Measuring Network Structure Metrics as a Proxy for Socio-political Activity in Social Media

Anonymous

***Abstract*—Social surveys have been used by researchers and policy makers as an essential tool for understanding social and political activities in society. Social media has introduced a new way of capturing data from large numbers of people. Unlike surveys, social media deliver data more rapidly and cheaply. In this paper, we aim to rapidly identify socio-political activity in South Africa using proxy data from social media. We measure and analyse scalar properties of a network created by user interactions on Twitter. Our experimental results show that network diameter and reciprocity have statistical significance in determining socio-political activity.**

***Index Terms*—Social Media, Social Networks Analysis, RDF Graphs, Socio-political, Network Metrics**

I. Introduction

A socio-political problem is any condition or behaviour that has negative consequences for large numbers of people in a society and is generally recognized as a condition that needs to be addressed [10]. Sociological surveys have been used by researchers and policy makers as a tool for collecting and monitoring socio-political activities. Social media has provided a new way of delivering abundant data on almost any topic. The real-time and wide-spread nature of social media has quickly made it become a lens of perspective [13]. Research has shown that social media has become pivotal in shaping the political discourse around the world [2], [3], [4]. A report released by Portland[1] shows that Africa is tweeting about political issues more than USA and UK. This is indicative of the adoption of social media in Africa as a platform for political discourse. In this work, we aim to identify socio-political activities in South Africa by analysing the network structure created by user interactions on social media.

A network is represented as a graph of vertices and edges. Two popularly used network formats are Resource Description Framework (RDF) [18] and Social Networks [16]. RDF is a family of World Wide Web Consortium (W3C) specifications designed to model information that is implemented in Web resources. Objects in a network are represented in the form subject-predicate-object expressions, known as triples. The concepts in an RDF graph are formalised by a controlled vocabulary called an ontology. SPARQL [21] is a query language for retrieving semantic information from RDF graphs. Social Networks represent objects with asymmetric relations. Social Networks proposes graph algorithms for computing metrics that characterise the overall structure of a network. In our work, we leverage the expressivity power of RDF and Social Networks algorithms to create a network abstraction of the tweets.

Previous studies have shown that it's plausible to gain a tremendous amount of insights by analysing networks in social media. Network analysis has been used significantly in understanding the topology of user relationships in social media for viral advertisement [17]. The topology of retweet and mention networks on Twitter has been analysed to understand communication between communities with different political orientations in the United States [4]. The retweet network structure has also been analysed as part of deciphering public opinions on UK's decision to leave or remain in the European Union (EU) [15]. Social media was seen has a driving force in 2011 Arab Spring uprisings in Africa (Egypt, Tunisia and Libya) and the Middle East. Among many analyses done on the large amount of data produced by these uprisings, the interaction network created by users has been analysed to understand participation of directly affected citizens and onlookers [14].

In this paper, we analyse the network formed by message forwarding and mentions on South African Twitter. We use existing ontologies — SIOC[2](an ontology for describing information from online communities) and SemSNA(an ontology for describing Social Networks indices) [5] — as a basis for conceptualising user interactions on Twitter. We also use Social Networks algorithms to measure network diameter, nodes, edges, density, clustering coefficient, assortativity, reciprocity, average path length and network modularity.

The contributions of this paper are two-fold:

- Firstly, we identify metrics that are indicators of socio-political activity in social media.
- Secondly, we extend the SIOC ontology to describe forward and mention interactions on Twitter. We also extend the SemSNA ontology to include density, assortativity, clustering coefficient and reciprocity concepts.

The paper is organized as follows. Section II describes our methodology for topic classification, network representation and the techniques used for comparing network structures. Section III reports on the experimental results. In section IV we review literature related to our work. Finally, in Section V we give the conclusion.

[1]https://portland-communications.com/publications/how-africa-tweets-2015/

[2]https://www.w3.org/Submission/sioc-spec/

II. Data and Methods

A. Data Collection and Topic Classification

We created a dataset of 1 million tweets by crawling the public Twitter API. We partitioned our dataset into 10 sub-datasets according to the 10 hashtags (shown in Table I) used to download the data. Users on Twitter use hashtags (#) to indicate a topic in a message. Therefore, in this paper, a topic corresponds to all the tweets downloaded for each hashtag. In order to do further analysis on the data, we trained a multi-label classifier to associate topics with three labels as follows:

- $\mathrm{S_p}$: Socio-political topic.
- E: Entertainment topic.
- P: Topic about people.

In multi-label classification, topics are associated with a set of labels $Y \subseteq L$. We use a multi-label classifier for topic categorization because social media posts can belong to more than one conceptual class [19]. For example, a post on the 2013 Boston marathon can include the hashtags #BostonMarathon and #TerroristBombing. We have used a k-Nearest Neighbor (KNN) algorithm adpated for multi-label classification [1]. Some topics in the dataset were assigned more than one label as shown in Table I. We use the following labels to tag topics with more than one label:

- E-$\mathrm{S_p}$: A topic labeled as entertaiment and socio-political.
- P-$\mathrm{S_p}$: A topic labeled as people and socio-political.
- E-P: A topic labeled as entertainment and people.
- E-P-$\mathrm{S_p}$: A topic labeled as entertaiment, people and socio-political.

TABLE I
Topics and Classifications

Hashtag	Description	Class
#FeesMustFall	Protesting university students against the increase in school fees.	$\mathrm{S_p}$
#SASSA	Voice out against the decision made by South Africa Social Security Agency (SASSA) to stop paying social grants.	$\mathrm{S_p}$
#ZumaMustFall	Match against the reshuffle of cabinet made by President Zuma.	$\mathrm{S_p}$
#Isibaya	South African TV soap opera.	E-$\mathrm{S_p}$
#DateMyFamily	South African TV reality show.	E
#MissSA	South Africa beauty pageant held on March 26, 2017.	E
#OurPerfectWedding	South African TV reality show.	E
#RIPJoeMafela	Death of Joe Mafela. Joe Mafela was a South African actor and singer.	P
#AhmedKathrada	Death of Ahmed Kathrada. Ahmed Kathrada was a South African anti-apartheid activist.	P-$\mathrm{S_p}$
#OscarPistorius	Discuss the trial of a South African sprint runner Oscar Pistorius accused of killing his girlfriend.	P-$\mathrm{S_p}$

To train our classifier, we downloaded news articles tagged with Political, Entertainment and People from South African online news media: news24[3], IOL[4] and timeslive[5]. We used these tagged articles as a basis for categorising topics in the dataset. Tweets are shorter (140 characters) than news articles. To compensate for this difference, we combined all the tweets in a topic into a single document and use the document as input to a classifer. The classifier associate each document with a set of labels $Y \subseteq L$, where L= $\left\{\mathrm{S_p},\mathrm{E},\mathrm{P},\mathrm{E\text{-}S_p},\mathrm{E\text{-}P},\mathrm{E\text{-}P\text{-}S_p}\right\}$.

[3]http://www.news24.com/SouthAfrica
[4]http://www.iol.co.za/news
[5]http://www.timeslive.co.za/

B. Network Representation and Measurement

A network is represented as a graph of vertices and edges. In this work, we used two popular network representation formats — Resource Description Framework (RDF) [18] and Social Networks [16] — to create a network abstraction of the topics in the dataset. We leverage the expressivity power of RDF to describe the relationships among Twitter users. Though RDF is well suited for semantic analysis of a network, it is limited in giving insights into the overall structure. For this reason, we used Social Networks for measuring and analysing structural properties of a network.

We create and analyse networks in four steps:

1) We first create a network of users and tweets using RDF. Several ontologies already exist to conceptualise concepts in online social networks like Twitter [6]. In this work, we used an existing RDF ontology, SIOC. SIOC relates a user to a tweet through the creator_of property as shown in Fig. 1. We have extended the Post concept in the ontology to include the has_forwarder and has_mention properties.
2) The second step is to use SPARQL [21] to extract a forward and mention network from the RDF graph using the has_forwarder and has_mention properties respectively. We also extracted the forward-mention combined network by using polymorphic SPARQL queries [5] using the super-property: sioc:related_to. In a forward network, we define an edge from user A to user B if A forwarded a message from B. In a mention network, an edge connects A to B if A mentioned B.
3) The third step is to measure structural properties of the networks and extend the RDF graph created in step 1 with the measures. We used SemSNA [5] for conceptualizing network metrics. SemSNA is an ontology designed to model social networks metrics in RDF graphs. Fig. 2 shows part of the SemSNA ontology. We have extended the ontology to describe network density, assortativity, clustering coefficient and reciprocity. In this paper, our primary focus is to compare networks across topics in the dataset using a consistent set of network structural properties such as, network diameter, density, number of nodes, number of edges, reciprocity, assortativity, clustering coefficient, average path length and network modularity [16].

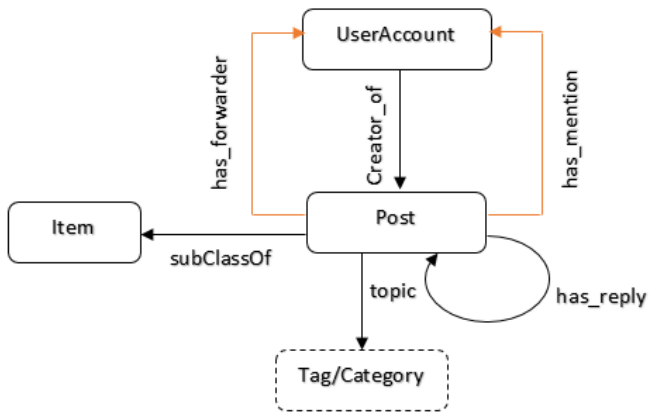

Fig. 1. Extended SIOC.

The network properties are defined as follows:
Network diameter: Every node N has a shortest path to other nodes in a network called geodesic. A network diameter is the longest geodesic distance in a network.
Network density: Density D is calculated as a ratio $D = 2E/N(N-1)$, where E is the number of edges and N is a number of nodes.
Number of nodes: Is a sum of all nodes in a network.
Number of edges: Is a sum of all edges in a network.
Network reciprocity: Is the proportion of mutually linked vertices in a directed network.
Network assortativity: Is the tendency of nodes to connect to others nodes with similar edge degrees.
Network clustering coefficient: Measures the probability that the adjacent vertices of a vertex are connected. In this paper, we calculate the global clustering coefficient based on triplets of nodes. A triplet consists of three connected nodes. A triplet can either be open (connected by two undirected ties) or closed (connected by three undirected ties). Global clustering coefficient is defined as: C =number of closed triplets/number of connected triplets of vertices (both open and closed).
Network average path length: Is the mean of the shortest distance between each pair of nodes in the network.
Network modularity: Measure the strength of cohesion in clusters/communities in a network. A network with high modularity has dense connections between nodes in clusters but sparse connections between nodes in different clusters.

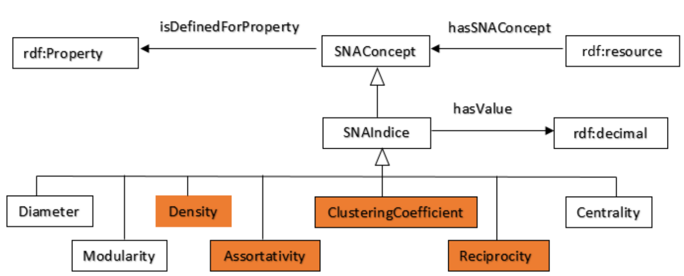

Fig. 2. Extended SemSNA.

4) The last step (described in more details in the next section) is to use the RDF graph created in step 1 and 3 to understand the relationships and characteristics among users, tweets and network metrics.

Table II gives an example of measures captured from the RDF graphs.

C. Comparing Networks across Topics

In this section, we discuss the methods used to identify patterns of network properties across different classes of topics. We used multiple regression models and correlation techniques to understand how different network parameters interact [24], [23], [14], [4], [25].

Correlation: Correlation looks at the global trend shared between two variables. For example, we want to understand if increase in the number of tweets leads to increase or decrease in network density. We used the Pearson correlation coefficient [22] to measure correlations among tweets and network properties. We compared variable correlations across topic classes to identify patterns unique to each class of topics.

Multiple regression modelling: Regression explains the causal relationship between variables. We use multiple regression to find a relationship between network properties (response variables) and topic classes (independent variables) [7], [8]. Let socio-political class S_{p} be X_1, entertainment class E be X_2 and people class P be X_3, then a multiple regression equation is given by:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3, \tag{1}$$

where Y is a network property. In this work, we used three types of multiple regression models,

- Linear model: We used a linear model for network metrics that are quantitative and normally distributed, e.g. network density.
- Logisitic model: We used presence-absence analysis to analyse network reciprocity. If reciprocity is present in a network - calculated as greater than 0 - we reduce it to 1, otherwise 0 if reciprocity is absent. We used multiple logistic regression for modelling reciprocity because the response variable is categorical with two possible binary outcomes. The response variable Y takes on values 0 or 1, and is modelled as a binomial distribution with probability $P(Y_{\mathrm{i}} = 1) = \pi_{\mathrm{i}}$,
- Poisson model: Poisson regression provides a model that describes how the mean μ response changes as a function of one or more independent variables. The Response Y is modelled as Poisson distribution that is $y_{\mathrm{i}} \sim Poisson\,(\mu_{\mathrm{i}})$ for $i = 1, ..., N$ where the expected count of y_{i} is $E(Y) = \mu$. We used Poisson regression to model network properties collected as counts (e.g, number of nodes, network diameter, number of edges etc.).

Both Logisitic and Poisson regressions belong to a family of regression models called generalized linear models [9], [10].

TABLE II
SAMPLE NETWORK METRIC MEASURES

class	$\mathbf{t_n}$	**n**	**m**	ϕ	**p**	**C**	**r**	$\mathbf{r^{\leftrightharpoons}}$	l_G	Q	**type**
P-S_p	857	394	423	3	0.005463634	0.0005309359	0.004471068	0	1.09636	0.5290786	forward
S_p	805	582	567	5	0.003353621	0.0007443553	-0.01618863	0.003527337	1.348158	0.7023055	combined
E	3476	447	302	2	0.003029665	0.01648352	0.06462597	0.006622517	1.103858	0.9404357	mention
P	456	140	118	2	0.01212744	0	-0.05746937	0	1.008403	0.8601336	forward
S_p	12631	1445	1707	9	0.00163617	0.009588494	-0.007949564	0.01405975	2.964084	0.8148845	combined
E-S_p	1736	427	431	3	0.004738815	0.001455604	-0.004947816	0.004640371	1.103858	0.8068594	mention
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

[a] The properties measured are: total number of tweets t_n; total number of vertices n; total number of edges m; network diameter ϕ; density p; clustering coefficient C; assortativity r; reciprocity $r^{\leftrightharpoons}$; average path length l_G and network modularity Q. The last column shows the type of a network.

III. Analysis of results

In this section, we report on the findings from our experiments. We sampled data from each topic in our dataset at a daily interval. Topics in the dataset had different sample sizes because some topics have a longer life-span on Twitter than others and some topics receive more attention from users than others. Table III shows the number of days each topic was sampled. Using RDF and Social Networks, we created a network abstraction of the tweets in each sample. We examined the correlations of different network metrics (shown in Table II) using the Pearson correlation coefficient. Our experiments yielded no correlation patterns across the topical classses. We also used regression models to analyse causal relationships between the parameters shown in Table II and the topical classes. In this paper, we only report on the network properties where socio-political class had a significant impact. The experiments showed socio-political class to have a significant impact on reciprocity and diameter. The remainder of the section reports our findings.

A. Reciprocity

Reciprocity takes on the values 0 if there is no reciprocity in the sampled network or 1 if reciprocity is present. Because reciprociy is categorical with two possible outcomes, we assumed reciprocity to have a binomial distribution. To determine topical classes with a significant impact on the presence of reciprocity in forward, mention and forward-mention networks, we fit the measures of reciprocity in the logistic equation given in "(1)".

According to the output, the models of the forward, mention and forward-mention networks are $logit(\pi) = -2.06467 + 1.62936X_1 - 0.01167X_2 + 0.23275X_3$, $logit(\pi) = -2.1314 + 1.7766X_1 + 0.8595X_2 + 0.7168X_3$ and $logit(\pi) = -0.3979 + 1.8019X_1 - 0.6340X_2 - 0.8450X_3$ respectively.

Before performing statistical tests on the variables of the regression, we checked the residual plots in order to examine the extent of deviation in the models [25]. The residual plot in forward network shown in Fig. 3 slants slightly to the positive and had points with deviance residuals greater than 1. The residue plot in mention network shown in Fig. 4 slants slightly to the negative with no points less than -1. The combined network slants slightly both to the positive and negative with no point with residual greater than 1 or less than -1. In our experiments, we set the threshold of deviance to 1 on the positive and -1 on the negative. Therefore, we removed all points in the forward network with residuals greater than 1 and fit the logistic regression again. Like in [25], the results did not change the pattern of the results. Therefore, all the three models were used to test the null hypotheses. We tested the null hypotheses $H_0 : \beta_1 = 0$, $H_0 : \beta_2 = 0$ and $H_0 : \beta_3 = 0$ using the z and p-values shown in Table IV. The statistical tests show that X_1 (socio-political class) has a significant impact on the probability of the presence of reciprocity in forward, mention and the combined networks. The results also show X_2 and X_3 having significant impact in mention and combined networks respectively. Although X_2 and X_3 have significance, X_1 has a significant higher z-value and lower p-value.

TABLE III
NUMBER OF SAMPLES IN EACH TOPIC

Topic	Number of days sampled	Total number of tweets
Topic 1	89	393153
Topic 2	37	63053
Topic 3	41	226641
Topic 4	32	26949
Topic 5	22	59316
Topic 6	19	48673
Topic 7	27	49592
Topic 8	21	14464
Topic 9	26	48283
Topic 10	9	69876

TABLE IV
RECIPROCITY Z AND P-VALUES

		z-value	**p-value**
forward	X_1	3.790	0.000151
	X_2	-0.030	0.975837
	X_3	0.591	0.554481
mention	X_1	4.557	5.19e-06
	X_2	2.258	0.0239
	X_3	1.862	0.0626
combined	X_1	4.756	1.97e-06
	X_2	-1.574	0.1156
	X_3	-2.095	0.0361

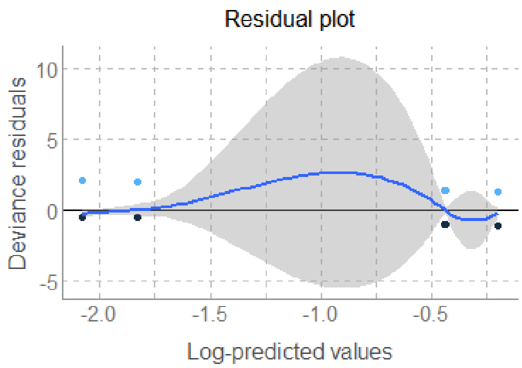

Fig. 3. Residual plot for forward network in logisitic model.

B. Diameter

Network diameter is the longest of all the calculated path counts, therefore we assume a Poisson distribution. To determine topical classes with a significant impact on the mean of the diameter in forward, mention and forward-mention networks, we fit the diameter measures in the Poisson equation given in “(1)”. The models of the forward, mention and forward-mention networks are, $g(\mu) = 0.94737 + 0.60717X_1 - 0.33138X_2 - 0.27890X_3$, $g(\mu) = 0.22911 + 0.88159X_1 + 0.25664X_2 + 0.39635X_3$ and $g(\mu) = -0.3979 + 1.8019X_1 - 0.6340X_2 - 0.8450X_3$ respectively.

Before testing the null hypotheses, we examined the residual plots for deviation in the models [25]. The residual plot in forward network shown in Fig. 5 slants slightly to the negative. The residue plot in mention network shown in Fig. 6 showed no major deviations. The forward-mention combined network slants slightly to the negative. All the three networks had no points with deviation greater than 1. Thus, all the three models were used to test the null hypotheses. We tested the null hypotheses $H_0 : \beta_1 = 0$, $H_0 : \beta_2 = 0$ and $H_0 : \beta_3 = 0$ using the z and p-values shown in Table V. The tests show all variables having significant impact on the mean of diameter in all the three networks though X_1 (socio-political class) has the highest z and p-values.

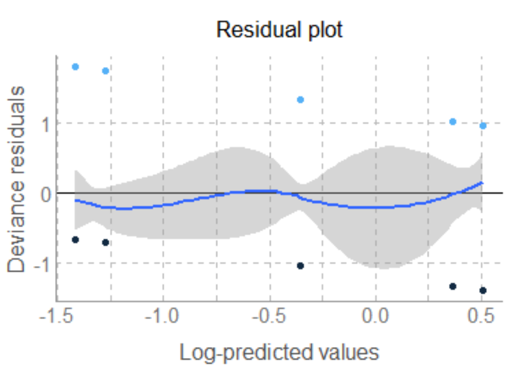

Fig. 4. Residual plot for mention network in logisitic model.

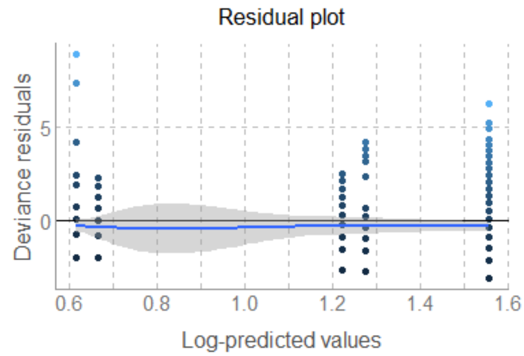

Fig. 5. Residual plot in forward network in Poisson model.

IV. Related Work

In the early days of social media, one of the first questions investigated was whether social media is a new medium of information sharing. Social Networks algorithms have been used for measuring topological characteristics of user networks. Network properties like total nodes, total edges, average degree, diameter, clustering coefficient, reciprocity, community structure, hubs and authorities have be used to measure user intention on Twitter, compare user interactions across continents and study the flow of information [24], [23]. The behavior of a user network on Twitter has been compared to a human social network and experimental results show a non-power-law distribution, a short effective diameter, and low reciprocity suggesting that Twitter is used more as an information sharing platform than a social network.

The wide-spread and real-time nature of Twitter has attracted policy makers to monitor public opinion for political strategic planning. Methodoligies have been designed for analysing the user network topology to predict political activity on Twitter [4], [14], [15]. The interaction network on Twitter (replying and retweeting) in the 2011 Arab Spring uprisings was analysed to understand participation of directly affected citizens and onlookers. Network features were used to train a classifier for predicting political alignment of Twitter users in 2010 U.S. midterm elections. Experimental results showed the classifier to outperform content-based classifiers.

TABLE V
Diameter Coefficients

		z-value	**p-value**
forward	X_1	5.951	2.67e-09
	X_2	-3.534	0.00041
	X_3	-2.901	0.00372
mention	X_1	8.634	2.0e-16
	X_2	2.714	0.00665
	X_3	4.282	1.85e-05
combined	X_1	6.566	5.16e-11
	X_2	-4.350	1.36e-05
	X_3	-3.799	0.000145

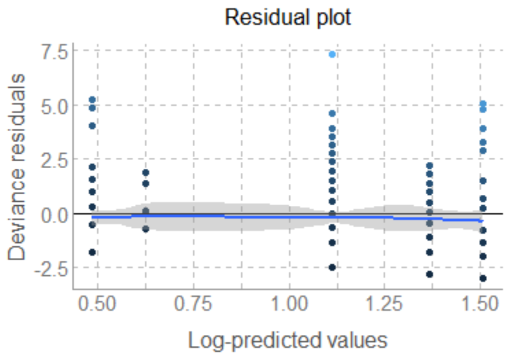

Fig. 6. Residual plot in mention network in Poisson model.

Topological analysis of a user network on Twitter has also been used as part of deciphering public opinions on UK's decision to leave or remain in the European Union (EU). In 2016, Twitter was used to decipher the 2016 U.S. presidential campaign. The follower-following network was used in part to analyse four dimensions of follower demographics: social status, gender, race and age [26]. Experimental results showed that the Trumpists were more polarized than the Clintonists, and were more likely to be either very young or very old and no gender affinity effected Clinton.

V. Conclusions

In this paper, we analysed 1 million tweets downloaded from South African social media. We associated topics with three labels — S_p (Socio-political), E (Entertainment) and P (People) — using a k-Nearest Neighbor multi-label classifier. We leveraged the expressivity power of Resource Description Framework (RDF) and Social Networks graph algorithms to model the complex relationships and characteristics of Twitter users, tweets and network properties in the forward and mention networks. Our experimental results show socio-political topics to significantly impact the presense of reciprocity in mention, forward and forward-mentione combined networks. The results also show that the mean of the diameter in forward, mention and forward-mention combined networks is significantly impacted by topics with socio-political alignment.

References

[1] Zhang, M. L., & Zhou, Z. H. (2005, July). A k-nearest neighbor based algorithm for multi-label classification. In Granular Computing, 2005 IEEE International Conference on (Vol. 2, pp. 718-721). IEEE.

[2] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. Icwsm, 10(1), 178-185.

[3] 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), 1-2.

[4] Conover, M., Ratkiewicz, J., Francisco, M. R., Gonalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on twitter. ICWSM, 133, 89-96.

[5] Ereteo, G., Buffa, M., Gandon, F., & Corby, O. (2009). Analysis of a real online social network using semantic web frameworks. The Semantic Web-ISWC 2009, 180-195.

[6] Ereteo, G., Buffa, M., Gandon, F., Grohan, P., Leitzelman, M., & Sander, P. (2008). A state of the art on social network analysis and its applications on a semantic web. In SDOW2008.

[7] Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. Journal of the american Statistical association, 97(460), 1090-1098.

[8] Asur, S., & Huberman, B. A. (2010, August). Predicting the future with social media. In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on (Vol. 1, pp. 492-499). IEEE.

[9] Dobson, A. J., & Barnett, A. (2008). An introduction to generalized linear models. CRC press.

[10] Venables, W. N., & Ripley, B. D. (2013). Modern applied statistics with S-PLUS. Springer Science and Business Media.

[11] Barkan, S. E. (2013). Social problems: Continuity and change. Flat World Knowledge, Incorporated.

[12] Dewing, M. (2010). Social media: An introduction (Vol. 1). Ottawa: Library of Parliament.

[13] Buntain, C., McGrath, E., Golbeck, J., & LaFree, G. (2016). Comparing Social Media and Traditional Surveys around the Boston Marathon Bombing. In # Microposts (pp. 34-41).

[14] Bruns, A., Highfield, T., & Burgess, J. (2013). The Arab Spring and social media audiences: English and Arabic Twitter users and their networks. American Behavioral Scientist, 57(7), 871-898.

[15] Khatua, A., & Khatua, A. (2016, December). Leave or Remain? Deciphering Brexit Deliberations on Twitter. In Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on (pp. 428-433). IEEE.

[16] Scott, J. (2017). Social network analysis. Sage.

[17] Chu, S. C. (2011). Viral advertising in social media: Participation in Facebook groups and responses among college-aged users. Journal of interactive advertising, 12(1), 30-43.

[18] Bratt, S. (2007). Semantic web and other W3C technologies to watch. Talks at W3C, January.

[19] Wang, J. (Ed.). (2008). Data warehousing and mining: Concepts, methodologies, tools, and applications: Concepts, methodologies, tools, and applications (Vol. 3). IGI Global.

[20] Perer, A., & Shneiderman, B. (2008, January). Systematic yet flexible discovery: guiding domain experts through exploratory data analysis. In Proceedings of the 13th international conference on Intelligent user interfaces (pp. 109-118). ACM.

[21] Prud, E., & Seaborne, A. (2006). SPARQL query language for RDF.

[22] Pearson, K. (1895). Note on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London, 58, 240-242.

[23] Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World wide web (pp. 591-600). ACM.

[24] Java, A., Song, X., Finin, T., and Tseng, B. (2007, August). Why we twitter: understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis (pp. 56-65). ACM.

[25] Davenport, S. W., Bergman, S. M., Bergman, J. Z., & Fearrington, M. E. (2014). Twitter versus Facebook: Exploring the role of narcissism in the motives and usage of different social media platforms. Computers in Human Behavior, 32, 212-220.

[26] Wang, Y., Li, Y., & Luo, J. (2016, March). Deciphering the 2016 US Presidential Campaign in the Twitter Sphere: A Comparison of the Trumpists and Clintonists. In ICWSM (pp. 723-726).