

Andreas Jungherr, Pascal Jürgens, and Harald Schoen. 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* 30(2): 229-234. doi: 10.1177/0894439311404119

**Preprint: Please cite published version!**

**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., & Welpel, I. M. “Predicting Elections with  
Twitter: What 140 Characters Reveal about Political Sentiment”**

Andreas Jungherr, University of Bamberg

Pascal Jürgens, University of Mainz

Harald Schoen, University of Bamberg

Keywords: computational social science, elections, predictions, social media,  
Twitter

Predicting election outcomes has become some kind of a cottage industry in the social sciences in previous years. In their paper “Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment” the authors Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner and Isabell M. Welpel (TSSW) join the ranks of election forecasters with an innovative proposal. In particular, they suggest that the relative frequency of party mentions on the microblogging platform Twitter predicts the outcome of German elections:

the mere number of tweets mentioning a political party can be considered a plausible reflection of the vote share and its predictive power even comes close to traditional election polls (Tumasjan et al. 2010: 183).

This claim is based on the observation that during a specific time period (August 13 and September 19, 2009) (180) “the relative volume of tweets mirrors the results of the federal election closely” (182 f.). Moreover, the authors find that the mean absolute error of their Twitter-based prediction is almost as small as the mean absolute error of popular German election polls (unfortunately without stating dates of the polls that they are comparing). They thus conclude that “Twitter can be seen as a valid real-time indicator of political sentiment” (184). This conclusion fits quite nicely into prior research suggesting that user inputs in online services are useful indicators of offline trends. Be it the tracking of influenza or consumer prices based on search terms (Ginsberg et al. 2009; Choi and Varian 2009), the inference of relationships between people based on their phone records (Eagle, Pentland and Lazer 2009), or the structure of TV events or the box-office of movies based on the real-time comments on Twitter (Shamma, Kennedy and Churchill 2010; Asur and Huberman 2010). These interactions between users of online services leave traces which in turn might allow us to gain a deeper understanding of the social dynamics that produced them (Lazer et al.

2009: 721). Thus the general idea that there might be a connection between political twitter messages and the offline political process is not as far-fetched as it might sound.

Yet, a prediction instrument is useful for the scientific community only if there are clear and well-grounded rules how to employ it. Otherwise scholars will not be able to utilize it to predict future elections. In this regard, however, TSSW's paper turns out to be somewhat lacking. In particular, the authors do not specify several details and make arbitrary choices without giving any substantive account. What is more, the results of their analysis appear to critically depend on these choices as we shall demonstrate empirically<sup>1</sup>.

### **And the Winner is – the Pirate Party**

In the run-up of the 2009 German election, we collected a data set comprising all twitter messages of users whose tweets at least once contained the name of a German party or a politically loaded keyword<sup>2</sup>. Unfortunately, TSSW are somewhat vague on their procedure of data collection. They do not specify whether they, e.g., queried the Twitter API, whether they scraped web-pages, whether they utilized RSS feeds of saved searches or whether they use third party applications. As a result, we do not know in which way their mode of data collection differs from the procedure we employed. That there is a difference in

data collection cannot be disputed, however, since the absolute number of party mentions we collected from August 13 to September 19 2009 is much larger than the number of mentions measured by TSSW (see bottom row in Table 1). This considerable difference suggests that analyses of Twitter messages should specify the details of data collection. In particular, a new method for election prediction cannot be widely applied if it is unclear how to replicate it.

Despite the difference in absolute numbers, however, the results on the relative frequencies of party mentions resemble each other quite closely. To be sure, the proportion of mentions of the SPD is considerably lower in our data than in the TSSW data, whereas the opposite holds for the proportion of mentions of the Greens. In general, however, the two distributions are quite similar. The mean absolute difference between both distributions is 2.18. When comparing the vote shares of the six parties included in the TSSW analysis to the proportion of Twitter mentions, the TSSW data exhibit a mean absolute difference of 1.65 percentage points, while our data yield a difference of 1.51 percentage points. We thus conclude that that our data set, while not identical, exhibits similar patterns as the TSSW data set. We are therefore confident that the analyses reported below would yield similar results when performed with the TSSW data, rather than ours.

Table 1 about here

The first issue we would like to raise concerns the identification of the parties whose mentions should be counted to create an election prediction. TSSW chose to include six parties, CDU, CSU, SPD, FDP, Die Linke, and the Grüne. They do not make transparent the reasoning behind this choice, however. Giving no account or rules, obviously, limits the applicability of the proposed instrument for election prediction in other elections. What is more, TSSW's actual choice does not fit nicely with the claims that "the mere number of messages reflects the election result" and "Twitter can be seen as a valid real-time indicator of political sentiment" (184). While these claims suggest that Twitter reflects political sentiment, irrespective of which political party is concerned, TSSW include only those parties that entered German parliament in 2009.

To test whether the results of TSSW's method depend upon which parties are included, we repeated our analysis including a seventh party, namely the Pirate Party. This party mounted considerable support in online forums, blogs and on Twitter in the run-up of the German election of 2009. The online presence of the supporters of the Pirate Party led to widespread speculation in German media and academia if online channels would profoundly change political participation in Germany (see, e.g., Bartels 2009; Becker 2009; Grabowsky 2009; Knop 2009; Teevs 2009; Zolleis, Prokopf, and Strauch 2010). Against this backdrop, excluding the Pirate Party is not an obvious choice.

As Table 2 indicates, including the Pirate Party alters the results considerably. Accordingly, the Pirate Party gained the largest proportion of party mentions and, given TSSW's line of reasoning, would have been the winner of the 2009 German federal election. It would be the major partner in a governing coalition and probably even provide the German chancellor. In the end, however, the Pirate Party managed to gather only some 2 percent of the votes. So, it is no surprise that the mean absolute error (MAE) of the prediction rises to more than 9 percentage points. As a result, adding a single political party to the analysis decreases the predictive power tremendously. The performance of the TSSW method thus depends critically upon the exclusion of certain parties for whom the instrument performs poorly – without specifying any criteria.

Table 2 about here

Yet, TSSW might object that before the 2009 federal election pundits and scholars alike deemed the Pirate Party as incapable of entering German parliament in the upcoming election. Given this knowledge, they – so the argument might go – deliberately excluded the Pirate Party from their analysis. This objection raises two issues, however. To begin with, the authors should make plainly clear the rules that led them to this decision. Otherwise their method cannot be applied to future elections. Moreover, this counterargument raises the question from which

source TSSW's pre-election knowledge concerning the Pirate's Party non-viability stemmed. They probably relied, *inter alia*, on the results of public opinion surveys and deemed them sufficiently valid indicators of the election outcome. If results of polls played a role in deciding upon the inclusion of particular parties, the TSSW method is dependent on public opinion surveys. It is thus not an independent tool for election prediction and can hardly "complement traditional methods of political forecasting" (184)<sup>3</sup>.

Consequently, we suggest that a valid methodological approach should be capable of explaining a substantial amount of the data and it should specify its limitations. For the study of electronic communication channels in general and Twitter in particular, this means that one should be careful not to start with a narrow scope. Otherwise, e.g., by excluding channel-specific aspects (such as the Pirate Party) which are potentially meaningful, one runs the risk of producing biased results.

### **What a Difference a Day Makes**

The second issue we would like to raise concerns the period during which TSSW gathered their data. For their analysis, TSSW collected Twitter messages containing the names of six political parties that were published between August 13 and September 19, 2009 (180). They do not give any account of why they decided to choose this period. Moreover, a closer look at the period does not give any hint at which reasoning led the authors to choose it. August 13 is a Thursday

with no particular significance in the campaign. Their data collection stops on a Saturday five weeks later, on September 19. This date, eight days before the election on September 27, is also a date with no obvious significance in the campaign. As with the choice of parties, the lack of explicit or implicit rules poses a problem for scholars who would like to apply the method to future elections. Leaving the rules for choosing a time frame unspecified would be no problem if counting party mentions in the Twittersphere led to identical results, irrespective of which period is chosen. To address this issue empirically, we calculated absolute errors (prediction based on mentions of the names of the parties in our Twitter data as compared to the actual election results) and the mean absolute error of the complete prediction for different time periods<sup>4</sup>. The results reported in Table 3 show that the absolute errors between predictions based on mentions of party names in our replication data are far from stable and vary for each party depending heavily on the time frame. The same is also true for the mean absolute error (MAE) of all predictions. Especially interesting is the fact that the mean absolute error of all predictions for an analysis of Twitter messages sent between August 13 and September 27, the day of the election, produces a mean absolute error of 2.13 percentage points which is significantly higher than the mean absolute error for the TSSW time frame (1.51). Including only eight additional days results in a considerable increase in the mean absolute difference. The results thus depend considerably on the period under scrutiny. It is thus particularly

unfortunate that TSSW did not specify any rule of how to determine the correct period.

Table 3 about here

As with the issue of the Pirate Party, we want to stress the importance of reasonably-chosen and well-documented sampling decisions. For any kind of aggregate data analysis, it is highly desirable that the measured results be as stable as possible against modifications of the data time frame. Data sets that exhibit significant variance and/or trends should be carefully investigated. The assumption that Twitter mood mirrors some kind of public sentiment should thus be tested against the worst cases of a data set. If any significant differences are found, it should be worthwhile to try and explain them.

## **Conclusion**

In their paper, TSSW claim that “the mere number of tweets mentioning a political party can be considered a plausible reflection of the vote share and its predictive power even comes close to traditional election polls”. Our analysis demonstrates that this conclusion is not warranted for several reasons. The authors did not specify well-grounded rules for data collection in general and the choice of parties and the correct period in particular. Scholars will thus not be able to

apply TSSW's method to future elections. Moreover, the performance of their indicator varies over time as well as it critically hinges upon which subset of the German party system is covered. The number of party mentions in the Twittersphere is thus not a valid indicator of offline political sentiment or even of future election outcomes. All in all, the TSSW paper did not make outcomes of German federal elections more predictable.

**Andreas Jungherr** is research fellow at the chair of political sociology of the University of Bamberg, Germany and may be contacted at [andreas.jungherr@gmail.com](mailto:andreas.jungherr@gmail.com).

**Pascal Jürgens** is a graduate from the mass communication department of the University of Mainz, Germany and may be contacted at [pascal.juergens@googlemail.com](mailto:pascal.juergens@googlemail.com).

**Harald Schoen** is professor of political science at the University of Bamberg, Germany and may be contacted at [harald.schoen@uni-bamberg.de](mailto:harald.schoen@uni-bamberg.de).

## References

Asur, S., & Huberman, B. A. (2010). *Predicting the Future with Social Media*.  
<http://arxiv.org/abs/1003.5699>.

- Bartels, H. (2009). *Die Piratenpartei: Entstehung, Forderungen und Perspektiven*. Berlin: Contumax-Verlag.
- Becker, M. (2009). "Das war wie ein Blitzschlag" *stern.de*. 11. September 2009.  
Available at: <http://www.stern.de/wahl-2009/bundestagswahl/piratenpartei-das-war-wie-ein-blitzschlag-1508537.html>.
- Choi, H., & Varian, H. (2009). *Predicting the Present with Google Trends*.  
Mountain View, California: Google Inc.
- Eagle, N., Pentland, A., & Lazer, D. (2009). "Inferring Social Network Structure using Mobile Phone Data". *Proceedings of the National Academy of Sciences* 106 (36): 15274-15278.
- Gelman, A., & King, G. (1993). "Why are American presidential election campaign polls so variable when votes are so predictable?" *British Journal of Political Science* 23 (4): 409-451.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). "Detecting influenza epidemics using search engine query data". *Nature* 457.
- Grabowsky, F. (2009). "Interview mit Jörg Schönenborn: Die Piratenpartei füllt ein Angebotsvakuum". *tagesschau.de*. 22. Juli 2009. Available at:  
<http://www.tagesschau.de/inland/schoenenbornpiraten100.html>.
- Jürgens, P., & Jungherr, A. (2011). "Wahlkampf vom Sofa aus: Twitter im Bundestagswahlkampf 2009". In: E. J. Schweitzer, & S. Albrecht (eds.).

*Das Internet im Wahlkampf: Analysen zur Bundestagswahl 2009.*

Wiesbaden: VS Verlag. (Forthcoming).

Knop, C. (2009). "Patent-Piraten auf Wählerfang". *FAZ.NET*. 18. September 2009. Available at:  
<http://www.faz.net/s/Rub4D6E6242947140018FC1DA8D5E0008C5/Doc~E60EB9066EF7C45BF9D79056B6A98D71B~ATpl~Ecommon~Scontent.html>.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). "Computational Social Science," *Science* 323 (5915): 721-723.

Shamma, D. A., Kennedy, L., & Churchill, E. F. (2010). "Conversational Shadows: Describing Live Media Events Using Short Messages". *Proceedings of the Fourth International AAI Conference on Weblogs and Social Media*. Menlo Park, California: The AAI Press. 331-334.

Teevs, C. (2009). "Piratenpartei greift offline an". *Spiegel Online*. 13. August 2009. Available at:  
<http://www.spiegel.de/politik/deutschland/0,1518,642348,00.html>.

Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welp, I. M. (2010). "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," *Proceedings of the Fourth International AAI Conference on*

*Weblogs and Social Media*. Menlo Park, California: The AAAI Press. 178-185.

Zolleis, U., Prokopf, S., & Strauch, F. (2010). *Die Piratenpartei: Hype oder Herausforderung für die deutsche Parteienlandschaft?* München: Hans-Seidel-Stiftung.

All links were accessible as of December 10, 2010.

## **Endnotes**

(1) In this response, we note, but do not address, the more theoretical issue that TSSW do not provide a causal link between tweets and votes. In this regard, they appear to follow the logic of scientific instrumentalism, rather than realism.

(2) The following keywords led to an inclusion of the messages of a twitter user in our data set: cdu, csu, spd, fdp, gruene/grüne, piraten, npd, linke, zensursula, bundestagswahl, petition, politik, btw09, wahl, sst, Linkspartei, union, tvduell. For a detailed discussion of this data set see Jürgens and Jungherr 2011).

(3) Of course, polls usually aim at reflecting public opinion at a given time, rather than at predicting election outcomes in the future (cf. Gelman and King 1993).

(4) To ease comparison to TSSW results, we excluded the Pirate Party from this analysis.

**Table 1: Aggregate counts of party mentions in TSSW data and in our data**

Party	No. of mentions (TSSW)	% of mentions	No. of mentions (replication)	% of mentions
CDU	30,886	30.1	51,165	28.6
CSU	5,748	5.6	8,346	4.7
SPD	27,356	26.6	40,419	22.6
FDP	17,737	17.3	30,710	17.2
LINKE	12,689	12.4	22,775	12.7
Grüne	8,250	8	25,410	14.2
Total	102,666		178,825	

**Table 2: Parties' vote shares and proportions of Twitter mentions including the Pirate Party**

Party	Election results	Share of Twitter messages (replication)
CDU	28.4	18.6
CSU	6.8	3.0
SPD	24.0	14.7
FDP	15.2	11.2
Linke	12.4	8.3
Grüne	11.1	9.3
Piraten	2.1	34.8

**Note:** Following TSSW, when calculating vote shares, we included only the votes cast for the seven parties under scrutiny.

**Table 3: Absolute Errors of Predictions based on Party Mentions in our Twitter Data as Compared to the Actual Election Results**

	13.8.-19.9 (TSSW)	13.8.-27.9	13.8.-19.9	20.8.-19.9	27.8.-19.9	3.9.-19.9	10.9.-19.9	17.9.-19.9
CDU	1.0	1.95	0.39	0.58	1.42	1.62	2.65	2.60
CSU	1.3	2.22	2.23	2.28	2.3	1.75	2.03	3.00
SPD	2.2	2.21	1.9	1.99	1.75	2.33	1.82	4.43
FDP	1.7	3.04	1.67	2.01	2.22	2.83	2.59	3.14
Linke	0.3	0.03	0.04	0.03	0.31	0.40	0.53	0.39
Green	3.3	3.31	2.81	2.81	2.93	2.47	3.38	6.51
MAE	1.6	2.13	1.51	1.62	1.82	1.90	2.17	3.34