, form tight-knit clusters in the online community. If the network structure of some users and their field of interest is known, it seems to be possible to infer preferences of connected users. In sum, the network structure of political microbloggers seems to reflect their party preference. In addition, we found that the network structure of topics can be used to indicate political opinion. With respect to the political topics discussed on Twitter, it could be shown that groups of ideologically similar microbloggers discuss different political

Table 5: Share of ideological similar users per mentioned politician

Party/Politician	Merkel (CDU)		Seehofer (CSU)		Steinmeier (SPD)		Exclusive Users
	Share	ABS	Share	ABS	Share	ABS	Share
CDU	23,68%	*6,23%*	7,69%	*9,76%*	22,49%	*5,04%*	17,45%
CSU	0,00%	0,49%	0,00%	0,49%	0,00%	0,49%	0,49%
SPD	44,84%	**13,89%**	28,21%	2,74%	46,75%	**15,80%**	30,95%
FDP	14,61%	**11,15%**	48,72%	**22,96%**	16,86%	*8,90%*	25,76%
LINKE	8,31%	2,55%	10,26%	0,61%	8,28%	2,58%	10,86%
GRUENE	8,56%	*5,92%*	5,13%	*9,36%*	5,62%	*8,86%*	14,49%

Party/Politician	Westerwelle (FDP		Gysi (LINKE)		Trittin (GRUENE)		Exclusive Users
	Share	ABS	Share	ABS	Share	ABS	Share
CDU	7,88%	*9,57%*	14,00%	3,45%	7,84%	*9,61%*	17,45%
CSU	0,49%	0,00%	0,00%	0,49%	0,00%	0,49%	0,49%
SPD	20,20%	**10,75%**	6,00%	**24,95%**	7,84%	**23,10%**	30,95%
FDP	51,72%	**25,96%**	12,00%	**13,76%**	33,33%	*7,57%*	25,76%
LINKE	11,82%	0,96%	68,00%	**57,14%**	9,80%	1,06%	10,86%
GRUENE	7,88%	*6,60%*	0,00%	**14,49%**	41,18%	**26,69%**	14,49%

Note: Deviations (ABS) of the overall share are marked in bold above 20%.

topics more intensely. These topics map to the political position of the respective party in the political landscape.

With respect to the politicians mentioned in the stream of microblogging messages, we found that the share of Twitter users in favor of a certain party can indicate strong opinion. It can be either agreement or disagreement with a certain politician. Overall, our findings lead to the conclusion that the data found in microblogging services like Twitter can be used to gain relevant insights into the political landscape offline.

Even though this study provides intriguing results, it does not come without some caveats. There are several open ends for potential further research. To begin with, the collection of messages on Twitter was restricted to the six major German parties (represented in the German parliament at that time) and did not catch any typographical errors. In addition, Leskovec et al. (2010) have shown that the distinction between positive and negative interactions in a social network can yield powerful insights. This differentiation is implied in our analysis of different parties, however, we do not explicitly examine the sentiment of individual connections in our network. This could be done by an analysis of individual chat messages.

Finally, our results show indications that Twitter can be used to analyze the network structure of discussion topics. Further research should use more properties of the Twitter data to achieve even better results. One possibility would be to take the type of a message (retweet, reply) into account and to follow the thread of a conversation.

Conclusion

On the one hand network data of microbloggers can be used to identify their preferences and extract cohesive subgroups. Our study was conducted in the context of political microblogging, but the results might be transferable to other fields like collaborative filtering.

On the other hand it could be shown that political microbloggers can be structured through the topics and politicians they discuss. We found that groups of ideologically similar microbloggers discuss different political topics more intensely. We found that an accumulation of mentions by users in favor of a party can indicate either a strong positive reaction with that person (e.g. when it is the leader of a party) or a notable dispute (for example, when this person is a member of an opposing party).

Our approach can extract information about semantic structure from an seemingly unstructured media like the microblogging service Twitter. Even though Twitter has no dedicated forums, such as separate bulletin boards, people can still be grouped around certain discussion topics. This method could be further developed to allow politicians and organizations to find lead users and opinion leaders for targeted advertising and PR.

References

Adamic, L. A., and Glance, N. 2005. The political blogosphere and the 2004 u.s. election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, LinkKDD '05, 36–43. New York, NY, USA: ACM.

Asur, S., and Huberman, B. A. 2010. Predicting the future with social media. *CoRR* abs/1003.5699.

Brustein, J. 2010. Nation's political pulse, taken using net chatter. new york times (http://www.nytimes.com/2010/11/-01/technology/01sentiment.html).

Java, A.; Song, X.; Finin, T.; and Tseng, B. 2007. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, WebKDD/SNA-KDD '07, 56–65. New York, NY, USA: ACM.

Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is Twitter , a Social Network or a News Media? Categories and Subject Descriptors. *Most* 591–600.

Leskovec, J.; Huttenlocher, D.; Kleinberg, J 2010. Signed networks in social media. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, 1361–1370. New York, NY, USA: ACM.

O'Connor, B.; Balasubramanyan, R.; Routledge, B.; and Smith, N. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *International AAAI Conference on Weblogs and Social Media*.

Shamma, D. A.; Kennedy, L.; and Churchill, E. F. 2009. Tweet the debates: understanding community annotation of uncollected sources. In *Proceedings of the first SIGMM workshop on Social media*, WSM '09, 3–10. New York, NY, USA: ACM.

Tumasjan, a.; Sprenger, T. O.; Sandner, P. G.; and Welpe, I. M. 2010. Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape. *Social Science Computer Review*.

Wakita, K., and Tsurumi, T. 2007. Finding Community Structure in Mega-scale Social Networks. In *Proceedings of the 16th international conference on World Wide Web*, 1275–1276. ACM.