MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-761

Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter

Xujuan Zhou[a], Enrico Coiera[a], Guy Tsafnat[a], Diana Arachi[a], Mei-Sing Ong[a,b], Adam G. Dunn[a]

[a] *Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University NSW, Australia*
[b] *Children's Hospital Informatics Program, Boston Children's Hospital, Boston MA, United States*

Abstract

The manner in which people preferentially interact with others like themselves suggests that information about social connections may be useful in the surveillance of opinions for public health purposes. We examined if social connection information from tweets about human papillomavirus (HPV) vaccines could be used to train classifiers that identify anti-vaccine opinions. From 42,533 tweets posted between October 2013 and March 2014, 2,098 were sampled at random and two investigators independently identified anti-vaccine opinions. Machine learning methods were used to train classifiers using the first three months of data, including content (8,261 text fragments) and social connections (10,758 relationships). Connection-based classifiers performed similarly to content-based classifiers on the first three months of training data, and performed more consistently than content-based classifiers on test data from the subsequent three months. The most accurate classifier achieved an accuracy of 88.6% on the test data set, and used only social connection features. Information about how people are connected, rather than what they write, may be useful for improving public health surveillance methods on Twitter.

Keywords:

Machine learning; Social media; HPV vaccines; Public health surveillance; Twitter messaging; Text mining.

Introduction

Social media surveillance applications that provide value to public health include surveying demographics, estimating population-wide sentiment about public health issues like vaccines, forecasting influenza outbreaks, and producing spatial indicators of language, behaviour, and mood [1-7]. One of the specific problems associated with using Twitter for online surveillance is the brevity and non-standard text structures of Twitter posts (tweets), which limit the text fragments that can be used to train classifiers, and may limit performance [8, 9].

We hypothesized that connections between users on social media may help to improve surveillance methods for the following reasons: (a) homophily – where people tend to form connections with others who share similar attributes or opinions [10-12]; (b) contagion of opinions – where social connections represent the conduits through which information flows, influencing and shaping opinions [13-15]; and (c) temporal dynamics – where user relationships may change more slowly than the content in the topics being discussed.

To test the hypothesis that social connections could improve the performance of opinion classification methods, we considered a classification task in the surveillance of anti-vaccine rhetoric about human papillomavirus (HPV) vaccines on Twitter. The growth of anti-vaccine rhetoric in the media is an international problem [16, 17]. HPV vaccines are a relatively recent introduction to the armament of public health, and uptake is highly variable by country, demographic, and location [18]. Anti-vaccine rhetoric specifically directed at HPV vaccines is present in media articles and websites [19-21], and this appears to have the capacity to alter vaccine acceptance and decision-making [22].

The aim of this study was to determine if information about social connections could be used to improve the performance of classifiers intended for ongoing use in public health surveillance, using anti-vaccine rhetoric as an example.

Methods

Study Data

English-language tweets (42,533 tweets) containing keywords related to HPV vaccines were collected between October 1, 2013 and March 31, 2014. We identified these tweets by searching for combinations of keywords (HPV, vaccine, Gardasil, Cervarix, vaccination, cervical, cancer) via repeated calls to the Twitter application programming interface (API), in accordance with the terms of service for Twitter developers. For each of the users responsible for the tweets (21,166 users), the sets of users they followed (*sources*), as well as the sets of users that followed them (*followers*), were accessed through separate API requests and recorded soon after the first time they tweeted about HPV vaccines in the six-month period.

We split the six months of data into two distinct but contiguous three-month periods and randomly sampled tweets for use in the classifier training (1050 tweets from October 2013 to December 2013) and testing (1100 tweets from January 2014 to March 2014). Two investigators (DA and AD) independently rated each tweet as having presented an anti-vaccine opinion or otherwise. Agreement between the investigators was 95% in the training period (Cohen's kappa 0.88; $p<0.001$), and 95% in the testing period (Cohen's kappa 0.86; $p<0.001$). Disagreements were resolved by discussion. Tweets were removed if they were deleted or the user had become protected or suspended, or if the text was identical after pre-processing. Final samples used included 884 and 907 tweets in the training and testing period, of which 247 (28%) and 201 (22%) were labelled as anti-vaccine, respectively.

Data pre-processing

We pre-processed the text (content) to remove punctuation, words that were unlikely to confer meaning (e.g. 'and'), and other non-word elements (e.g. URLs). The remaining text was used to produce sets of unigrams (one word) and bigrams (two contiguous words), which were then available for use in the classifier training. An example is given in Figure 1.

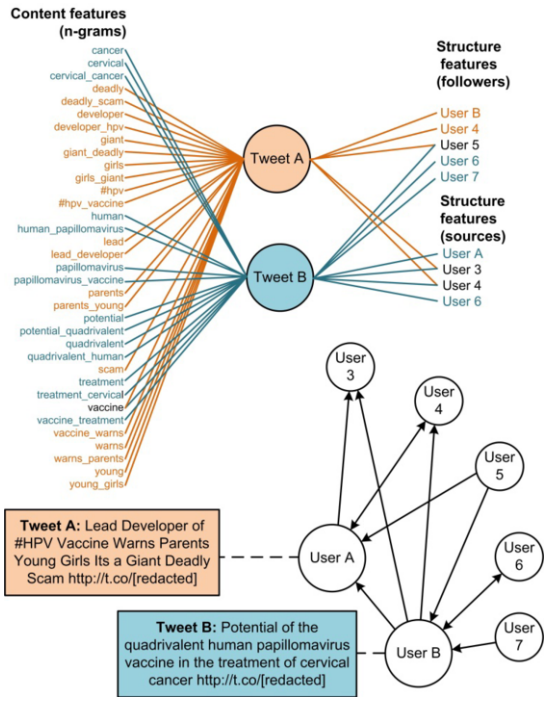

Figure 1– The text is decomposed into n-gram features (left). The follower network for the two users posting the tweets is decomposed into source and follower features (right)

We then determined the source and follower relationships among the set of 21,166 users who tweeted at least once about HPV vaccines. Social connection features were constructed directly from the follower relationships between users. An example of the decomposition is given in Figure 1.

Statistical analysis of content and connection features

Using the sample from the training period, we identified content and connection features that were significantly over-represented in one of the two classes by applying Fisher's exact tests to each feature and then a Bonferroni correction. To examine how low-frequency features might affect the performance of classifiers in the training and testing periods, we also relaxed the inclusion criteria to create alternative sets of features to be used as inputs to the classifier training. The first included all features for which the p-values were less than 0.05, and the second included the set of all features represented in at least three tweets in the sample.

The frequencies of the features exhibiting significant associations in period one were then compared to their frequencies in period two, to measure how the associations may degrade over time. For each of the sets of significant features, we calculated the proportion of features that retained a significant over-representation in the same class.

Classification algorithms

To demonstrate how the temporal variation in the content and connection information might affect the performance of supervised machine learning classifiers, we demonstrated the approach by constructing classifiers using support vector machine (SVM) methods. SVM classifiers have been widely applied in text-based classification [23, 24], and sentiment analysis [25, 26], and are considered the standard and the most appropriate classifier for unbalanced datasets and a large number of features. We chose to apply radial basis function kernels [27], and used consistent parameter values across all the classifiers in order to avoid retrospectively influencing the reporting of the performance.

Feature selection methods are heuristics that are used in the training of machine learning classifier to improve performance by ignoring features that are not useful, and including combinations of features that are best able to discriminate between classes. We applied a hybrid method of forward selection and backward elimination [28, 29].

Classifier construction and testing in the training period was undertaken using a ten-fold cross validation. Note that we determined the potential features using the statistical analysis covering the entire training period. In the testing period (the subsequent three months), the classifiers were applied directly to the full set of tweets from the period as a holdout set, in order to examine the temporal degradation. To compare the classifier performance from training to testing periods, we calculated the standard performance measures: precision, recall, accuracy, and F_1-score. The research project was approved by the Human Ethics Advisory Panels of The University of New South Wales and Macquarie University.

Results

Temporal degradation in content and connection features

From the set of 42,533 unique English-language tweets spanning six months, a random sample of 2150 were extracted and then manually labelled as anti-vaccine or otherwise. After pre-processing, 884 tweets remained in the sample in the first six months (training period), and these came from 877 unique users. Applying Fisher's exact tests and Bonferroni corrections, we identified text fragments and social connections that were found to be significantly more frequent in one class relative to the other (Table 1).

Table 1– The frequency of content and connection features compared across the two periods

Characteristic Set	Number of unique tweets	Number of anti-vaccine tweets (%)	Number of significant features	Features that were no longer significant (%)	Features that switched direction (%)	Features that were still significant (%)
Bigrams (content)						
Bonferroni-corrected	884	247 (28%)	25	24 (96%)	0	1 (4%)
p-value <0.05	884	247 (28%)	232	228 (98%)	2 (0.86%)	2 (0.86%)
Followers (connections)						
Bonferroni-corrected	877	220 (25%)	73	0	0	73 (100%)
p-value <0.05	877	220 (25%)	542	220 (41%)	0	322 (59%)
Sources (connections)						
Bonferroni-corrected	877	220 (25%)	82	2 (2.5%)	0	80 (98%)
p-value <0.05	877	220 (25%)	494	183 (37%)	0	311 (63%)

When the same features were then compared across classes in the tweets from the testing period, the comparison showed that only 1 of 24 (4%) of the content features were also significant in the subsequent three months (Figure 2). In comparison, 80 of 82 (98%) of the connection-based source features were also significant in the subsequent three months, as well as 73 of the 73 (100%) of the connection-based follower features (Table 1). The results show that very few text-based features retained their significant differences in the testing period, while social connections nearly always retained their significant differences in the testing period.

It might be expected that the reason why connection features are stable from one period to the next is because the same users are responsible for anti-vaccine tweets in both periods. However, among the users in the tweets sampled from the training period (877 users), and the testing period (797 users), only 4.1% of the users (66 of 1,608) appear in both samples. Extending this analysis to consider all original tweets in the two periods and not just the sampled sets, only 11.3% of users (2,382 of 21,166) posted tweets about HPV vaccines in both periods. The small overlap suggests that the connection features were stable across the two periods not as a consequence of tweets being posted by the same users, but because users posting about HPV vaccines for the first time in the six month period often followed the same accounts as other users who expressed similar opinions.

Classifier training and testing in period one

The classifiers trained using only connection features produced similar levels of accuracy (often with higher precision and lower recall) to the classifiers that were trained using only content features (Table 2). The best-performing classifier that only used connection features achieved an accuracy of 89.4% (95% CI 87.4-91.4), which was roughly equivalent to the best-performing classifier trained using only content features (89.8% accuracy; 95% CI 87.9-91.8). The overall best-performing classifier in the training period was constructed from both content and connection features (94.4% accuracy; 95% CI 93.1-96.3), and used 23 social connections and 28 text-based features.

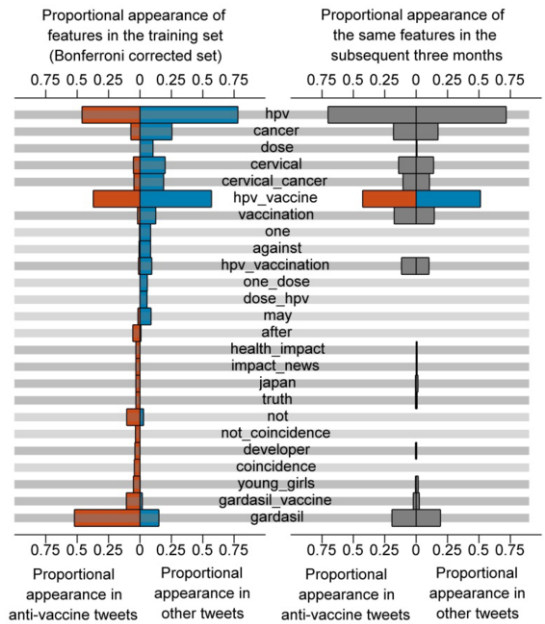

Figure 2– The proportional appearance of text fragments from the Bonferroni-corrected set of content features from the first three months (left), and the subsequent three months (right). Features with non-significant differences in the testing period are illustrated in grey

The performances of the classifiers that were constructed from connection-based source features were slightly better than classifiers from connection-based follower features. The accuracies of classifiers that selected from sources (86.2% to 88.0%) were slightly higher than their direct counterparts that were selected from followers (84.0% to 87.1%). The complete results are given in Table 2.

Table 2– The performances of classifiers trained to classify anti-vaccine tweets within the training period (the first three months)

Input Feature Set	Features selected	Precision	Recall	F1-score	Accuracy (95% CI)
Content: bigrams					
Bonferroni correction	11	0.74	0.56	0.63	82.0 (79.5-84.5)
p-value < 0.05	26	0.77	0.70	0.73	85.8 (83.5-88.1)
threshold = 3	37	0.88	0.74	0.82	89.8 (87.9-91.8)
Connections: followers					
Bonferroni correction	13	0.87	0.44	0.57	84.0 (81.6-86.4)
p-value < 0.05	21	0.89	0.48	0.62	85.5 (83.2-87.8)
threshold = 3	36	0.91	0.55	0.68	87.1 (84.9-89.3)
Connections: sources					
Bonferroni correction	18	0.88	0.55	0.67	86.7 (84.2-89.3)
p-value < 0.05	13	0.88	0.53	0.65	86.2 (83.9-88.5)
threshold = 3	28	0.88	0.60	0.71	88.0 (86.0-90.0)
Connections: followers, sources					
Bonferroni correction	23	0.86	0.57	0.68	87.0 (84.8-89.2)
p-value < 0.05	33	0.88	0.65	0.74	89.0 (86.9-91.1)
threshold = 3	39	0.88	0.67	0.76	89.4 (87.4-91.4)
Combined: bigrams, sources					
Bonferroni correction	17	0.86	0.63	0.72	87.8 (85.5-90.1)
p-value < 0.05	38	0.90	0.82	0.86	93.1 (91.3-94.9)
threshold = 3	42	0.91	0.84	0.87	93.8 (92.1-95.5)
Combined: bigrams, followers, sources					
Bonferroni correction	24	0.91	0.64	0.74	88.9 (86.7-91.1)
p-value < 0.05	51	0.94	0.84	0.88	94.4 (92.8-96.0)
threshold = 3	47	0.94	0.85	0.89	94.7 (93.1-96.3)

Classifier testing in period two

The performance of the classifiers was not sustained in the testing period, and the performance degradation observed from the training period to the testing period varied substantially across the classifiers (Table 3). The classifiers that included content features and had the highest accuracies in the training period exhibited the greatest degradation in performance when tested on tweets from the testing period.

Classifiers that used social connection information tended to perform similarly in the training and testing periods, with smaller changes in accuracy compared to the content-based classifiers (Table 3). These results are consistent with the statistical analysis of the features, which showed that the social connections were more consistently distributed across the two classes in the training and testing periods, compared to the text fragments.

Table 3 – The change in performance when applying the classifiers to the testing period (the subsequent three months)

Classifier	Accuracy (95% CI)	Accuracy change (%)
Bigrams (content)		
Bonferroni correction	85.2 (82.9-87.5)	3.2
p-value < 0.05	82.6 (80.1-85.1)	-3.2
threshold = 3	53.6 (50.4-56.9)	-36.2
Followers (connections)		
Bonferroni correction	86.0 (83.6-88.4)	2.0
p-value < 0.05	85.5 (83.1-87.9)	0.0
threshold = 3	84.1 (81.6-86.6)	-3.0
Sources (connections)		
Bonferroni correction	88.6 (86.4-90.8)	1.9
p-value < 0.05	81.6 (78.9-84.3)	-4.6
threshold = 3	87.3 (85.0-89.6)	-0.7
Bigrams, sources (both)		
Bonferroni correction	88.6 (86.4-90.8)	0.8
p-value < 0.05	82.2 (79.5-84.9)	-10.9
threshold = 3	87.1 (84.8-89.4)	-6.7

The two best performing classifiers were capable of distinguishing anti-vaccine tweets from all other tweets with 88.6% accuracy in the testing period. One was trained using only social connections and the other was trained using social connections and text fragments.

Discussion

We demonstrated that social connection information can be used to improve classifiers capable of identifying anti-vaccine opinions for HPV vaccines on Twitter, addressing the temporal degradation associated with using content features alone. While we have examined this phenomenon for only one topic, the results suggest that this approach may help to reduce the frequency at which social media surveillance systems would need to be updated through manual intervention.

Previous attempts at using social network information as features in supervised machine learning for Twitter classification have demonstrated reasonable performance – the best reported accuracies on various datasets were 68% and 73% using information from replies and retweets [30, 31], and between 58% and 77% using follower connections [32-34]. We believe our study is the first to demonstrate the difference in temporal degradation across classifiers constructed from content and social connection features.

The results suggest that information about who users follow, rather than who follows them, may be more useful for predicting the direction of their expressed opinions. A plausible explanation for this comes from the friendship paradox [35]. For any given user posting a tweet about HPV vaccines, the users they follow are likely to have more followers on average. More popular and influential users are expected to be better connected to the communities that tweet about HPV vaccines and as a consequence, may produce more useful features. The results may also suggest that there is a core of users that may be influential in vaccine information communities and that their followers tend to express similar opinions as a consequence of homophily or contagion [12, 14].

Limitations

There were several important limitations to the experiments reported here. Firstly, rate-limited access to Twitter precluded the instantaneous collection of user information each time we captured a relevant tweet, so calls to the API were staggered throughout the period and the information was collected only once for each user. However, given the stability of the social connections and the relatively small proportion of users that tweeted in both periods, this limitation is unlikely to have affected the conclusions. In addition, we did not apply any query expansion methods or evaluate the overall quality of the search terms we used.

Secondly, alternative feature space constructions and selection methods could have been chosen to produce classifiers, and these may have yielded different results. Specifically, there may be combinations of time-invariant text fragments that could out-perform our most accurate classifier (88.6% accuracy).

Finally – and importantly – we prospectively chose to demonstrate the results of the statistical analysis by implementing one type of classifier (SVMs using a radial basis function kernel) and fixed the parameter values to balance precision and recall in an unbalanced sample. If we had chosen other parameter values, different kernels, or other less appropriate machine learning algorithms, the results may have been different. However, since we tested the significant associations for all content and connection features independently of the classifier training and testing, our conclusions are largely independent of, but confirmed by, the construction of the classifiers.

Conclusion

For the task of classifying tweets about HPV vaccines as anti-vaccine or otherwise, information about the social connections between users provided a useful addition to the content of what people write. In particular, we showed that it is possible to use information about the users that people follow online to help predict their opinions. Our findings also suggest some potential avenues for the development of opinion forecasting – prospectively predicting the opinions of individuals and populations based on their social connections, rather than reactively classifying their opinions based on what they write.

References

[1] Burger JD, Henderson J, Kim G, and Zarrella G. Discriminating gender on Twitter, In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Edinburgh, United Kingdom, 2011; pp. 1301-9.

[2] Salathé M, and Khandelwal S. Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. PLoS Comput Biol 2011: 7: e1002199.

[3] Signorini A, Segre AM, and Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. During the influenza A H1N1 pandemic. PLoS ONE 2011: 6: e19467.
[4] Collier N, Son N, and Nguyen N. OMG U got flu? Analysis of shared health messages for bio-surveillance. J Biomed Semantics 2011: 2(S9).
[5] Chew C, and Eysenbach G. Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. PLoS ONE 2010: 5: e14118.
[6] Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, and Vespignani A. The Twitter of Babel: mapping world languages through microblogging platforms. PLoS ONE 2013: 8: e61981.
[7] Dodds PS, Harris KD, Kloumann IM, Bliss CA, and Danforth CM. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. PLoS ONE 2011: 6: e26752.
[8] Saif H, He Y, and Alani H. Alleviating data sparsity for Twitter sentiment analysis. In: 2nd Workshop on Making Sense of Microposts, 21st International Conference on the World Wide Web. Lyon, France, 2012; pp. 2-9.
[9] Zhang K, Cheng Y, Xie Y, Honbo D, Agrawal A, Palsetia D, Lee K, Liao W-K, and Choudary A. SES: Sentiment elicitation system for social media data. In: 11th International Conference on Data Mining Workshops. 2011; pp. 129-136.
[10] Coiera E. Social networks, social media, and social diseases. BMJ 2013: 346: f3007.
[11] Centola D. An experimental study of homophily in the adoption of health behavior. Science 2011: 334: 1269-72.
[12] McPherson M, Smith-Lovin L, and Cook JM. Birds of a feather: homophily in social networks. Ann Rev Sociology 2001: 27: 415-44.
[13] Dietz K, Epidemics and rumours: a survey. J R Stat Soc 1967: 130: 505-28.
[14] Iyengar R, Van den Bulte C, and Valente TW. Opinion leadership and social contagion in new product diffusion. Marketing Science 2011: 30: 195-212.
[15] Coleman J, Katz E, and Menzel H. The diffusion of an innovation among physicians. Sociometry 1957: 20: 253-70.
[16] Leask J. Target the fence-sitters. Nature 2011: 473: 443-5.
[17] Gangarosa EJ, Galazka AM, Wolfe CR, Phillips LM, Gangarosa RE, Miller E, and Chen RT. Impact of anti-vaccine movements on pertussis control: the untold story. Lancet 1998: 351: 356-61.
[18] Fisher H, Trotter CL, Audrey S, MacDonald-Wallis K, and Hickman M. Inequalities in the uptake of human papillomavirus vaccination: A systematic review and meta-analysis. Int J Epidemiol 2013: 42: 896-908.
[19] Larson HJ, Smith D, Paterson P, Cumming M, Eckersberger E, Freifeld CC, Ghinai I, Jarrett C, Paushter L, Brownstein JS, and Madoff LC. Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyse public concerns about vaccines. Lancet Infect Dis 2013: 13: 606-13.
[20] Madden K, Nan X, Briones R, and Waks L. Sorting through search results: a content analysis of HPV vaccine information online. Vaccine 2012: 30: 3741-6.
[21] Mahoney ML, Tang T, Ji K, and Ulrich-Schad J. The digital distribution of public health news surrounding the human papillomavirus vaccination: a longitudinal infodemiology study. JMIR Public Health Surveill 2015: 1: e2.
[22] Sotiriadis A, Dagklis T, Siamanta V, Chatzigeorgiou K, and Agorastos T. Increasing fear of adverse effects drops intention to vaccinate after the introduction of prophylactic hpv vaccine. Arch Gynecol Obstet 2012;285:1719-24.
[23] Joachims T. Transductive inference for text classification using support vector machines. In: Proceedings of the 16th International Conference on Machine Learning. Bled, Slovenia, 1999; pp. 200-9.
[24] Sebastiani F. Machine learning in automated text categorization. ACM Comput Surv 2002: 34: 1-47.
[25] Pang B, Lee L, and Vaithyanathan S. Thumbs up?: Sentiment classification using machine learning techniques. In: Proceedings of the 7th Conference on Empirical Methods in Natural Language Processing. Philadelphia, United States, 2002; pp. 79-86.
[26] Mullen T, and Collier N. Sentiment analysis using support vector machines with diverse information sources. In: Proceedings of the 9th Conference on Empirical Methods in Natural Language Processing. Barcelona, Spain, 2004; pp. 412-8.
[27] Keerthi SS, and Lin C-J. Asymptotic behaviors of support vector machines with gaussian kernel. Neural Comput 2003: 15: 1667-89.
[28] Kohavi R, and John GH. Wrappers for feature subset selection. Artificial Intelligence 1997: 97: 273-324.
[29] Guyon I, and Elisseeff A. An introduction to variable and feature selection. J Machine Learning Res 2003: 3: 1157-82.
[30] Jiang L, Yu M, Zhou M, Liu X, and Zhao T. Target-dependent twitter sentiment classification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland, United States, 2011; pp. 151-60.
[31] Boutet A, Kim H, and Yoneki E. What's in your tweets? I know who you supported in the UK 2010 general election. In: Proceedings of the 6th International AAAI Conference on Weblogs and Social Media. Dublin, Ireland, 2012; pp. 411-4.
[32] Speriosu M, Sudan N, Upadhyay S, and Baldridge J. Twitter polarity classification with label propagation over lexical links and the follower graph. In: Proceedings of the 1st Workshop on Unsupervised Learning in NLP, 16th Conference on Empirical Methods in Natural Language Processing. Edinburgh, Scotland, 2011; pp. 53-63.
[33] Hu X, Tang L, Tang J, and Liu H. Exploiting social relations for sentiment analysis in microblogging. In: Proceedings of the 6th ACM International Conference on Web Search and Data Mining. Rome, Italy, 2013; pp. 537-46.
[34] Tan C, Lee L, Tang J, Jiang L, Zhou M, and Li P. User-level sentiment analysis incorporating social networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Diego, United States, 2011; pp. 1397-1405.
[35] Feld SL. Why your friends have more friends than you do. Am J Sociol 1991: 96: 1464-77.

Address for correspondence

Xujuan Zhou and Adam G. Dunn, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, 2109, NSW, Australia. Emails: susan.zhou@mq.edu.au; adam.dunn@mq.edu.au