*Information, Communication & Society*, 2015
http://dx.doi.org/10.1080/1369118X.2015.1069871

Routledge
Taylor & Francis Group

# Wiki-worthy: collective judgment of candidate notability

Drew B. Margolin[a*], Sasha Goodman[b], Brian Keegan[b], Yu-Ru Lin[c] and David Lazer[b,d]

[a]*Department of Communication, Cornell University, Ithaca, NY 14850, USA;* [b]*Department of Political Science and College of Computer and Information Science, Northeastern University, Boston, MA 02115, USA;* [c]*School of Information Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA;* [d]*John F. Kennedy School of Government, Harvard University, Cambridge, USA*

(*Received 9 January 2015; accepted 30 June 2015*)

The use of socio-technical data to predict elections is a growing research area. We argue that election prediction research suffers from under-specified theoretical models that do not properly distinguish between 'poll-like' and 'prediction market-like' mechanisms understand findings. More specifically, we argue that, in systems with strong norms and reputational feedback mechanisms, individuals have market-like incentives to bias content creation toward candidates they expect will win. We provide evidence for the merits of this approach using the creation of Wikipedia pages for candidates in the 2010 US and UK national legislative elections. We find that Wikipedia editors are more likely to create Wikipedia pages for challengers who have a better chance of defeating their incumbent opponent and that the timing of these page creations coincides with periods when collective expectations for the candidate's success are relatively high.

**Keywords:** election prediction; crowdsourcing; Wikipedia; politics; social media; communication studies



Social media and other socio-technical systems provide researchers with unique opportunities to observe collective activity (Lazer et al., 2009). One area of growing interest is the assessment of collective political opinion and the prediction of election outcomes. Thus far, however, election predictions based on socio-technical aggregates have had inconsistent success. Early work suggested that the volume of Twitter messages mentioning a party reflected both its popularity and its success in upcoming elections (Tumasjan, Sprenger, Sander, & Welpe, 2010); however, this work was subsequently criticized for biases in the population and temporal samples employed (Jungherr, Jurgens, & Schoen, 2012) An analysis of Google search data also showed only a weak correspondence between search volume and election outcomes (Gayo-Avello, Metaxas, & Mustafaraj, 2011). In the 2008 US presidential race, sentiment in Twitter streams was closely correlated with public opinion surveys of consumer confidence and political opinion, but did not track pre-electoral polls (O'Connor, Balasubramanyan, Routledge, & Smith, 2010). Nonetheless, DiGrazia, McKelvey, Bollen, and Rojas (2013) found that candidate mentions on Twitter were predictors of election outcomes in US elections. In a follow-up paper, McKelvey, DiGrazia, and Rojas (2014) show that a primary predictor of electoral success is 'buzz' among non-political elites, that is, the

*Corresponding author. Email: dm658@cornell.edu

extent to which a candidate draws chatter from the public. Unankard, Li, Sharaf, Zhong, and Li (2014) also show that elections can be predicted from sentiment using 'sub-event' analysis.

We argue that these conflicting findings are partly the result of an underspecified theoretical model that explains *why* data from socio-technical systems should reflect prospective election outcomes. The absence of an explicit theoretical explanation makes it difficult to identify the boundary conditions and adjustments required to generalize findings from one dataset to another. When predicting elections from socio-technical system behavior, researchers should differentiate between two distinct theoretical models: *polls*, in which individual behavior is viewed as a representation of that individual's personal attitudes regarding a topic; and *prediction markets*, in which an individual's behavior is viewed as a representation of their expectations about the collective attitudes of others (Arrow et al., 2008; Sunstein, 2006). We do not suggest distinguishing these models because we expect one to be superior. There may be cases where both mechanisms operate independently and others in which one may be dominant. By distinguishing them, however, researchers can anticipate the influence of each within a particular dataset and calibrate their methods and interpretations accordingly.

To highlight the importance of prediction market-like behavior in a socio-technical system, we show the correlation between election outcomes and a socio-technical behavior that is regulated to restrict the influence of poll-like mechanisms: the creation of articles for political candidates on Wikipedia. Drawing on the Wikipedia history for major party challengers running against incumbents in the 2010 US and UK legislative elections, we show that while the majority of challengers do not have Wikipedia pages during the campaign, contributors are more likely to create pages during the campaign for challengers who eventually win. We then provide evidence that both the act and timing of page creation are associated with a candidate's expected prospects for victory. We suggest that these results reflect prediction market-like behavior in which individuals weigh the cost and benefit of contributing based on their assessment of the likely outcome of a future event.

Though our results provide evidence that contributions to Wikipedia are motivated by expectations for electoral success, these expectations appear to be largely based on information already available in the political information environment. Specifically, we find that Wikipedia article creation contributes minimally to improvements in prediction accuracy over a baseline fundamentals model, with these contributions confined to a small number of 'competitive' elections. Thus, rather than presenting a new prediction algorithm, this paper contributes to the basic science of articulating a theoretical relationship between election outcomes and socio-technical behavior with the goal of encouraging the development of more sophisticated, theoretically based models.



## Models of election prediction

### *Importance of models*

Theoretical models specify a causal path between variables for which there is an observed association (Morgan & Winship, 2007). These paths help to identify the boundary conditions under which the association should be expected to generalize. For example, in predicting disease incidence, if Google searches for 'flu' correlate with the number of cases of flu *because* most searches are performed by people who have symptoms, it follows that if the mix of reasons for 'flu' searches changes, this method will become less reliable (Lazer, Kennedy, King, & Vespignani, 2014).

Election predictions from socio-technical systems rarely make their theoretical models explicit. Most predictive models measure sentiment and stated intentions of large numbers of individuals, implying that their underlying model is that of a political survey. Unfortunately, proper

surveys require representative sampling and/or adjustments to produce reliable estimates of public opinion (Hayes, 2005), and representative sampling is difficult to achieve with most socio-technical data because participation in most systems is both non-random and self-selected (Heckman, 1979; Lin, Margolin, Keegan, & Lazer, 2013). Perhaps for this reason, election prediction studies do not explicitly claim that they are approximating polling methodology.

Furthermore, claims to approximate polls are often unnecessary for making successful predictions. Political scientists have been successful at predicting elections using *fundamentals models,* in which economic or other structural indicators of a voting population's general satisfaction or mood are used to predict the success of incumbent political candidates (Nannestad & Paldam, 1994). Socio-technical aggregates, such as 'buzz' on social media (McKelvey et al., 2014), may similarly reflect a property of the body politic that is distinct from individuals' sentiments about specific candidates but nonetheless correlated with election outcomes.



### *Prediction markets*

An alternative theoretical model is the prediction market (Sunstein, 2006). Whereas polls work by aggregating a representative sample of individuals' personal attitudes, prediction markets work by aggregating individuals' estimates about the attitudes of the public as a whole (Sunstein, 2006). In essence, where polls ask 'what do you think about this candidate?', prediction markets ask 'what do you think the majority of other people think about this candidate?'

In general, the success of prediction markets for predicting election outcomes is comparable to that of polls (Berg, Nelson, Rietz, 2007; Rothschild, 2009). The two methods have distinct boundary conditions, however. Prediction markets do not rely on representative sampling. Rather, the key conditions for prediction markets are that: (1) individuals have access to some information about others' attitudes; and (2) a mechanism is in place to pressure those with poorer information to participate less actively than those with superior information. In ideal prediction markets, this mechanism is the financial risk taken when betting on an election outcome. Financial incentives are not required, however. So long as those with better information about the election outcome are rewarded for voicing their prediction and those with poorer information are discouraged from doing so, the system as a whole will solicit relatively accurate predictions (Sunstein, 2006).

This condition can be met, to varying degrees, in socio-technical systems. Within social media, for example, individuals are often sensitive to the demands of their audience (Marwick & boyd, 2011). In particular, some topics are only legitimate subjects of conversation in some circles or sub-communities (Cattani, Ferrani, & Allison, 2014). Reputational risk can then play the role that financial risk plays in classic prediction markets, as those who deviate from the prescribed set of topics may lose reputation within or attention from the community. It thus behooves participants to consult knowledge of a topic's status within the audience *as a whole* before taking a risk to venture a claim about it.

The outcome of an upcoming election is likely to play a role in topic selection because likely winners are more relevant to the audience. A topic has high *information utility* when it 'can aid individuals in making future decisions, in political contexts and beyond' (Knobloch-Westerick & Kleinman, 2011, p. 171). Information about candidates likely to win an election has higher utility than information about candidates whose prospects are effectively moot. These candidates can be described as having *political relevance* because they have or are expected to gain power (Schweitzer, 2012).

Political relevance and information utility guide the attention of both audiences and traditional media. Incumbent politicians obtain more media coverage than challengers (Hopmann, de Vreese, & Albaek, 2011; Tresch, 2008), while candidates who are more likely to win their election receive increased scrutiny in the form of tougher questions from

interviewers (Gnisci, Van Dalen, & Di Conza, 2014). In response to learning that a candidate is increasing in public support, individuals construct more thorough evaluations of candidates and build arguments to support their opinion of the candidate (Mutz, 1997).

Information about the likely outcomes of elections is also generally easily accessible due to the media's tendency to provide 'horse-race' coverage (Mutz, 1995). In fact, prediction market participants tend to rely on published polls for their information and thus market predictions tend to lag polls (Jackman, 2015). Thus, when there are strong incentives to behave and speak in a relevant way, many participants in socio-technical systems should be able to discriminate between relevant and moot candidates, while those who cannot are encouraged to keep quiet.

## Normative pressures and the selection of discourse on Wikipedia

### *Wikipedia as fact collector*

Like its contemporary innovations in communication technology, Wikipedia is an open access platform in which anyone may contribute (Anthony, Smith, & Williamson, 2009). However, unlike other socio-technical systems such as Twitter or Facebook, Wikipedia has an explicit purpose: the accumulation of verifiable facts from reliable sources presented in a neutral point of view (Wikipedia, 2014a, 2014b). Every Wikipedia article is thus assembled according to the same set of general rules with the same basic purpose.

Publication on Wikipedia is governed by a strict *gate-keeping* process (Barzilai-Nahon, 2008). Contributors may gather and provide information from a wide variety of sources, but any contribution that does not meet Wikipedia's standards is likely to be edited and removed. Authors who continually violate Wikipedia's norms can have their reputation damaged and, eventually, be locked out of making further contributions to the site. Thus, when Wikipedians contribute to an article, they take a modest but real reputational risk.

The strict enforcement of these norms has led Wikipedia to demonstrate a surprising level of timeliness in updating information in response to current and breaking news events (Keegan, 2013) as well to its becoming a site of active editing of information about political candidates and elections (Brown, 2011; Neff et al., 2013). The site is also commonly used by journalists in compiling their own coverage (Messner & South, 2011). All of this is achieved without a central editor or specific strategy (beyond fact accumulation). Rather, the site is successful because the norms are largely effective, enforced, and followed (Reagle, 2010).

### *What is wiki-worthy? Rules for article creation on Wikipedia*

The first regulatory norm imposed by Wikipedia is the insistence that all information in Wikipedia be *verifiable*. All contributions to Wikipedia should carry a citation to a credible, external source that corroborates the claim's content. This rule discourages frivolous contributions that can overwhelm the system of collective evaluation (Margolin & Monge, 2013). Content contributors must provide research to minimize the effort that other editors must invest to test their assertions.

Wikipedia also imposes an additional burden called the *notability* standard. This standard governs which topics may have a stand-alone article. Notability limits the collective burden of evaluating verifiable but potentially irrelevant information:

> Wikipedia's concept of notability applies this basic standard to avoid indiscriminate inclusion of topics. Article and list topics must be notable, or 'worthy of notice'. Determining notability does not *necessarily* depend on things such as fame, importance, or popularity – although those may enhance the acceptability of a subject that meets the [subject specific] guidelines. (Wikipedia, 2014c)

These two norms – verifiability and notability – put reputational pressure on contributors. Specifically, when a contributor creates a Wikipedia page for a topic, such as a political candidate, they do so with the expectation that others will share their judgment that the topic is notable. This expectation then justifies the contributor's research efforts to make the claims verifiable. When a candidate's notability is questionable, it is not worth finding verifiable information about them because it is very likely that the article will be taken down, anyway. Thus, in a coarse manner, Wikipedia's rules create a 'market' in which reputational risks encourage individuals to 'bet' their research efforts only when they expect that other editors in the community will accept the candidate's notability.

### *Notability, article creation, and election outcomes*

The Wikipedia notability criterion is applied differently for different subject matter. For politicians, those who hold national or regional office are automatically considered notable. By contrast, mere candidates for such offices, as well as holders of local office only, are not (Wikipedia, 2014d). Rather, to become notable, these candidates must be judged as 'famous', 'important' or 'popular' by the community.

Consistent with these rules, the majority of challenger candidates in national elections do not have their own Wikipedia pages prior to the election. Out of 349 challengers from the two major parties in the 2010 US races for the House of Representatives, 194 had never had a Wikipedia article before or as of the day of the election. Nonetheless, at least 54 major party challengers had articles created about them during the 12 months prior to the election. This variation reflects the collective judgment of Wikipedians about each candidate's notability based on the political information environment (Mutz, 1995; Schweitzer, 2012).

Do a candidate's prospects for electoral success play a role in the assessment of notability? According to Wikipedia policy, they should not. The guidelines state 'notability is not temporary', and therefore a candidate's notability should not be contingent on their future success (Wikipedia, 2014c). Nonetheless, the fact that 'fame, importance, or popularity … may enhance the acceptability of a subject [as notable]' suggests that there may be a correlation between notability and a candidate's prospects (Wikipedia, 2014c). Some Wikipedians may also be aware of cases where, based on fundamental factors such as a party's history of success in a district, a candidate is likely to prevail. Creating an article for such a candidate would both serve the public's need for information in the short term and be likely to pass the notability standard in the longer term when the candidate takes office. By contrast, campaigns that have little chance of success will neither appear notable in the short term nor pass the notability standard over time.

## Testing for the prediction market mechanism

These arguments suggest two related propositions: (1) that there will be an association between the creation of Wikipedia articles and electoral outcomes; and (2) that this association will be at least partly attributable to prediction market-like mechanisms in which individuals' behavior is motivated by expectations of election outcomes. First, to test for this association, we describe candidates who have articles created for them during a campaign as being 'wiki-worthy' and formalize our first hypothesis as follows:

> H1: Wiki-worthiness will be associated with electoral success (candidates that have Wikipedia articles created about them during an election campaign are more likely to win elections).

Next, we attempt to isolate the role of expectations in producing this association. There are a number of plausible alternative explanations for an association between wiki-worthiness and

electoral success. First, the causal order may be the opposite of that specified by our model. That is, rather than articles being created because Wikipedians observe that a candidate has a substantial probability of success, the creation of the article during the campaign may itself significantly increase this probability. Another possibility is that the relationship between wiki-worthiness and electoral success is due to some third factor that both increases verifiability or notability and is associated with candidate success. In particular, both attention from national news media and money spent by the campaign should each lead to more information being available about a candidate, increasing verifiability. These factors are also likely to be associated with success (DiGrazia et al., 2013; Erikson & Palfrey, 1998; Grossman & Helpman, 2001). Similarly, whether a candidate has held local office prior to the election may be associated with notability and also with success (Jacobsen, 1989). If expectations for election outcomes are at least partly responsible for the association between wiki-worthiness and candidate success, the relationship should remain significant when controlling for these factors:

H2: Controlling for endogeneity and alternative (non-expectations based) causes, wiki-worthiness will be associated with electoral success.

Another way to address the role of expectations in the process of article creation is to examine when articles are created. The expectations for a candidate's success can vary over the course of the campaign year. When new information, such as new polls or the development of a scandal or controversy, emerges, the projected outcome of an election may be adjusted (Campbell, 2012). If expectations are one of the mechanisms behind article creation, then:

H3: During the campaign year, a Wikipedia article will be more likely to be created for a candidate when expectations for the candidate's success are higher.

The primary aim of our analysis is to explain the mechanism that leads to the association between Wikipedia article creation and electoral success. Nonetheless, to the extent to which there is a consistent relationship between Wikipedia article creation and electoral success, this raises the question of whether including article creation as a variable in prediction models can improve model performance over baseline models. We thus ask:

RQ1: Controlling for other factors, are candidates deemed wiki-worthy more likely to win elections?



## Method

### *Data and measures*

#### *Candidate selection*

Data for the study were drawn from the 2010 US Congressional elections and the 2010 UK parliamentary elections. We restrict the analysis to candidates from 'major' national parties: Democratic and Republican parties in the United States; Liberal Democrats, Conservative, and Labour parties in the United Kingdom. This limits the chances that observed effects are due to a bias against particular small parties. We also restrict our main analysis to 'challenger' candidates. We define a major party 'challenger' as a candidate who is running on behalf of a major party for a seat that is currently occupied by an incumbent individual who is running (again) for that seat ($N = 349$ in United States; $N = 952$ in United Kingdom). We focus on these candidates because their opponent's Wikipedia status is the same across all cases since all incumbents

possess articles prior to the campaign. Special elections that did not occur on Election Day (2 November 2010) in the United States were not included.

### Election variables

The data for the US House races were drawn from the Federal Elections Commission report of election results (FEC, 2010) and include the name, party, incumbency status, and number of votes received for all registered candidates in each race for the House of Representatives in the 112th Congress. The same variables are obtained for the UK parliamentary races from the Guardian's report of election results (UK election results, 2013).

*Vote-share*. The percentage of the total votes in a race that a candidate received.

*Winner*. Whether the candidate received the greatest share of votes in their respective race.

*Prior party vote-share.* The vote-share received by the candidate who ran for the same seat from the challenger's party in the prior election cycle for that district.

### Wikipedia variables

Wikipedia makes data available about its content and users publicly available through the MediaWiki API[1] as well as other public data. Every version of an existing Wikipedia article since 2002 (i.e. those that have not been deleted) is archived along with meta-data including the date and time of any edit to the article. For each candidate with a currently existing (i.e. not deleted) Wikipedia article, the article creation date was used to assign the following dummy codes to each challenger:

*Long-standing page*. Article was created more than 365 days before the election.

*Page created during campaign*. Article was created less than 365 days before the election. This measure excludes deleted pages, for which creation dates are not available.

*Page created during campaign robust*. The same as page created during campaign with the addition of all deleted pages (i.e. assuming that they were all created during this time period).

*Page created month*. The month (from 1 to 12) prior to the election that a page received during the campaign was created (e.g. a page created in December, 2009 = month 2).

*Page-views*. The cumulative number of requests for viewing or editing that a candidate's article received via the Wikipedia website over the last 4 months of the campaign, reflecting the number of times an article is requested for reading by unregistered editors.[2]

### Campaign variables

*Mentions in news media*. To estimate the quantity of verifiable material available to Wikipedia contributors, we count the number of times each candidate is mentioned in the top 25 newspapers by national circulation over the last 4 months of the campaign. These data were collected through the Daylife query APIs.[3] Articles were considered to mention a candidate if the string '[first name] <space>[last name]' was printed in the article along with a reference to the office they were seeking '[House] or [Rep.]' or their political party. This high recall method should minimize underestimation of a candidate's verifiability, thus creating a more rigorous challenge for our hypotheses. The mention counts represent the total number of articles in which the candidate's name appeared in these sources within the specified time range.

*Campaign expenditure*. The total amount of money spent by each candidate's campaign is drawn from the FEC report field TTL_DISB and divided by 1000 for ease of interpretation. Data on the timing of this spending were not available.

*Local office holder*. Whether the candidate has ever held a local elected political office (Jacobson, 1989). These data, provided courtesy of David Rohde, were only available for the US candidates.

*Party*. In US mid-term elections, the performance of candidates from the President's party is often correlated with the president's approval rating (Campbell, 2012). We thus mark each US challenger with a dummy code indicating whether they are in the President's party (Democrat = 1) or the other party (Republican = 0).

*Recent Cook expectations*. The Cook Report[4] publishes expected outcomes for US Congressional races on a weekly basis. Cook creates these expectations based on a complex integration of information available in the political information environment and is known as a reliable representation of expert opinion on likely race outcomes (Campbell, 2012).

The Cook Report categorizes each race into one of eight ordinal categories: Solid Democratic, Likely Democratic, Lean Democratic, Democratic Toss Up, Republican Toss Up, Lean Republican, Likely Republican, and Solid Republican. These scores were transformed to reflect the prospects of success for the challenging candidate, ranging from 'Solid Incumbent' (scored as 1) to 'Solid Challenger' (scored as 8). For the periods observed no challenger ever received a greater than 'Toss Up Challenger' (scored as 5) rating. The last report issued in each month was used to assess each candidate's prospects for success at the beginning of the next month.



## Model specification

To test Hypothesis 2, we run logistic regressions predicting both article creation and election outcomes. In addition to the control variables likely to be associated with Wikipedia's standards, we include two other control variables that are likely to be associated with electoral success: prior vote-share and the challenger's party, as a check of robustness to insure that the observed relationship is not the artifact of some other, less well understood mechanism. To test Hypothesis 3, we employ the Cook Report assessment as an operationalization of expectations in a hazard model for the timing of article creation.

To address RQ1, we include all controls used in the logistic regressions as each is generally a leading indicator of electoral success. To evaluate our prediction models, we use 200 iterations where a random 80% of cases are drawn for training and the remaining 20% are used for testing. We then compare models relying on control variables to those relying on control variables with wiki-worthiness included as a predictor.

Victory in an election is the result of an interdependent, multi-level process that involves all of the candidates in a race. In each of our US elections, there is only one major party challenger in each district, minimizing inter-dependency between cases. In the UK, however, there are three major parties. Thus, each challenger is competing against both an incumbent and, in many cases, another challenger who is in our data. We thus perform our UK analyses for each party separately as well as report results for all parties together. All analyses were performed using R-3.0.1 (R Core Team, 2013). Tables were produced using the stargazer package (Hlavac, 2014).

## Results

Our arguments focus on challengers to incumbents who did not have 'long-standing articles.' Nonetheless, for the sake of completeness, we briefly provide information about candidates with long-standing articles as well as candidates who are running in open seat elections.

Candidates with long-standing articles achieved notability prior to the campaign. This preexisting notability may reflect the candidate's relevance and may correlate with their success, but the creation of the article was not likely due to an expectation of this success. There were 71 of these candidates in the 2010 US election, with 26 (37%) of them eventually winning. By contrast, of the remaining 278 challengers who did not have an article at least one year before the election, only 30 (11%) were eventually victorious, a significant difference ($\chi^2 = 28.0$, df = 1, $p < .001$). Similarly, in the UK, there were 49 candidates with long-standing articles, with 24 (49%) winning their races. By contrast, only 49 (5%) of the 903 candidates without long-standing articles were victorious ($\chi^2 = 109.8$, df = 1, $p < .001$).

Open seat races are those in which there is no incumbent candidate. These races thus contain multiple candidates who may prove wiki-worthy during the campaign. In the United States, in 2010, there were 67 candidates who ran for open seats that did not have a Wikipedia article prior to the campaign year. Among these 67 candidates, 29 received Wikipedia articles during the campaign, 20 (69%) of whom were victorious. By contrast, only 5 (13%) of the 38 who did not receive articles won their race ($\chi^2 = 19.58$, df = 1, $p < .001$). In the United Kingdom, in 2010, there were 427 candidates who ran for open seats that did not have a Wikipedia article prior to the campaign year. Within these 427 candidates, 18 received Wikipedia articles during the campaign, 16 (89%) of whom were victorious. By contrast, only 103 (25%) of the 409 who did not receive articles won their race ($\chi^2 = 34.81$, df = 1, $p < .001$). Though the Conservative party took the majority of these (10 articles with 9 winners), both the Labour (6 articles, 6 winners) and Liberal (3 articles, 2 winner) showed a similar pattern. These results provide an initial indication of the relationship between wiki-worthiness and candidate success.



### *Hypothesis 1*

To test Hypothesis 1, we examined challengers (to incumbents) who lacked Wikipedia articles one year before the election. In the United States, 54 candidates out of these 278 candidates became 'wiki-worthy', that is, it could be confirmed that an article was created about them during the year of the campaign. Of these, 24 (44%) were victorious, compared with only 5 out of the 194 remaining candidates (3%), a significant difference ($\chi^2 = 71.7$, df = 1, $p < .001$).

To be conservative in our analysis, we also include the 30 candidates about whom pages have been deleted. By assuming that these pages were created during the campaign, our analysis treats these overwhelmingly losing candidates as wiki-worthy, contrary to our theorized effect. Nonetheless, the new proportions (25 out of 84 wiki-worthy compared with only 5 out of the 194 with no page) constitute a significant difference ($\chi^2 = 45.0$, df = 1, $p < .001$).

Results for the UK elections are murkier. For articles that could be confirmed to be created during the campaign year, 3 out of 8 candidates (38%) were victorious, compared with only 46 out of 822 candidates (6%) who had no article created during this time (and had no pre-existing article) ($\chi^2 = 14.52$, df = 1, $p < .001$). However, this result was no longer statistically significant when the 73 deleted articles – all of which belonged to losers – were included. If we assume all 73 of these articles were created during the campaign year, candidates about whom articles were created have a lower chance of winning (4% vs. 6%) though the difference is not statistically significant ($\chi^2 = 0.514$, df = 1, $p = .47$). Analysis of the results by party showed that the effect only held for the Conservative party, as neither the Labour party (zero articles created during the campaign year for 196 candidates) nor the Liberal Democrats (two articles created for 379 candidates) had meaningful data. Thus, there is some support for Hypothesis 1 in the UK data, but it is much weaker and relies on very small samples.

### *Hypothesis 2*

*Endogeneity*

Because Wikipedia articles can disseminate information about a candidate, candidates with articles may have a better chance of winning because of the article itself. Wiki-worthy candidates thus may be more successful irrespective of whether the article's creation reflects an expectation of their success. To test for this explanation, we examine whether there is a relationship between the attention received by candidate's Wikipedia article and their success. Candidate success, measured by both eventual victory and eventual vote-share, was regressed on the number of page-views for each candidate's article beginning 120 days prior to the date of the election for which such data could be obtained (using only candidates with pages that existed 120 days prior to the election, including those created a year or more before) ($N = 98$ in US, $N = 49$ in UK). Total page-views are neither positive nor significant predictors of either candidate victory or eventual vote-share, and the R-squared in each regression is less than 1%, suggesting that the article's presence is not responsible for the association between its existence and the candidate's success.

*Common causes*

For Hypothesis 2 to be supported, the relationship between wiki-worthiness (article creation) and success must also hold after controlling for three factors: news media coverage, campaign spending, and whether the candidate held local office. We ran two sets of logistic regressions, one in which the outcome variable was the election success (winner) and one in which the outcome variable was wiki-worthiness (page created campaign year). As expected, when examined on their own, each of our potential common causes shows a significant relationship to both electoral success and wiki-worthiness. However, including these factors does not eliminate the statistical relationship between wiki-worthiness and electoral success. Table 1(a) shows the results for all election outcomes (including those where candidates had long-standing pages). As can be seen in column 4, when controlling for these common causes, wiki-worthy candidates are more likely to win elections. The relationship also holds when additional predictive information (President's party, prior party vote-share) is included. Results, shown in Table 1(b), were similar when predicting article creation (excluding those with long-standing pages). These results, together with the test for endogeneity, support Hypothesis 2.

### *Hypothesis 3*

Hypothesis 3 predicted that during the campaign year, a Wikipedia article will be more likely to be created for a candidate when expectations for the candidate's success are higher.

We test for this association using a hazard model of the likelihood of an article being created in a given month. We include the same predictors as in our full model and then add the Cook Report expectations for the race prior to that month. Table 2 shows these results. In the first column, the significant parameters from the logistic regression model are included only. The second column shows the results when the time varying covariate 'Recent Cook expectations' is included. As before, local office holding ($p < .05$) remains a significant predictor of an article being created, while campaign spending remains very close to significance ($p = .053$). A candidate's eventual victory, however, is no longer significant ($p = .23$). At the same time, the expectations for the candidate immediately prior to that month are a significant indicator of an article being created in the subsequent month. Models using a candidate's eventual vote-share rather than the dichotomous winner variable provided substantively the same result:

Table 1. (a) Electoral success and (b) article creation during campaign year.

| | Dependent variable: | | | | | | |
|---|---|---|---|---|---|---|---|
| | Odds of candidate victory | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| (a) | | | | | | | |
| Mentions in news media | 0.508*** | | | 0.375** | 0.382** | 0.478** | 0.472* |
| | (0.098) | | | (0.142) | (0.143) | (0.185) | (0.187) |
| Campaign expenditure | | 0.001*** | | 0.001*** | 0.001** | 0.0005 | 0.0004 |
| | | (0.0002) | | (0.0002) | (0.0002) | (0.0003) | (0.0003) |
| Local office holder | | | 1.752*** | 1.484*** | 0.993* | 1.530** | 1.352* |
| | | | (0.323) | (0.402) | (0.445) | (0.509) | (0.563) |
| President's party | | | | | −4.076*** | −4.467*** | −4.353*** |
| | | | | | (1.010) | (1.141) | (1.133) |
| Prior party vote-share | | | | | 9.557*** | 11.056*** | 10.623*** |
| | | | | | (2.420) | (2.644) | (2.691) |
| Long-standing page | | | | | | | 0.599 |
| | | | | | | | (0.858) |
| Page created campaign | | | | 1.037* | | 1.578** | 1.935* |
| | | | | (0.410) | | (0.532) | (0.758) |
| Constant | −2.886*** | −2.818*** | −2.144*** | −4.346*** | −6.245*** | −7.754*** | −7.829*** |
| | (0.314) | (0.272) | (0.193) | (0.549) | (0.996) | (1.288) | (1.310) |
| Observations | 349 | 312 | 349 | 284 | 278 | 254 | 254 |
| Log likelihood | −138.275 | −109.914 | −137.819 | −85.059 | −73.067 | −57.454 | −57.206 |
| Akaike inf. crit. | 280.549 | 223.828 | 279.638 | 180.119 | 158.134 | 128.908 | 130.412 |

| | Dependent variable: | | | | |
|---|---|---|---|---|---|
| | Odds that page is created for candidate during campaign | | | | |
| | (1) | (2) | (3) | (4) | (5) |
| (b) | | | | | |
| Mentions in news media | 0.251** | | | 0.047 | 0.005 |
| | (0.089) | | | (0.120) | (0.126) |
| Campaign expenditure | | 0.002*** | | 0.001*** | 0.001** |
| | | (0.0003) | | (0.0004) | (0.0004) |
| Local office holder | | | 1.974*** | 1.661** | 1.611** |
| | | | (0.470) | (0.565) | (0.574) |
| President's party | | | | −0.223 | 0.055 |
| | | | | (0.451) | (0.465) |
| Prior party vote-share | | | | 2.336 | 0.909 |
| | | | | (2.098) | (2.264) |
| Winner | | | | | 1.647* |
| | | | | | (0.756) |
| Constant | −1.274*** | −1.456*** | −1.030*** | −2.282*** | −1.905** |
| | (0.221) | (0.208) | (0.150) | (0.694) | (0.732) |
| Observations | 254 | 219 | 254 | 192 | 192 |
| Log likelihood | −152.188 | −117.216 | −146.228 | −97.683 | −94.976 |
| Akaike inf. crit. | 308.376 | 238.431 | 296.456 | 207.367 | 203.952 |

*$p < .05$.
**$p < .01$.
***$p < .001$.

Table 2. Predicting the creation of a Wikipedia article during campaign year.

| | *Dependent variable:* | |
|---|---|---|
| | Hazard that page is created for candidate during a given month | |
| | (1) | (2) |
| Mentions in news media | 0.080 | 0.047 |
| | (0.098) | (0.102) |
| Campaign expenditure | 0.0003** | 0.0002 |
| | (0.0001) | (0.0001) |
| Local office holder | 1.003** | 0.889* |
| | (0.344) | (0.359) |
| President's party | 0.315 | 0.427 |
| | (0.430) | (0.419) |
| Prior party vote-share | 0.189 | −2.209 |
| | (2.028) | (2.466) |
| Winner | 0.999* | 0.490 |
| | (0.416) | (0.492) |
| Recent cook expectation | | 0.514* |
| | | (0.248) |
| Observations | 2,343 | 2,343 |
| $R^2$ | 0.017 | 0.019 |
| Max. possible $R^2$ | 0.186 | 0.186 |
| Log likelihood | −220.734 | −218.444 |
| Wald test | 45.790*** (df = 6) | 53.810*** (df = 7) |
| LR test | 40.091*** (df = 6) | 44.669*** (df = 7) |
| Score (Logrank) test | 58.245*** (df = 6) | 68.891*** (df = 7) |

*$p < .05$.
**$p < .01$.
***$p < .001$.

recent expectations, as measured by the Cook Report, are a significant predictor of article creation, supporting Hypothesis 3.

### *Research question 1*

Research question 1 asks whether Wikipedia article creation is a useful predictor of upcoming electoral success over and above typical predictors. Improving performance for these elections is difficult because most challenges to incumbents in US Congressional elections are 'uncompetitive', making the result largely known in advance (Campbell, 2012). In our data, one year prior to the election, the Cook Report identified 270 out of our 349 races as cases as 'Solid Incumbent'. Of these 270, 260 (96%) were won by the incumbent. Of the remaining 79 races, which we will call 'competitive', 43 (58%) were won by the challenger. Thus, the bulk of our cases can be accurately predicted without the aid of any control variables simply by picking the incumbent. Thus, to assess the predictive value of Wikipedia article creation, we compared models both for all races and for competitive races only.

Table 3 shows the results of these predictions. The models are compared for mean differences in overall accuracy, true positive rate (the portion of true winners accurately detected by the model), and the true negative rate (the portion of true losers accurately detected by the model). Results show that when predicting all races Wikipedia page creation information does not significantly improve baseline models. For competitive races, Wikipedia page creation information adds

Table 3. Prediction model results.

| Means (and standard errors) of 200 runs | | | |
|---|---|---|---|
| Model | Accuracy | True positive rate | True negative rate |
| *All races* ($N = 349$) | | | |
| Controls | 0.88 | 0.646 | 0.936 |
| | (0.003) | (0.01) | (0.003) |
| Controls + Page created campaign | 0.883 | 0.654 | 0.941 |
| | (0.003) | (0.012) | (0.002) |
| Mean improvement with article creation | 0.003 | 0.008 | 0.005 |
| *Competitive races* ($N = 79$) | | | |
| Controls | 0.792 | 0.936 | 0.571 |
| | (0.007) | (0.006) | (0.015) |
| Controls + Page created campaign | 0.814 | 0.899 | 0.705 |
| | (0.007) | (0.009) | (0.013) |
| Mean improvement with article creation | 0.022* | −0.037** | 0.134*** |

*$p < .05$.
**$p < .01$.
***$p < .001$.

a modest though significant improvement. Accuracy for competitive races is 2% better when wiki-worthiness is included. In particular, while the baseline models of competitive races appear to favor guessing that challengers will win, correctly identifying 94% of winners but only 57% of losers, models that include wiki-worthiness are less sanguine about the prospects of challengers who do not receive articles, correctly identifying slightly fewer winners (90%), but gaining in identifying losers (71%).

## Discussion

### *Review of findings*

Our results show a consistent relationship between a candidate's notability within Wikipedia and expectations for their success in an upcoming election. In both the United States and the United Kingdom, candidates for whom stand-alone Wikipedia pages were created were more likely to win their elections. Further analysis on US data revealed that this effect at least partially reflects *expectations* of a candidate's victory.

The official description of Wikipedia's verifiability and notability standards does not state that candidates that are expected to win an upcoming election meet these standards. Rather, the correlation between wiki-worthiness and electoral success appears to be a matter of how Wikipedia's contributors interpret the standards. Specifically, Wikipedians seem to feel that a candidate with better prospects for success is more 'notable', that is, relevant to Wikipedia's users and mission.

Consistent with other findings for prediction markets (Jackman, 2015), these judgments appear to largely reflect information already observable in the political information environment. Inclusion of Wikipedia article creation does not significantly improve election prediction over baseline fundamentals models when applied to all races, and only modestly improves predictions of competitive races. Rather, the association appears to reflect Wikipedians' responses to the political information environment.

### *Implications for further research*

Our findings suggest that Wiki-worthiness reflects the judgments individuals make about the information other members of the community possess and are willing to act on, rather than the

aggregations of their personal intentions. Is this response specific to Wikipedia or present in other socio-technical systems? While the explicit enforcement of a 'notability standard' is unique to Wikipedia, the theoretical model that predicts the observed association applies more broadly. It is thus plausible that in any socio-technical system, or sub-community within a system where relevance and information utility play an important role, *worthiness* and expected success will be correlated. It is not clear if or when such dynamics generalize to Twitter, Facebook or other systems, but if these dynamics do infiltrate these environments, they will interfere with purely 'poll'-like analyses that assume that individuals' contributions simply reflect their own, personal opinions.

Our analysis does not invalidate the 'poll' model, but rather suggests that it should be complemented with analysis from a 'prediction market' perspective. In many datasets, multiple processes may operate; such mechanisms from a combination of models should be examined. In particular, a more comprehensive approach in which sentiment is measured after controlling for relevance might improve the performance of more poll-like algorithms.

More theoretically, these findings suggest the possibility of interesting self-reinforcing feedback loops within news media and socio-technical systems. In particular, much as the spiral of silence discourages individuals from articulating views they perceive to be unpopular, thus extending the collective perception of this lack of popularity (Noelle-Neiman, 1974), the perception that candidates (or possibly issues) 'can't win' may discourage collective research about them, leading them to be less well recognized and understood, and thus perceived as less legitimate or viable options. This 'spiral of mootness' may subtly act to discourage the productive discussion of political alternatives. Further research might consider whether such spirals are harmful to democratic processes and, if so, how they might be mitigated.

In addition, we find that the wiki-worthiness effect is robust to raw mentions of a candidate in the news media; however, it is not clear if the expectations of relevance are built from news media reports or other sources of information about public opinion. If news media are responsible, it is also not clear if these expectations are conveyed indirectly through framing or directly through the reporting of poll numbers or other explicit statements about candidate prospects. Further research might examine whether expectations of relevance can be 'induced' through different framings of campaign news without explicit mentions of such expectations.

### *Limitations*

First, as in any study of observational data, inferences regarding causality are tendentious. Researchers should thus consider the implications of the claims in this paper for other Wikipedia datasets, other socio-technical systems, and conventional news media and how they might be tested within them.

Second, the inferences drawn with respect to alternative explanations, particularly with regard to mentions in the news media and campaign spending, are merely suggestive. These measures are only approximations of the source material available for verification on a political candidate. If more precisely measured, differences in the kind of national media coverage, or in the form and extent of local media coverage, might further refine the interpretation of the observed effect.

In addition, as we have argued, one of the benefits of using a theoretical model as a basis for predictions is the ability to use the theory to help identify relevant boundary conditions. In this case, we note that in many political systems, individuals cast votes for parties as a whole, with candidates officially selected only after the election is complete and the party's seats are allocated. In these cases, our mechanism will have limited applicability, because individuals' expectations about election outcomes cannot be directed toward particular, individual candidates.

Furthermore, Wikipedia is an evolving socio-technical system. The inferences drawn in this paper are based on the fact that in 2010, Wikipedia allowed its editors and participants to freely choose whether to create articles for challengers in the US and UK elections. Changes or 'drifts' in Wikipedia norms or the extent of its use might shift these relationships (Lazer et al., 2014).

**Acknowledgement**

The authors would like to thank David Rohde for his generosity in supplying data on candidate quality.

**Funding**

This initial conceptualization of this research and the analysis of the data from the United Kingdom was funded by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D12PC00285. The analysis of the US data was supported by Northeastern University. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

**Disclosure statement**

No potential conflict of interest was reported by the authors. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBE, or the US Government.

**Notes**

1. http://en.wikipedia.org/w/api.php, https://www.mediawiki.org/wiki/API:Main_page.
2. http://dumps.wikimedia.org/other/pagecounts-raw/.
3. http://developer.daylife.com/docs.
4. http://cookpolitical.com/house/charts/race-ratings.

**Notes on contributors**

Drew Margolin is an assistant professor in the Department of Communication at Cornell University. His research focuses on the mechanisms that shape the production of individual and institutional discourse on the Internet and social media.

Sasha Goodman is a post-doctoral research fellow in the Political Science Department at Northeastern University and an affiliate of the Harvard Institute for Quantitative Social Science. His research focuses on organizational theory and political organizations.

Brian Keegan is a post-doctoral research fellow at in the Political Science Department at Northeastern University and an affiliate of the Harvard Institute of Quantitative Social Science. He researches social networks and online collaboration using computational social science.

Yu-Ru Lin is an assistant professor in the School of Information Sciences at the University of Pittsburgh. Her work focuses on social and political networks as well as computational and visualization methods for understanding network data.

David Lazer is a professor in the Political Science Department and the College of Computing and Information Science at Northeastern University and a Visiting Scholar at the Harvard Kennedy School. He works in the intersection of computational social science, political science, and social networks.

## References


Anthony, D., Smith, S. W., & Williamson, T. (2009). Reputation and reliability in collective goods: The case of the online encyclopedia Wikipedia. *Rationality and Society*, *21*(3), 283–306. doi:10.1177/1043463109336804

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., … Zitzewitz, E. (2008). The promise of prediction markets. *Science*, *320*(5878), 877–878. doi:10.1126/science.1157679

Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American Society for Information Science and Technology*, *59*(9), 1493–1512. doi:10.1002/asi.20857

Berg, J., Nelson, F., & Rietz, T. (2007). Prediction market accuracy in the long run. *International Journal of Forecasting*, *24*(2), 285–300.

Brown, A. R. (2011). Wikipedia as a data source for political scientists: Accuracy and completeness of coverage. *PS: Political Science & Politics*, *44*(02), 339–343. doi:10.1017/S1049096511000199

Campbell, J. E. (2012). Forecasting the presidential and congressional elections of 2012: The trial-heat and the seats-in-trouble models. *PS: Political Science & Politics*, *45*(04), 630–634. doi:10.1017/S104909651200090X

Cattani, G., Ferriani, S., & Allison, P. D. (2014). Insiders, outsiders, and the struggle for consecration in cultural fields: A core-periphery perspective. *American Sociological Review*, *79*(2), 258–281. doi:10.1177/0003122414520960

DiGrazia, J., McKelvey, K., Bollen, J., & Rojas, F. (2013). More tweets, more votes: Social media as a quantitative indicator of political behavior. *PLoS ONE*, *8*(11), e79449. doi:10.1371/journal.pone.0079449

Erikson, R. S., & Palfrey, T. R. (1998). Campaign spending and incumbency: An alternative simultaneous equations approach. *The Journal of Politics*, *60*(02), 355–373.

FEC. (2010). *Federal Elections 2010: Election results for the US Senate and the US house of representatives*. Retrieved December 19, 2012, from http://www.fec.gov/pubrec/fe2010/federalelections2010.shtml

Gayo-Avello, D., Metaxas, P. T., & Mustafaraj, E. (2011). *Limits of electoral predictions using twitter*. ICWSM. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewPDFInterstitial/2862/3254

Gnisci, A., Van Dalen, A., & Di Conza, A. (2014). Interviews in a polarized television market: The Anglo-American watchdog model put to the test. *Political Communication*, *31*(1), 112–130. doi:10.1080/10584609.2012.747190

Grossman, G. M., & Helpman, E. (2001). *Special interest politics*. Cambridge, MA: MIT Press.

Hayes, A. F. (2005). *Statistical methods for communication science*. Mahwah, NJ: Lawrence Erlbaum.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society*, *47*, 153–161.

Hlavac, M. (2014). *Stargazer: LaTeX code and ASCII text for well-formatted regression and summary statistics tables. R package version 5.0*. Retrieved from http://CRAN.R-project.org/package=stargazer

Hopmann, D. N., de Vreese, C. H., & Albaek, E. (2011). Incumbency bonus in election news coverage explained: The logics of political power and the media market. *Journal of Communication*, *61*(2), 264–282. doi:10.1111/j.1460-2466.2011.01540.x

Jackman, S. (2015). All that glitters: The betting markets and the 2013 Australian Federal election. To appear. In C. Johnson & J. Wanna (Eds.), *The Abbott's gambit: The 2013 Australian federal election* (pp. 143–159). Canberra: Australian National University Press.

Jacobson, G. C. (1989). Strategic politicians and the dynamics of US House elections, 1946–1986. *The American Political Science Review*, *83*(3), 773–893.

Jungherr, A., Jurgens, P., & 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*, *30*(2), 229–234. doi:10.1177/0894439311404119

Keegan, B. C. (2013). A history of news work on Wikipedia. In *Proceedings of ninth international symposium on Wikis and open collaboration*. Hong Kong, China. New York, NY: ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2491062

Knobloch-Westerwick, S., & Kleinman, S. B. (2011). Preelection selective exposure: Confirmation bias versus informational utility. *Communication Research*, *39*(2), 170–193. doi:10.1177/0093650211400597

Lazer, D. M., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google flu: Traps in big data analysis. *Science*, *343*(6176), 1203–1205.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., … Van Alstyne, M. (2009). Computational social science. *Science*, *323*(5915), 721–723.
Lin, Y., Margolin, D., Keegan, B., & Lazer, D. (2013). *Voices of victory: A computational focus group framework for tracking opinion shift in real time*. Proceedings of ACM conference on world wide web (WWW) conference, Rio.
Margolin, D. B., & Monge, P. R. (2013). Conceptual retention in epistemic communities. In P. Moy (Ed.), *Communication and community* (pp. 1–22). New York, NY: Hampton Press.
Marwick, A., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*, 114–113.
McKelvey, K., DiGrazia, J., & Rojas, F. (2014). Twitter publics: How online political communities signaled electoral outcomes in the 2010 US house election. *Information, Communication & Society*, *17*(4), 436–450. doi:10.1080/1369118X.2014.892149
Messner, M., & South, J. (2011). Legitimizing Wikipedia: How US national newspapers frame and use the online encyclopedia in their coverage. *Journalism Practice*, *5*(2), 145–160.
Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. New York: Cambridge University Press.
Mutz, D. C. (1995). Effects of horse-race coverage on campaign coffers: Strategic contributing in presidential primaries. *The Journal of Politics*, *57*(04), 1015. doi:10.2307/2960400
Mutz, D. C. (1997). Mechanisms of momentum: Does thinking make it so? *The Journal of Politics*, *59*(01), 104. doi:10.2307/2998217
Nannestad, P., & Paldam, M. (1994). The VP-function: A survey of the literature on vote and popularity functions after 25 years. *Public Choice*, *79*(3–4), 213–245.
Neff, J. J., Laniado, D., Kappler, K. E., Volkovich, Y., Aragón, P., & Kaltenbrunner, A. (2013). Jointly they edit: Examining the impact of community identification on political interaction in Wikipedia. *PLoS ONE*, *8*(4), e60584. doi:10.1371/journal.pone.0060584
Noelle-Neiman, E. (1974). The spiral of silence: A theory of public opinion. *Journal of Communication*, *24* (2), 44–51.
O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, *11*, 122–129.
R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/
Reagle, J. M. (2010). *Good faith collaboration: The culture of Wikipedia*. Cambridge, MA: MIT Press.
Rothschild, D. (2009). Forecasting elections comparing prediction markets, polls, and their biases. *Public Opinion Quarterly*, *73*(5), 895–916.
Schweitzer, E. J. (2012). The mediatization of e-campaigning: Evidence from german party websites in state, national, and European parliamentary elections 2002–2009. *Journal of Computer-Mediated Communication*, *17*(3), 283–302. doi:10.1111/j.1083-6101.2012.01577.x
Sunstein, C. R. (2006). *Infotopia: How many minds produce knowledge*. New York: Oxford University Press.
Tresch, A. (2008). Politicians in the media: Determinants of legislators' presence and prominence in Swiss newspapers. *The International Journal of Press/Politics*, *14*(1), 67–90. doi:10.1177/1940161208323266
Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. (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). Menlo Park, CA: The AAAI Press.
UK election results: Data for every candidate in every seat. (2013). *The Guardian*. Retrieved March 6, 2013, from http://www.guardian.co.uk/news/datablog/2010/may/07/uk-election-results-data-candidates-seats
Unankard, S., Li, X., Sharaf, M., Zhong, J., & Li, X. (2014). Predicting elections from social networks based on sub-event detection and sentiment analysis. In *Web information systems engineering–WISE 2014* (pp. 1–16). New York: Springer.
Wikipedia. (2014a). Retrieved April 14, 2014, from https://en.wikipedia.org/wiki/Wikipedia:Introduction
Wikipedia. (2014b). Retrieved April 14, 2014, from https://en.wikipedia.org/wiki/Wikipedia:Five_pillars
Wikipedia. (2014c). Retrieved April 14, 2014, from http://en.wikipedia.org/wiki/Wikipedia:Notability
Wikipedia. (2014d). Retrieved April 14, 2014, from http://en.wikipedia.org/wiki/Wikipedia:Notability_(people)#Politicians