Small Worlds with a Difference: New Gatekeepers and the Filtering of Political Information on Twitter

Pascal Jürgens
University of Mainz, Germany
Colonel-Kleinmann-Weg 2
55099 Mainz
0049 6131 / 392-5638
pascal.juergens@uni-mainz.de

Andreas Jungherr
University of Bamberg, Germany
Feldkirchenstraße 21
96045 Bamberg
0049 951 / 863-2736
andreas.jungherr@gmail.com

Harald Schoen
University of Bamberg, Germany
Feldkirchenstraße 21
96045 Bamberg
0049 951 / 863-2738
harald.schoen@uni-bamberg.de

ABSTRACT

Political discussions on social network platforms represent an increasingly relevant source of political information, an opportunity for the exchange of opinions and a popular source of quotes for media outlets. We analyzed political communication on Twitter during the run-up to the German general election of 2009 by extracting a directed network of user interactions based on the exchange of political information and opinions. In consonance with expectations from previous research, the resulting network exhibits small-world properties, lending itself to fast and efficient information diffusion. We go on to demonstrate that precisely the highly connected nodes, characteristic for small-world networks, are in a position to exert strong, selective influence on the information passed within the network. We use a metric based on entropy to identify these New Gatekeepers and their impact on the information flow. Finally, we perform an analysis of their input and output of political messages. It is shown that both the New Gatekeepers and ordinary users tend to filter political content on Twitter based on their personal preferences. Thus, we show that political communication on Twitter is at the same time highly dependent on a small number of users, critically positioned in the structure of the network, as well as biased by their own political perspectives.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences – *sociology.*

General Terms

Algorithms, Human Factors, Theory.

Keywords

Bundestagswahl 2009, Entropy, Gatekeepers, Network Analysis, Political Communication, Twitter.

1. TWITTER, SMALL WORLDS AND NEW GATEKEEPERS DURING THE CAMPAIGN OF 2009

During the 2009 German general election Twitter was used prominently by politicians and politically interested users to discuss and comment on political events and the campaign [1]. These public Tweets were then often used by commentators and the media to evaluate the appeal of a given candidate or party to the online audience based on the number of positive or negative tweets connected to that person or party [2]. Thus, via the traditional media, selected political Twitter messages reached a much larger audience than politically interested Twitter users. This observation raises several questions concerning the nature and dynamics of Twitter communication: Are all political opinions equally likely to achieve visibility in the political twittersphere? If not, what determines the visibility of political messages on Twitter? Is the visibility of political messages on Twitter dependent on the selection choices of a few well placed Twitter users?

To answer these questions, we analyzed data on communication networks between politically vocal Twitter users starting June 18 and ending September 30, 2009, the months directly preceding the parliamentary election. Our analysis will show that this network falls into the class of so-called small-world networks as introduced by Watts and Strogatz [3]. One of the properties of small-world networks most commonly commented on is the possibility of information to spread quickly through the entire network even though most members of the network might not be linked to each other directly. This is possible through the existence of a few vertices that connect components of a given network where otherwise structural holes would appear. We will demonstrate a side effect to the role of these vertices as fast distribution hubs of information. Through their critical position in the network the selection choices of these vertices - on which message to retweet or which party to endorse in their original tweets - strongly influences the distribution of political opinions that are most visible in the whole conversation network and which political opinions disappear from view.

In effect, they are able to block, or at least severely hinder, selected political information from reaching the whole conversation network and thus from achieving public prominence.

Adapting a concept from mass communication research, we call these vertices New Gatekeepers. In this paper we will show their importance in the structure of the network of political conversations on Twitter in the run-up to the German parliamentary election of 2009. We will also demonstrate the extent to which the New Gatekeepers exhibited political bias in the Tweets they chose to publish on Twitter. We will add to the literature on communication dynamics on Twitter by showing that the structure of communication networks of politically interested Twitter users has important implications for the prominence of political opinions in the twittersphere.

2. THE DATA

On Twitter, users routinely attribute keywords to a given tweet by preceding it with a # sign. Using the Twitter API, we collected the messages of all users who used one of the following (cf. table 1) politically relevant #keywords at least once between 18 June 2009 and 30 September 2009, three days after the election on 27 September 2009.

Based on these criteria, we identified 33,048 Twitter users who used one of these keywords at least once during the period under scrutiny. During this interval these users posted roughly 10 million Tweets including messages without political keywords. Of these messages, we utilize public messages directed at other users (@reply) and retweets (RTs) to construct a network of manifest communication activity between users [4]. On Twitter, users communicate with others on multiple topics. Thus political conversations bleed over into nonpolitical contexts and are broadcast to multiple audiences [5]. To identify dynamics of political conversation networks we have to select a context-specific subgroup of our political interested users, i.e. a subgroup of users who are actually conversing on political topics.

Table 1. List of politically charged #keywords whose use led to the inclusion in our data set[1]

#cdu	#csu	#union	#spd
#fdp	#gruene / grüne	#piraten	#npd
#linke	#linkspartei	#bundestagswahl	#btw09
#wahl	#sst	#politik	#tvduell
#zensursula	#petition		

3. A NETWORK OF POLITICAL CONVERSATIONS

To construct our network of political conversations we identified all users in our original data set who sent or received a public message (@reply) or retweeted a message containing at least one of the political keywords documented in table 1. Thus we were able to identify a topic-specific community on Twitter that converses on politics and the campaign. This community can be represented by a network that in turn lends itself to quantitative analysis, which can identify critical elements in the structure of the conversations in this community.

In doing this we identify the Twitter accounts of users as vertices in a network. Public messages (@replies) between users and retweets of other users' messages containing political keywords serve as directed edges between these vertices. This results in a network of 8,609 vertices sharing 22,416 edges. This network falls into the class of so-called small-world networks as introduced by Watts and Strogatz [3]. It displays the characteristic short average path length (L=4.9885) and high clustering coefficient (C=0.0200) when compared to a random network of comparable parameters (which has a L of 5.6736 and C of 0.0005). Popular accounts attribute to networks with comparable structural characteristics high efficiency in the distribution of information. We will show that in this case this network structure leads to a different phenomenon, the biased filtering of information based on the selections of Twitter users in critical positions in the structure of the political conversation network.

4. THE NEW GATEKEEPERS

In an attempt to advance the identification of different sets of key players in networks, Stephen Borgatti formulated two problems. 1. Identify the vertices in a network whose elimination would leave a residual network behind with the least possible cohesion (KPP-Neg); 2. Identify the vertices in a network that are maximally connected to all other vertices (KPP-Pos) [7]. In short, the KPP-Pos problem refers to the ability of certain vertices to spread information because of their high connectivity in the network while the KPP-Neg problem focuses on those actors who, because of their structural position between otherwise lightly interconnected parts of the network, are able to disrupt the flow of information.

Applied to political conversation networks, the crucial questions are: 1. KPP-Neg: Who are the users who, based on their position in the structure of the communication network, are particularly capable of influencing the visibility of a tweet or a political information to the network as a whole by not retweeting or referencing it; 2. KPP-Pos: Who are the users who, based on their position in the structure of the communication network, are particularly capable of actively spreading political information through a network.

While prior research focused on Twitter users who were highly influential in spreading information [8], [9], we focus on users who had the strongest possibility to block or disrupt the flow of information. We chose this perspective because in political conversations information is not neutral, but value-laden, affectively charged, and biased [10], [11]. It is thus reasonable to assume that users, instead of retweeting or posting political information regardless of its content, choose to actively select which message to post and which message to ignore.

Based on their position in the conversation networks these users are able to influence the visibility and distribution of political messages and information on Twitter. In mass communication research a similar role is attributed to journalists. Journalists are able to choose which elements of reality to include in their stories. Journalists become Gatekeepers who decide which part of reality the mass media audience experiences through the filter of the news [12]. This perspective represents a departure from the view (popularly utilized for example in epidemic models) that one type of information flows in parallel through a network as long as there are ties, no matter how long they are. Instead, social networks

[1] These keywords document the abbreviations of the major political parties in Germany plus the Pirate Party whose supporters proved to be exceptionally vocal online [6]. In addition to party names the list includes the German term for election, popular abbreviations for several elections in 2009 and a number of keywords representing contagious issues during the campaign.

usually relay different messages in a serial fashion. The analysis of this type of flow on networks necessitates an approach that is able to capture the "ease" with which messages reach their recipients.

Building on this concept, we call those users in our political conversation network with the ability to disrupt the flow of political information based on their position in the structure of the network New Gatekeepers. To identify these users by their structural position in the network, we adapted an algorithm based on work by Borgatti [13], [7] and Ortiz-Arroyo [14]. This algorithm measures the impact which the elimination of a given set of vertices has on the entropy value of the whole network. Specifically, Ortiz-Arroyo [14] proposes a metric combining Shannon's [15] entropy and the probability distribution of centrality across all vertices. He defines the centrality entropy of a graph as (2.8):

$$H_{ce}(G) = -\sum_{i=1}^{n} \gamma(v_i) \times \log_2 \gamma(v_i)$$

Here, γ represents the probability distribution of centrality as the number of shortest paths originating from one vertex divided by the total number of shortest paths:

$$\gamma(v) = \frac{spaths(v_i)}{spaths(v_1, v_2, \ldots, v_M)}, \quad spaths(v_1, v_2, \ldots, v_M) > 0$$

The (non-normalized, ergo graph-specific) value of H_{ce} represents the "ease" with which information can flow through the network. Next, each vertex is in turn removed from the graph and the global centrality entropy re-calculated, yielding the impact of that specific vertex in terms of the KPP-Neg problem:

"Centrality entropy provides information on the degree of reachability for a vertex in the graph. In a fully connected graph the removal of any vertex will have the same effect on centrality entropy as when any other vertex is removed. All vertices are equally important for the flow of information. […] Contrarily, in partially connected graphs, those vertices whose removal will split the graph in two or more parts or that will reduce substantially the number of geodesic paths available to reach other vertices when removed, will have a higher impact in decreasing the total centrality entropy." [14, p.36]

If a particularly well-connected and important node is removed, potential messages have to be routed through other Twitter users. This strategy sounds logical from the perspective of the network, but is highly unlikely to succeed in the absence of intentional co-ordination of structure-conscious users. A drop in centrality entropy can accordingly be linked to the decreased chances of messages traversing the network.

We implemented the algorithm (2.2) [14, p.39] in the python programming language utilizing the igraph network analysis module [16]. After verifying our implementation against Ortiz-Arroyo's examples, we modified the algorithm slightly so that it also works on directed graphs: Our probability distribution of centrality only takes into account edges originating from a vertex. Since communication on Twitter is highly asymmetric in the relationship between the messages a user sends and those she receives (especially so with regard to prominent users with high in-degree), this is a crucial prerequisite.

Data on the entropy impact of individual vertices was collected in the combinatorial method described above by removing one vertex at a time and re-calculating H_{ce} for the entire graph. Since this approach relies on a breadth-first-search of shortest paths which is repeated n times for a graph with n vertices, the overall enterprise has a time complexity of about $O(n^3)$. The calculations were split over several machines and re-aggregated following the popular map-reduce pattern in order to compensate for this encumbrance.

The New Gatekeepers are identified in the algorithm by the degree to which the ease of information flow within the graph is degraded when they are removed. More precisely, New Gatekeepers are those users whose removal decreases information spread within the network by a particularly large degree.

5. NEW GATEKEEPERS DURING THE CAMPAIGN OF 2009

The base connectivity entropy H_{ce} of our network of political communication during the German General Election of 2009 was 12.1596. Removing vertices one by one, we found that they had an impact that ranged from as little as .0000009 to as much as .0938486. Compare this to the example network of Ortiz-Arroyo [14 p.40]: it has only 19 nodes and should be much more susceptible to interruptions, yet the largest impact on entropy is about the same (.1).

Furthermore, the number of nodes with such a large impact is rather small: the second largest drop in connectivity entropy is only 0.0659668. We conclude that there existed few users in the German twittersphere which played a crucial role in information diffusion, enabling them to function as Gatekeepers to political content.

The entropy-based algorithm requires an arbitrary cutoff value - similarly to agglomerative clustering methods - for the identification of key nodes. In our case, we chose to select the top 100 impact nodes as Gatekeepers, about 1% of the total population. Taking into account the quick decline of their effect, an even lower number might be equally suitable (compare figure 1 below).

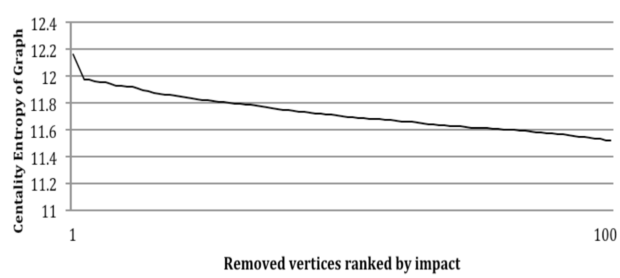

Figure 1: Centrality entropy of the graph upon subsequent removal of 100 gatekeepers

When removed together they decreased the network's connectivity entropy by 0.6418447. To illustrate the destructive effect in traditional network terms: the number of strongly and weakly connected components increased from 5881 to 6574 and from 374 to 1350, respectively. At the same time, the average path length increased from 4.9885 to 5.9110.

Having identified New Gatekeepers, we turn to the question of how these users select information to spread in the network. In offline communication, traditional Gatekeepers were shown to select to present those elements of reality in their stories in accordance with personal, social and economical selection criteria [17]. A similar question arise with regard to the political twittersphere: Do Twitter users in general and New Gatekeepers in particular post political messages whose political position more or less reflects the political position of those messages that the users receive? Or, do the messages they published or chose to retweet show a political position significantly skewed from the political position of the messages which they receive? In short, do tweets accurately reflect the opinion space around a user, or are they a biased selection of these opinions based on choices by the Twitter user? The nature of our data set allows for a precise analysis of this question.

During the campaign of 2009 German users coded their Twitter messages according to their political convictions by including specific #keywords in their messages. If in agreement with a political party, they posted the party label followed by a + (#cdu+; #spd+). If in critique of a party, they posted the party label followed by a - (#cdu-; #spd-). For every user of our conversation network we computed the exact count of these positive or negative mentions of political parties contained in Twitter messages in her direct neighborhood in the network (the vertices a given vertex is directly linked to by an edge). We then computed the thus marked counts of positive or negative mentions of political parties in the Tweets the user herself published. In a next step we tested if the distribution of these counts showed significant differences with a chi square test. If true, we took this as evidence of bias. We used this procedure for all users in our political conversation network.

Table 2 exemplarily shows the bias of one New Gatekeeper in our data set. The table shows the relationship of outgoing and incoming messages in support of seven political parties for the Twitter account of the German politician Volker Beck, Member of Parliament for the Green Party (@volker_beck). The table shows that in the direct neighborhood of Volker Beck's network a total of 1,269 messages in support of the FDP (liberals), containing the keyword #fdp+, were posted. We also see that Volker Beck did not post any message, containing #FDP+, in support of the FDP. Thus there is a bias, a systematic difference between the political opinions voiced in Volker Beck's direct information environment and the political opinion Volker Beck chose to express in his own Tweets.

Table 2. Political bias of the Twitter account @volker_beck by the German politician Volker Beck (Bündnis 90/Die Grünen)

	Outgoing[2]	Incoming[3]
CDU (conservatives)	0	511
CSU (conservatives)	0	58
SPD (social democrats)	0	876
FDP (liberals)	0	1,269
Grüne (green party)	19	630
Linke (socialists)	0	267
Piraten (pirate party)	10	4,944

Of the users in the conversation network, roughly 3,000 had enough in- and outgoing Twitter messages to statistically test for a bias. For negative mentions of parties, 2,866 users had a significant bias ($p < .001$) while a mere 201 users did not exhibit a significant bias. It should be noted that a bias could be identified for all of the testable gatekeepers (N=97). The situation is quasi identical for positive mentions of parties: 3,835 users with a significant bias stand against only 34 users without significant bias. Again, all of the gatekeepers (N=100) were found to have sent political messages with a significantly different distribution from their directly connected neighbors.

Consequently the messages of many Twitter users and those of nearly all critically placed New Gatekeepers do not accurately reflect the distribution of political opinion on Twitter but only the political opinion of that user. Add to that the heavy influence of a user's structural position in the conversation network on the visibility of her opinion and we have a situation in which personal opinions of Twitter users are heavily amplified by the technological bias of the structure of conversation networks on Twitter. Journalists reporting on the political opinion on Twitter might thus only be reporting on the political opinion of a few individuals critically positioned in the structure of the conversation network.

6. NEW MEDIA, TECHNOLOGY, AND THE EMERGENCE OF NEW GATEKEEPERS

Our paper thus contributes to the literature on small-world networks since we were able to show that in communication networks small-world structures do not necessarily lead to the fast and impartial distribution of information through network graphs. Instead, these networks are highly dependent on the willingness of users in critical network positions to distribute potentially expressive messages. Thus these individuals become information filters. This is especially relevant in the case of information distribution in highly partisan communication contexts, as for example politics. In our paper we showed the existence of New Gatekeepers, Twitter users in critical structural position in

[2] Tweets in support of a political party published by Volker Beck.

[3] Tweets in support of a political party published in Volker Beck's direct network neighborhood.

communication networks with the very real ability to disrupt the flow of information through the network. We also showed that these users do not tend to be impartial judges but instead tend to post messages whose political position significantly diverges from the messages these users themselves perceived in their direct communication environment on Twitter. Thus, New Gatekeepers are not neutral distribution hubs that indiscriminately distribute the information that reaches them. Instead they are very conscious curators of, in our case, political information. This analysis also shows that the examination of network structure in empirically measured communication networks holds strong potential for research into communication dynamics.

7. REFERENCES

[1] Jungherr, A. 2009. Twitternde Politiker: Zwischen buntem Rauschen und Bürgernähe 2.0. In Bieber, C., Eifert, M., Groß, T., and Lamla, J. (eds.). *Soziale Netze in der digitalen Welt: Das Internet zwischen egalitärer Beteiligung und ökonomischer Macht*. Campus Verlag, Frankfurt am Main. 99-127.

[2] heise online. 2009. Schäfer-Gümbel schlägt Koch – aber nur im Netz. Available online: http://www.heise.de/newsticker/meldung/Schaefer-Guembel-schlaegt-Koch-aber-nur-im-Netz-197987.html.

[3] Watts, D. J., and Strogatz, S. H. 1998. Collective dynamics of 'small-world' networks. *Nature* 393, 440-442. DOI- http://dx.doi.org/10.1038/30918.

[4] Jürgens, P., and Jungherr, A. 2011. Wahlkampf vom Sofa aus: Twitter im Bundestagswahlkampf 2009. In: Schweitzer, E. J., and Albrecht, S. (eds.). *Das Internet im Wahlkampf: Analysen zur Bundestagswahl 2009*. VS-Verlag, Wiesbaden, 201-225. DOI- http://dx.doi.org/10.1007/978-3-531-92853-1_8.

[5] Marwick, A., and boyd, d. 2011. I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media and Society* 13, 96-113. DOI- http://dx.doi.org/10.1177/1461444810365313.

[6] Jungherr, A., Jürgens, P., and Schoen, H., 2012. Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. 'Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment'. *Social Science Computer Review*. Forthcoming. DOI- http://dx.doi.org/10.1177/0894439311404119.

[7] Borgatti, S. P. 2006. Identifying sets of key players in a social network. *Computational & Mathematical Organization Theory* 12 (1), 21-34. DOI- http://dx.doi.org/10.1007/s10588-006-7084-x.

[8] Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. P. 2010. Measuring User Influence in Twitter: The Million Follower Fallacy. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media* (Washington, DC, May 23 – 26, 2010). ICWSM. AAAI Press, Menlo Park, CA, 10-17.

[9] Bakshy, E, Hofman, J. M., Mason, W. A., and Watts, D. J. 2011. Everyone's an Influencer: Quantifying Influence on Twitter. *Proceedings of the Fourth International Conference on Web Search and Data Mining* (Hong Kong, China, February 09 – 12, 2011). WSDM 2011.

[10] Steenbergen, M. R., and Lodge, M. 2003. Process Matters: Cognitive Models of Candidate Evaluation. In: MacKuen, M. B., and Rabinowitz, G. (eds.). *Electoral Democracy*. University of Michigan Press, Ann Arbor, 125-171.

[11] Taber, C. S. 2003. Information Processing and Public Opinion. In: Sears, D. O., Huddy, L., and Jervis, R. (eds.). *Oxford Handbook of Political Psychology*. Oxford University Press, Oxford, 433-476.

[12] Shoemaker, P. J., Vos, T. P., and Reese, S. D. 2009. Journalists as Gatekeepers. In: Wahl-Jorgensen, K., and Hanitzsch, T. (eds.). *The Handbook of Journalism Studies*. Routledge, New York et al., 73-87.

[13] Borgatti, S. P.,2005. Centrality and network flow. *Social Networks* 27, 55-71. DOI - http://dx.doi.org/10.1016/j.socnet.2004.11.008.

[14] Ortiz-Arroyo, D. 2010. Discovering Sets of Key Players in Social Networks. In: Abraham, A., Hassanien, A.-E., and Snásel , V. (eds.). *Computational Social Network Analysis*. Springer Verlag, Dordrecht et al., 27-46. DOI- http://dx.doi.org/10.1007/978-1-84882-229-0_2.

[15] Shannon, C.E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27, 379–423, 623–656.

[16] Csárdi, G, and Nepusz, T.,2006. The igraph software package for complex network research. *InterJournal Complex Systems*, 1695.

[17] Shoemaker, P. J., Eichholz, M., Kim, E., and Wrigley, B. 2001. Individual and Routine Forces in Gatekeeping. *Journalism and Mass Communication Quarterly* 78(2), 233-246.