A method to evaluate the reliability of social media data for social network analysis

Derek Weber[1,2*], Mehwish Nasim[3-6], Lewis Mitchell[5,6], Lucia Falzon[2,7]
[1]School of Computer Science, University of Adelaide, Adelaide, Australia
[2]Defence Science and Technology Group, Adelaide, Australia
[3]Data61, Commonwealth Scientific and Industrial Research Organisation, Adelaide, Australia
[4]Cyber Security Cooperative Research Centre, Adelaide, Australia
[5]ARC Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS), Adelaide, Australia
[6]School of Mathematical Sciences, University of Adelaide, Adelaide, Australia
[7]School of Psychological Sciences, University of Melbourne, Melbourne, Australia
[*]derek.weber@{adelaide.edu.au,dst.defence.gov.au}

***Abstract*—In order to study the effects of Online Social Network (OSN) activity on real-world offline events, researchers need access to OSN data, the reliability of which has particular implications for social network analysis. This relates not only to the completeness of any collected dataset, but also to constructing meaningful social and information networks from them. In this multidisciplinary study, we consider the question of constructing traditional social networks from OSN data and then present a measurement case study showing how the reliability of OSN data affects social network analyses. To this end we developed a systematic comparison methodology, which we applied to two parallel datasets we collected from Twitter. We found considerable differences in datasets collected with different tools and that these variations significantly alter the results of subsequent analyses. Our results lead to a set of guidelines for researchers planning to collect online data streams to infer social networks.**

I. Introduction

It is assumed that the social networks present on online social networks (OSNs) can inform the study of information dissemination and opinion formation, contributing to an understanding of offline community attitudes. Though such claims are prevalent in the social media literature, there are serious questions about their validity due to an absence of SNA theory on online behaviour, the mapping between online and offline phenomena, and the repeatability of such studies. In particular, the issue of reliable data collection is fundamental. Collection of OSN data is often prone to inaccurate boundary specifications due to sampling issues, collection methodology choices, as well as platform constraints.

Previous work has considered the question of data reliability from a sampling perspective [1]–[4], biases [5]–[8] and the danger of making invalid generalisations using "big data" approaches lacking nuanced interpretation of the data [9], [10]. Analyses of incomplete networks exist [11], but this paper specifically considers the questions of data reliability for SNA, considering not only the significance of online interactions to discover meaningful social networks, but also how sampling and boundary issues can complicate analyses of the networks constructed. Through an exploration of modelling and collection issues, and a measurement study examining the reliability of simultaneously collected, or *parallel*, datasets, this multidisciplinary study addresses the following research question:

> *How do variations in collections affect the results of social network analyses?*

Our work makes the following contributions:

1) Discussion of the challenges mapping OSN data to meaningful social and information networks;
2) A methodology for systematic dataset comparison;
3) Recommendations for the use and evaluation of social media collection tools; and
4) Two original social media datasets collected in parallel, and relevant analysis code, available at *https://github.com/weberdc/socmed_sna*.

Further detail is available in [12].

II. Social networks from social media

Using SNA to explore social behaviours and processes from OSN data presents many challenges. Most easily accessible OSN data consists of timestamped interactions, rather than details of long-standing relationships, which form the basis of SNA theory. Online interactions offer a window onto online behaviour only, and any implications for offline relations and behaviour are unclear. Additionally, although interactions on different OSNs are superficially similar, how they are implemented may subtly alter their interpretation. Beyond modelling and reasoning with the data is the question of collection – accessing the right data to construct meaningful social networks is challenging. OSNs provide a limited subset of their data through a variety of mechanisms, balancing privacy and competitive advantage with openness and transparency.

A. Interactions and relationships online

The concepts and tools provided by SNA assume an established social network based on stable relationships, even when they are dynamic [13]–[15]. As the typical equivalent (*friend* and *follower* relations) are costly to obtain, may be

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/$31.00

stale, and differ in semantics between OSNs, we look to direct interactions which can reveal the extent, direction and recency of the flow of information (and arguably influence) between actors [16]. Each OSN presents its features in unique ways, resulting in communities with unique interaction cultures; the intended audience of a particular interaction may vary depending on the OSN (e.g., replying to a politician's tweet may signal to the politician, the other respondents, or the poster's followers), meaning that care must be taken when comparing superficially similar networks from different OSNs.

B. Social Network Analysis theory

Relationships between individuals in a social network may last for extended periods of time, vary in strength, and be based upon a variety of factors, not all of which are easily measurable. Because of the richness of the concept of social relationships, data collection is often a qualitative activity, involving directly surveying community members for their perceptions of their direct relations and then perhaps augmenting that data with observation data such as recorded interactions (e.g., meeting attendance, emails, phone calls). It is tempting to believe that this richness should be discoverable in the vast amount of interaction data available from OSNs, but there are issues to consider:

1) links between social media accounts may vary in type and across OSNs — it is unclear how they contribute to any particular relationship;
2) what is observed online is only a partial record of a relationship's interactions, where interactions may occur via other OSNs or online media, or entirely offline; and
3) collection strategies and OSN constraints may also hamper the ability to obtain a complete dataset.

C. Challenges obtaining OSN data

Because OSNs provide data via proprietary Application Programming Interfaces (APIs), they control *how much* data is available, their *types*, and the *precision or flexibility of queries*, and typically charge for extra or more specific access, which raises questions of repeatability [5]. Studies of Twitter's 1% Sample API have raised questions regarding the representativeness of the data it provides [3], [4], which then "affects the networks of communication that can be reconstructed from the messages sampled" [2, p.17]. Analyses of human-generated social media relying on "big data" approaches may also miss fine grained nuances, such as sarcasm and disgust [10], and many studies fail to consider that not all social interactions between a pair will occur online [17] and that passive users may be influenced to act offline [18].

Although OSN APIs often include both streaming and retrieval models, there are two primary conceptual collection approaches: 1) focusing on one or more seed users and investigating their neighbourhoods; and 2) relying on keywords to capture discussion around a topic. Morstatter *et al.* [19] used both these approaches in their study of the 2017 German election: an initial keyword-based collection was conducted for 11 days to identify the most active accounts, the usernames of which were then used as keywords in a six week collection.

III. Methodology

Our initial hypothesis was that if the same collection strategies were used at the same time, then each OSN would provide the same data, regardless of the collection tool used. Consequently, social networks built from such data using the same methodology should be highly similar, in terms of both network and node level measurements. Our methodology consisted of these steps:

1) Conduct simultaneous collections on an OSN using the same collection criteria with different tools.
2) Compare statistics across datasets.
3) Construct sample social networks from the data collected and compare network-level statistics.
4) Compare the networks at the node level.
5) Compare the networks at the cluster level.

A. Data collection

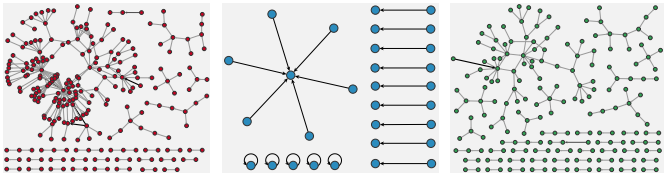
Twitter was chosen as the source OSN due to the availability of its data, the fact that the data it provides is highly regular [3], and because it has similar interaction primitives to other major OSNs. In the interests of comparing a proprietary collection tool with a baseline two tools were chosen:
Twarc (https://github.com/DocNow/twarc) is an open source library which wraps Twitter's API, and provided the baseline.
RAPID [20] is a social media collection and data analysis platform for Twitter and Reddit. It enables filtering of OSN live streams, as well as dynamic topic tracking, meaning it can update filter criteria in real time, adding terms popular in recent posts and removing unused ones.

Both tools facilitate filtering Twitter's Standard live stream with keywords, providing datasets of tweets as JSON objects.

B. Constructing social networks

A social network is constructed from dyads of pair-wise relations between nodes, which in our case are Twitter accounts. The node ties denote intermittent relations between accounts, inferred from observed interactions [14], [15]. Here, we consider three social networks built from interaction types common to many OSNs: 'mention networks', 'reply networks', and 'retweet networks' (e.g., retweets are analogous to Facebook shares or Tumblr reposts, and replies analogous to comments on posts on Reddit) (see Figure 1). These networks are all weighted directed networks: an arc from u to v implies account u mentioned/replied to/retweeted account v, and the edge weight corresponds to the frequency of that occurrence in the dataset. Although retweets are not necessarily direct interactions [5], they can be used to determine reach.

Other network types can be constructed based on direct or inferred relations, including shared use of hashtags or URLs, reciprocation or minimum levels of interaction, or friend/follower connections. As our focus is on the social relationships implied by direct interactions we will focus only on the above three types of network construction here, however similar issues will certainly arise in other applications.



(a) Mentions. (b) Replies. (c) Retweets.

Fig. 1. Sample networks of accounts built from 5 minutes of Twitter data. Nodes may appear in one or more networks, depending on their behaviour during the sampled period. While all are dominated by a large component, the mention network has higher cohesiveness. The similarity between the mention and retweet networks is because retweets mention the account being retweeted. The diagrams were constructed with *visone* (https://visone.info).

C. Analyses

1) Dataset statistics: To compare the parallel datasets we examined the following features:

Absolute counts: accounts, tweets, retweets, quotes, replies, URLs, hashtags, and mentions; and

Highest counts: tweeting account, mentioned account, retweeted tweet, replied-to tweet, used hashtags and URLs.

Based on these figures, we account for major discrepancies, which can guide post-processing (e.g., spam filtering).

2) Network statistics: The following network statistics are used to assess differences in the constructed networks: number of nodes, edges, average degree, density, mean edge weight, component count and the size and diameter of the largest, Louvain [21] cluster count and the size of the largest, reciprocity, transitivity, and maximum k-cores.

3) Centrality values: Centrality measures offer a way to consider the importance of individual nodes within a network. We consider degree, betweenness, closeness and eigenvector centralities for the mention and reply networks only (edges in retweet networks are not necessarily direct interactions [5]).

Given the set of nodes in each corresponding pair of networks is not guaranteed to be identical, it is not possible to directly compare the centrality values of each node, so instead we rank the nodes in each network by the centrality values, take the top $1,000$ from each list, further constrain the lists to only the nodes common to both lists, and then compare the rankings with the Kendall τ coefficient. We also calculate Spearman's ρ coefficient as a confirmation measure.

4) Cluster comparison: The final step is to consider the clusters discoverable in the mention, reply and retweet networks and compare their membership. We use Louvain clustering [21] and the Adjusted Rand index [22] for this.

IV. Experiment

A. Data collection

To obtain a moderately active portion of activity, we collected data from Twitter's Standard live stream relevant to an Australian television panel show that invites its viewers to participate in the discussion live using the hashtag #QandA. A particular broadcast in 2018 was chosen due to the expectation of high levels of activity given the planned discussion topic. As a result, the filter keywords used were 'qanda' (to catch @qanda too) and two terms that identified a panel member (available on request). We collected two parallel datasets:

Part 1: Four hours starting 30 minutes before the hour-long programme, to allow for contributions from the country's major timezones; and

Part 2: From 6am to 9pm the following day, capturing further related online discussions.

Twarc acted as the baseline collection as it provides direct access to Twitter's API, while RAPID was configured to use *co-occurrence keyword expansion* [20], meaning it would progressively add keywords to the original set if they appeared sufficiently frequently (five times in ten minutes). This expanded dataset was referred to as 'RAPID-E' and was filtered back to just the tweets containing the original keywords (and labelled 'RAPID') for comparison with the 'Twarc' dataset. We expected the moderate activity observed would not breach rate limits, and thus RAPID should capture all tweets captured by Twarc. This was not the case (Table I).

All data were treated according to approved University of Adelaide ethics protocols #170316 and H-2018-045.

TABLE I
SUMMARY DATASET STATISTICS.

	Dataset	All Tweets	Unique Tweets	Retweets	All Accounts	Unique Accounts
Part 1	Twarc	27,389	11,481	14,191	7,057	2,090
(20:00-00:00)	RAPID	15,930	22	8,744	4,970	3
	RAPID-E	17,675	1,767	9,767	5,547	527
Part 2	Twarc	15,490	4,089	10,988	5,799	1,128
(06:00-21:00)	RAPID	11,719	318	8,051	4,708	37
	RAPID-E	23,583	12,180	13,679	8,854	4,007

TABLE II
DETAILED DATASET STATISTICS.

	Part 1		Part 2	
	RAPID	Twarc	RAPID	Twarc
Quotes	325	1,203	498	1,232
Replies	1,446	2,067	1,715	1,731
Tweets with hashtags	10,043	15,591	3,912	3,961
Tweets with URLs	2,470	4,029	3,106	4,074
Most prolific account	Account a_1	Account a_1	Account a_2	Account a_3
Tweets by most prolific account	103	146	57	68
Most retweeted tweet	Tweet t_1	Tweet t_1	Tweet t_2	Tweet t_2
Most retweeted tweet count	260	288	385	385
Most replied to tweet	Tweet t_3	Tweet t_3	Tweet t_4	Tweet t_4
Most replied to tweet count	55	121	58	58
Tweets with mentions	11,314	18,253	10,472	13,514
Most mentioned account	Account a_4	Account a_4	Account a_4	Account a_4
Mentions of most mentioned account	2,883	3,853	2,753	2,752
Hashtags uses	15,700	23,557	7,672	7,862
Unique hashtags	1,015	1,438	960	1,082
Most used hashtag	#qanda	#qanda	#qanda	#qanda
Uses of most used hashtag	10,065	15,644	2,545	2,549
Next most used hashtag	#auspol	#auspol	#auspol	#auspol
Uses of next most used hashtag	1,381	2,103	1,652	1,349
URLs uses	913	1,650	1,602	2,411
Unique URLs	399	560	658	790
Uses of most used URL	49	128	71	81

B. Comparison of collection statistics

The first striking difference between the datasets is the greater number of tweets and accounts collected by Twarc than with RAPID (Table I). This occurred in both parts, although

TABLE III
PART 1 NETWORK STATISTICS.

	RETWEET		MENTION		REPLY	
	RAPID	Twarc	RAPID	Twarc	RAPID	Twarc
Nodes	3,234	4,426	4,535	6,119	1,184	1,490
Edges	7,855	12,327	13,144	19,576	1,231	1,631
Average degree	2.429	2.785	2.898	3.199	1.040	1.095
Density	0.001	0.001	0.001	0.001	0.001	0.001
Mean edge weight	1.113	1.151	1.268	1.300	1.175	1.267
Components	74	95	86	108	164	192
Largest component	3,061	4,115	4,326	5,819	829	1,081
- Diameter	12	12	10	11	15	15
Clusters	93	115	109	134	186	219
Largest cluster	318	540	731	1,348	169	229
Reciprocity	0.004	0.007	0.025	0.025	0.106	0.099
Transitivity	0.026	0.034	0.065	0.063	0.024	0.021
Maximum k-core	11	14	13	16	2	3

TABLE IV
PART 2 NETWORK STATISTICS.

	RETWEET		MENTION		REPLY	
	RAPID	Twarc	RAPID	Twarc	RAPID	Twarc
Nodes	3,594	4,591	5,198	6,205	1,492	1,507
Edges	7,344	10,110	14,802	18,184	1,560	1,576
Average degree	2.043	2.202	2.848	2.931	1.046	1.046
Density	0.001	0.000	0.001	0.000	0.001	0.001
Mean edge weight	1.096	1.087	1.245	1.222	1.099	1.098
Components	118	176	123	179	196	201
Largest component	3,308	4,085	4,854	5,612	1,073	1,080
- Diameter	12	11	10	10	15	15
Clusters	138	197	158	210	221	226
Largest cluster	471	727	1,090	1,513	122	123
Reciprocity	0.004	0.004	0.024	0.025	0.072	0.071
Transitivity	0.027	0.026	0.084	0.079	0.016	0.016
Maxmium k-core	9	10	11	14	3	3

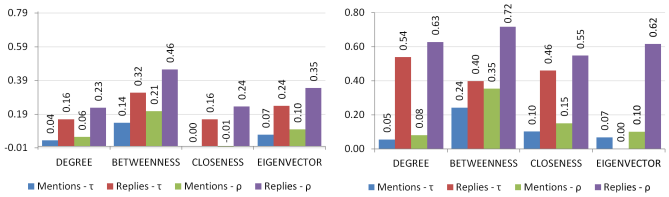
RAPID-E collected more than Twarc in Part 2 but not in Part 1. Filtered back to the keywords, many tweets were simply missed, but not many extra (i.e., unique to RAPID) were captured. Discussions with RAPID's developers revealed it captures but then discards tweets in which the keywords do not appear in textual fields (e.g., the message body, account profile or name). Based on manual inspection, RAPID-E captured relevant tweets in Part 2 missed by Twarc; it is unknown why.

Table II reveals that although feature counts vary significantly, many of the most common values are the same (e.g., most retweeted tweet, most mentioned account, most used hashtags). Notably, although the most prolific account is different in Part 2, the most mentioned account is the same for both Parts 1 and 2, potentially implying that account has had similarly high influence in both parallel datasets. Furthermore, both datasets shared almost all the same top ten hashtags.

C. Comparison of network statistics

Given the differences in datasets, we expect differences in the derived social networks (Tables III and IV) [11]. Each network is dominated by a single large component, comprising over 90% of nodes in the retweet and mention networks, and around 70% in the reply networks. The distributions of component sizes appear to follow a power law, resulting in corresponding high numbers of detected clusters.

Structural statistics like density, diameter (of the largest component in disconnected networks), reciprocity and transitivity may offer insight into social behaviours such as influence and information gathering. The high component counts in all networks lead to low densities and correspondingly low transitivities, as the potential number of triads is limited by the connectivity of nodes. That said, the largest components were consistently larger in the Twarc datasets, but the diameters of the corresponding largest components from each dataset were remarkably similar, implying that the extra nodes and edges were in the components' centres rather than on the periphery. This increase in internal structures improves connectivity and therefore the number of nodes to which any one node could pass information (and therefore influence) or, at least, reduces the length of paths between nodes so information can pass more quickly. The similarities in transitivity imply the increase may not be significant, however, with networks of these sizes. Reciprocity values may provide insight into information gathering, which often relies on patterns of to-and-fro communication as a person asks a question and others respond. Interestingly, the only significant difference in reciprocity is in the Part 1 retweet networks, with the Twarc dataset having a reciprocity nearly double that of the RAPID dataset (though still small). The Twarc dataset includes 60% more retweets than the corresponding RAPID dataset and 40% more accounts (Table I), which may account for the discrepancy. Given the network sizes, the reciprocity values indicate low degrees of conversation, mostly in the reply networks. Interestingly, mean edge weights are very low (1.3 at most), implying that most interactions between accounts in all networks happen only once, despite these being corpora of issue-based discussions.



(a) Part 1. (b) Part 2.

Fig. 2. Centrality ranking comparisons using Kendall τ and Spearman's ρ.

D. Comparison of centralities

Centrality measures can tell us about the influence an individual has over their neighbourhood. If networks are constructed from partial data, network-level metrics (e.g., radius, shortest paths, cluster detection) and neighbourhood-aware measures (e.g., eigenvector and Katz centrality) may vary and not be meaningful [11]. The Kendall τ and Spearman's ρ coefficients were calculated comparing the corresponding lists of nodes, each pair ranked by one of the four centrality measures (Figure 2). Although somewhat proportional, it is notable how different the coefficient values are, especially in Part 2. While Twarc produced more tweets than RAPID (Table I), and more unique accounts, the corresponding mention

and reply node counts are not significantly higher (Tables III and IV). In fact, the node counts in the Part 1 reply networks are correspondingly lower than in the Part 2 reply networks, even though both Part 2 datasets were smaller. Edge counts in the mention networks were very different (Twarc had many more) but were quite similar in the reply networks.

The biggest variation was in the mention networks from Part 1 (Table III), due to the large number of extra mentions from Twarc. It is notable that the Kendall's τ was low for all mention networks (Figure 2), especially for degree and closeness centrality. It is worth noting the minor differences in the degree and immediate neighbours of nodes impacts degree and closeness centralities significantly, and, correspondingly, their relative rankings. In contrast, rankings for betweenness and eigenvector centrality, which rely more on global network structure, remained relatively stable.

E. Comparison of clusters

We finally compare the networks via largest clusters. The ARI scores confirm that the reply clusters were most similar for Parts 1 and 2 (0.738 and 0.756, respectively), possibly due to the small size of the reply networks. The mention and retweet clusters for Part 2 were more similar than those of Part 1 (0.437 and 0.468 compared to 0.320 and 0.350), possibly due to the longer collection period. In Part 1, there is a chance the networks are different due to RAPID's expansion strategy. Changes to filter keywords may have collected posts of other vocal accounts not using the original keywords, at the cost of the posts which did.

V. Conclusion

We have summarised difficulties in modelling social media data for SNA and presented a method for comparing social media data as sources, evaluating it against two simultaneously collected datasets. The difference in number of accounts captured affects many measures the most, as this defines the maximum number of nodes in the accounts we created. Extra tweets resulted in extra edges, affecting network structure measures such as centrality and k-core maxima, but most prevalent content values remained similar. For example, the most mentioned account (arguably an indication of influence) remained the same in both parallel datasets.

The differences revealed in the Q&A datasets are a warning to OSN researchers to: be mindful of biases in their collection tools; construct filter conditions carefully to avoid over- or under-collection; and to monitor data integrity (e.g., check for connection failures).

A firm understanding of the data and how it was obtained may be vital depending on the nature of research being investigated, especially when they relate to sources of influence.

Acknowledgements

The work has been partially supported by the Cyber Security Research Centre Limited whose activities are partially funded by the Australian Government's Cooperative Research Centres Programme.

References

[1] F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley, "Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose," in *ICWSM*. AAAI Press, 2013, pp. 400–408.

[2] S. González-Bailón, N. Wang, A. Rivero, J. Borge-Holthoefer, and Y. Moreno, "Assessing the bias in samples of large online networks," *Social Networks*, vol. 38, pp. 16–27, 2014.

[3] K. Joseph, P. M. Landwehr, and K. M. Carley, "Two 1%s don't make a whole: Comparing simultaneous samples from Twitter's streaming API," in *SBP*, ser. Lecture Notes in Computer Science, vol. 8393. Springer, 2014, pp. 75–83.

[4] J. H. Paik and J. Lin, "Do multiple listeners to the public Twitter sample stream receive the same tweets?" in *TAIA*, 2015.

[5] D. Ruths and J. Pfeffer, "Social media for large studies of behavior," *Science*, vol. 346, no. 6213, pp. 1063–1064, 2014.

[6] R. Tromble, A. Storz, and D. Stockmann, "We don't know what we don't know: When and how the use of Twitter's public APIs biases scientific inference," in *SSRN*, 2017, pp. 1–26.

[7] J. Pfeffer, K. Mayer, and F. Morstatter, "Tampering with Twitter's sample API," *EPJ Data Science*, vol. 7, no. 1, p. 50, 2018.

[8] A. Olteanu, C. Castillo, F. Diaz, and E. Kıcıman, "Social data: Biases, methodological pitfalls, and ethical boundaries," *Frontiers in Big Data*, vol. 2, p. 13, 2019.

[9] D. Lazer, R. Kennedy, G. King, and A. Vespignani, "The Parable of Google Flu: Traps in Big Data Analysis," *Science (New York, N.Y.)*, vol. 343, pp. 1203–5, 03 2014.

[10] Z. Tufekci, "Big questions for social media Big Data: Representativeness, validity and other methodological pitfalls," in *ICWSM*. The AAAI Press, 2014.

[11] H. Holzmann, A. Anand, and M. Khosla, "Delusive pagerank in incomplete graphs," in *COMPLEX NETWORKS (1)*, ser. Studies in Computational Intelligence, vol. 812. Springer, 2018, pp. 104–117.

[12] D. Weber, M. Nasim, L. Mitchell, and L. Falzon, "A method to evaluate the reliability of social media data for social network analysis," *arXiv preprint arXiv:2010.08717*, 2020.

[13] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*. Cambridge University Press, 1994, vol. 8.

[14] M. Nasim, "Inferring social relations from online and communication networks," Ph.D. dissertation, Computer and Information Science, 2016.

[15] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *Science*, vol. 323, no. 5916, pp. 892–895, 2009.

[16] J. P. Bagrow, X. Liu, and L. Mitchell, "Information flow reveals prediction limits in online social activity," *Nature Human Behaviour*, vol. 3, no. 2, pp. 122–128, 2019.

[17] T. Venturini, A. Munk, and M. Jacomy, "Actor-network vs network analysis vs digital networks are we talking about the same networks?" in *Digital STS: A Handbook and Fieldguide*, V. Ribes and J. Vertesi, Eds. New York: Palgrave Macmillan, 2018, pp. 510–524.

[18] L. Falzon, C. McCurrie, and J. Dunn, "Representation and analysis of Twitter activity: A dynamic network perspective," in *ASONAM*. ACM, 2017, pp. 1183–1190.

[19] F. Morstatter, Y. Shao, A. Galstyan, and S. Karunasekera, "From *Alt-Right* to *Alt-Rechts*: Twitter analysis of the 2017 German Federal Election," in *WWW (Companion Volume)*. ACM, 2018, pp. 621–628.

[20] K. Lim, S. Jayasekara, S. Karunasekera, A. Harwood, L. Falzon, J. Dunn, and G. Burgess, "RAPID: Real-time Analytics Platform for Interactive Data Mining," in *ECML/PKDD (3)*, ser. LNCS, vol. 11053. Springer, 2018, pp. 649–653.

[21] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.

[22] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193–218, Dec 1985.