*Article*

# Dialoguing with Data and Data Reduction: An Observational, Narrowing-Down Approach to Social Media Network Analysis

**Jingrong Tong [1],\* and Landong Zuo [2]**

[1] Department of Journalism Studies, The University of Sheffield, Sheffield S1 4DT, UK
[2] Independent Researcher, London W1F 8PR, UK; landong.zuo@gmail.com
\* Correspondence: j.tong@sheffield.ac.uk

**Abstract:** In this article, we propose an observational, narrowing-down approach to analysing social media networks and developing research design by the joint use of computational algorithms and researchers' inductive exploration and interpretive explanations. The Brexit referendum on Twitter study is used to illustrate how we applied this approach in practice. In this study, observation helped us combine the strengths of computational statistical analysis and modelling and of inductive inquiries. Computational algorithms and tools including Elasticsearch, Kibana and Gephi provided us with an "ethnographic field" where we were able to inductively observe the relationships among users and to reduce the amount of data down to a level in which we could intuitively understand these relationships. In traditional observational studies, talking to human subjects and observing their interactions in a research site are important to ethnographers. Likewise, it is useful for social science researchers to dialogue with data, observe human relationships embodied in the data and reconstructed by computational tools, and understand these relationships through closely examining a small batch of meaningful data that is extracted from large-scale data. In this case study, adopting the proposed approach, we found the importance of political disagreement leading to a tale of two politicians, in which pro-Brexit users denounced @David_Cameron but legitimised @Nigel_Farage.

**Keywords:** Twitter; observation; computational algorithms; inductive analysis and approach; the Brexit referendum; political disagreement

## 1. Introduction

Social media offers great potential for researchers to investigate how people communicate and connect. However, the increasing ubiquity and enormity of social data have triggered debates among social science scholars in relation to how to study social media networks.

Traditional qualitative research methods used to study human relations and networks, such as observation, usually collect data about human interactions and involve an inductive exploration of the data. An inductive approach is data-driven and exploratory with the aim of building theory from exploring data. This means that by collecting and qualitatively exploring empirical data, researchers discover patterns in the data and interpret their meanings and implications for theory (Bryman 2012). An inductive analysis means "approaches that primarily use detailed readings of raw data to derive concepts, themes, or a model through interpretations made from the raw data by an evaluator or researcher" (Thomas 2006, p. 238). These methods and approaches have been adopted to collect and examine social media qualitatively. In related studies, scholars saw social media as a virtual site in which they qualitatively observed and interacted with their research objects (see, for example, Paccagnella 1997; Postill and Pink 2012; Hine 2000; Kozinets 2010; McEwan and Sobre-Denton 2011). These studies aim to find qualitative evidence and cultural meanings from human interactions on social media. While they are welcome, their approaches may not suit the analysis of large-scale social data. Computational tools are needed for inductively observing social networks and human interactions captured in large-scale social data.

In an iterative research process, through observation of human interactions and networks, revealed and visualised by computational statistical analysis and modelling, researchers can build up the knowledge of the research object under scrutiny. Their accumulated knowledge facilitates their judgement and enables them to design the research in a way of conducting close readings of data.

Scholars (see, for example, Lazer et al. 2009; Berry 2011, 2012) identify and celebrate "the computational turn" in social sciences and humanities studies. Computational algorithms have been developed to explore and statistically analyse large-scale social data. They work quite well to model and detect patterns in data (Fortunato 2010; Huang and Sun 2014; Takahashi et al. 2015). In particular, in social network analysis, computer algorithms can statistically calculate and model the connections between numerous users. However, in the cases where these computational algorithms and modelling are used to test pre-defined hypotheses scientifically from a deductive approach, the results of computational inquiries may be too rigid to reveal the compelling aspects of the dataset. A flexible, iterative research process and an inductive exploration of the data will be needed to mitigate the rigidity of hypothetico-deductivism. Qualitative, inductive inquiries can allow researchers the flexibility to develop and refine the research design in the process of data exploration and help them gain a rich understanding of the meanings of the interactions between users revealed by computational inquiries. Therefore, it would be helpful if we can combine the use of computational algorithms and the involvement of researchers' inductive exploration and interpretive explanations in researching social data. A question arises as to how to combine them.

With this in mind, in this article, we propose an observational, narrowing-down approach to analysing social media networks by the joint use of computational applications and inductive inquiries. The Brexit referendum on Twitter study will be used to illustrate how we used this approach in practice. We will first review the related work conducted in the field about social network analysis before introducing the proposed approach. We will then discuss the case study and how the approach was applied. The article concludes with reflections on this approach's usefulness and limitations and outlines some suggestions for future research. This research contributes to the literature of social network analysis and big data analysis by contending the importance of involving qualitative, inductive inquiries and reasoning along with the use of computational algorithms through observation and data reduction in large-scale social media analysis.

### 1.1. Social Network Studies and the Current Challenges

With a root in sociology, social network analysis (SNA) examines human relationships and interactions (connections) between actors (nodes) in order to understand the social structure and the roles of actors in that structure (Wasserman and Faust 1994). Qualitative research methods such as observation are commonly used in social network analysis in disciplines such as sociology and anthropology (Scott 2017). Recently, with the rise of social media, scholars increasingly applied SNA to analysing social media with the assistance of computer tools, although SNA has been criticised for only offering "static snaps while neglecting the networks dynamics" (Bruns and Stieglitz 2013; Bruns 2011). The literature is mostly based in the field of computer science. Over recent years, it has gradually expanded to that of social science.

There are three ways of using SNA in a small but growing number of social media studies. Firstly, SNA is used to understand social actors' influence and roles on social media (see, for example, Jörgens et al. 2016) or to identify opinion leaders in networks (see, for example, Dubois and Gaffney 2014; Wukich and Steinberg 2013; Xu et al. 2014). SNA studies also explore social media users' social networking strategies during events such as elections, crisis and disasters (see, for example, Yoon and Park 2014; Murthy and Longwell 2013; Samuel-Azran and Hayat 2017; Ogan and Varol 2017). Finally, SNA is employed to identify and comprehend the formation of networks, communities or campaigns on social

media (see, for example, Chatfield et al. 2015; Grandjean 2016; Antonakaki et al. 2016; Himelboim et al. 2017; Bonini et al. 2016; Lin et al. 2008).

These social network studies used computational algorithms to measure network metrics, such as density, modularity, degree, betweenness, centralisation, and clustering coefficient, so as to identify and understand communities formed on social media (see, for example, Himelboim et al. 2017; Hansen et al. 2011; Grandjean 2016). In terms of methodological approaches, a group of studies took a deductive approach, i.e., they developed research questions and hypotheses from the literature review and used them to guide their research design and data analysis. In these studies, while the design of these inquiries was research problem-oriented, the results of the network metrics were intended to answer specific research questions effectively or test hypotheses, as exemplified in the studies of predicting opinion leaders (Xu et al. 2014; Samuel-Azran and Hayat 2017; Wukich and Steinberg 2013; Yoon and Park 2014; Dubois and Gaffney 2014; Murthy and Longwell 2013). However, deductive studies are criticised for being too prescriptive, and there is a risk that the priorly determined conceptual frameworks or specific research questions may be invalid for understanding the data (Eriksson and Lindström 1997). They may also be unable to interpret unanticipated but meaningful patterns observed in data analysis (Quinn and Dunham 1983). This problem may even be worse in social data analysis, as the large scale of social data may restrict researchers from understanding what is most worth researching.

Some studies took a more explorative approach, exemplified by those that used statistical measures to form ideas about communities and networks (Dugue and Perez 2014; Antonakaki et al. 2016). Although stressing on the importance of exploring data, these studies mostly rely on computational methods and largely exclude the involvement of researchers' inductive reasoning. Other studies adopted mixed methods, comprising computational social network analysis and qualitative methods such as content analysis of a small number of tweets, so as to gain a deep understanding of the topic (Jörgens et al. 2016; Bonini et al. 2016; Chatfield et al. 2015). However, these studies still mainly test their research models or assumptions arising from the literature review (see, for example, Chatfield et al. 2015). They are not genuinely exploratory or inductive studies, which require scholars to be data-driven and develop research based on the exploration of empirical evidence in data analysis. Given the complexity of social networks and relationships, an inductive approach will be useful. It enables researchers to explore the data to grasp the meanings and discover the knowledge of the relationships. Therefore, it will be helpful to develop an approach that can accommodate inductive inquiries and apply computational algorithms in research.

*1.2. The Promise Offered by an Observational, Narrowing-Down Approach*

Some more recent studies (see, for example, Burgess and Matamoros-Fernández 2016; Bruns et al. 2020) have combined the use of computational tools and qualitative content analysis methods in exploring large-scale social media data. We would like to further take this explorative approach by arguing for the importance of both observation and data reduction in researching large-scale social data. They are important not only for data mining and analysis but also for research design. Observation is one of the common methods that are used to collect data about and understand the relationships and interactions among actors in social network analysis (Wasserman and Faust 1994). In traditional observational studies, observation happens in the stage of data collection. Observation means researchers—either as "a participant observer or a direct observer"—observe the social relations of a group and its members and record and analyse their observation systematically (Scott 2017). Although substantial time is required for fieldwork, observation enables researchers to obtain "an understanding of the cultural meanings of relationship" (Scott 2017, p. 45).

Observation has been used in studying interactions and relationships on the Internet in general and on social media in particular (Hine 2000; Postill and Pink 2012; Kozinets 2010).

These studies see social media as a research site where researchers virtually observe or participate in the activities of the observees; they recognise the intertwinement of the online and offline activities and the relationships of these participants. Therefore it is "digital socialities" that virtual ethnographic researchers should analyse (Postill and Pink 2012, p. 127). In other terms, researchers observe online interactions and relationships, such as "following"/"followed", "sharing" and "like", on particular social media platforms and on some occasions they engage with online interactions, such as getting into the "following" relationship or sharing posts posed by participants.

These virtual ethnographic studies and approaches are suitable for qualitative researchers to collect data through observation qualitatively—either online or offline—and analyse the data to form an understanding of the behaviours of their participants. These studies point to the importance of virtual observation. We contend that, for social network studies involving the analysis of large-scale social data, it is also vital to use observation in the data exploration and analysis process for the purpose of data mining, data reduction and research design.

In the virtual ethnographic studies discussed above, researchers observe interactions and relationship presented on the Internet by experiencing or witnessing them. However, when analysing large-scale social data, the interactions and relationships among social media users are captured by the data but remain invisible to researchers, unless they use computer tools to model, map, and then observe them. Such observation occurs in the data mining and analysis stage rather than in the stage of data collection. Besides, observation is aimed not only at understanding relationships but also at developing and refining research design.

The mapping and modelling process is dynamic, recursive and inductive. In the iterative process, the interactions occur between researchers and data. In other terms, the data has already been collected, and the objects of the observation are social networks and statistical figures generated and presented by computational tools in iterative queries. Therefore, researchers need to identify and understand such interactions among participants captured in the data by observing the findings from the statistical analysis and modelling generated by computer application tools. They can then use this understanding to inform their research design, formulate research questions and narrow down the scale of the data to focus on a small batch of data which they can handle qualitatively. This narrowing-down process is that of data reduction, which refers to a process of reducing the amount of data so that researchers can make sense of the data (Bryman 2012). Data reduction is crucial for both quantitative and qualitative research. In social network analysis, part of the aim of observation is to gradually narrow down the data to a manageable scale, which allows researchers to have detailed, interpretive readings of data.

There are two noticeable benefits of using observation in analysing large-scale social data. Firstly, it enables researchers to inductively explore the data and have the flexibility to design the research. That is to say, researchers can develop the next step's analytical tasks and research questions based on their understanding obtained from the observation of the results of previous computational inquiries. This flexibility releases researchers from being prohibited by pre-decided research questions or theoretical frameworks but to be led by the data to raise research questions and identify meaningful relationships and interactions. Secondly, while observing the relationships and interactions visualised by computer tools, researchers are able to use their inductive reasoning to identify and focus on the relationships and interactions of a relatively small group of users, which are deemed as meaningful.

*1.3. Six Stages to Dialogue with Data by Using Computational Tools*

This article proposes six stages to interact with and narrow down data by using computational methods and inductive inquiries so that researchers can develop an inductive understanding of social networks in large-scale social data. As the research goes on, researchers can use the knowledge gained in the process to design research aimed at

answering research questions that can help the researchers interpret the data meaningfully. These stages are discussed below.

1. Overall exploration of data

In this stage, data exploration aims to gain a general understanding of data. Key aspects of the data should be explored and observed, such as trends over time in the data, volumes of posts, most popular posts, most popular users, most active users, and average retweets and likes.

2. Identifying key social media users

The findings from the first stage should be able to direct us to key social media users. The analysis of the most popular and active users, for example, can tell us who they are so that we can focus on them for the social network analysis in the next stage. The data of the identified key users need to be manually checked and cross-checked against information coming from other sources such as media coverage, to determine whether or not these are users most relevant to the topic under scrutiny and are not spammers. If some of them are irrelevant or even spammers, the related data need to be removed from the database, followed by repeating the analysis in the first stage, so that "genuine" key users can be identified.

3. Initial social network analysis

After identifying and selecting key social media users, researchers can conduct an initial social network analysis. In this stage, some exploration needs to be made so that researchers can observe the overall interactions between users. The exploration aims to identify key nodes in these networks. Among others, key metrics include weight (of a node or an edge), in-degree, out-degree, betweenness centrality, clustering coefficient, graph density (Gama and Gama 2012; Cherven 2015). The values of weight and degree tell us about the levels of activity and popularity of a node. Betweenness centrality shows the extent of a bridging role played by a node in the network, while the values of clustering coefficient and graph density reveal the level of completeness and the density level of a network.

4. Key node identifier

By observing the above-discussed key statistics, such as the weight of a node and an edge, and the out-degree and the in-degree scores, researchers should be able to identify key nodes in these networks. In this stage, the manual checking and cross-checking conducted in the first stage should be repeated to spot nodes, which are prominent but irrelevant to the topic or are actually spammers. By now, the understanding of the data gained should be able to help shape and formulate the research questions.

5. Mapping the networks surrounding identified key nodes

After having decided key nodes, researchers can focus on analysing and observing these key nodes' networks. This analysis can give researchers more information about the users who closely follow or interact with these key nodes. In this stage, researchers draw attention to a subset of data, embodying these networks and nodes, extracted from the original large-scale dataset.

6. Qualitative analysis

To gain a more qualitative understanding of the nature of the networks and key nodes' followers, researchers can conduct a qualitative analysis of the background and tweets of those who followed and interacted with key nodes. Depending on the project's need, researchers can choose to use appropriate qualitative methods such as thematic analysis and discourse analysis to analyse the content of the tweets. If necessary, researchers may also want to use computational analysis to complement the qualitative analysis.

After going through the proposed six stages, researchers will be able to extract small-scale, meaningful data from large-scale social data and focus on it so that they can answer research questions developed in the whole process. This subset of data can map the core structure of social networks surrounding key users captured in the entire dataset. As being

small-scale, this batch of data can be explored in great detail. However, a focus on key users may rule out the data associated with less prominent users from the analysis. Besides, it would be wrong to assume and claim that this small batch of data can represent all data in the dataset.

These six stages resemble an observational process, where researchers can dialogue with data and zoom in to explore a smaller size of data with the assistance of computational applications. Their inductive inquiries and observation of the outcome of statistical calculations and modelling can help them develop intuitive insights into the relationships and interactions between users captured in the data. Their gained insights can help design their research. In the process, researchers are observers, whose judgement is led by data, while the objects of their observation are the patterns found in the data. In so doing, researchers can combine the practical and theoretical dimensions of "the computational turn" (Burgess and Bruns 2015), and gain a meaningful understanding of the data.

### 1.4. The Brexit Referendum on Twitter Study[1]

#### 1.4.1. Data Description

To illustrate the six stages in practice, in this section, we present our study of analysing tweets about the United Kingdom (UK)'s 2016 EU referendum as a case study. The referendum took place on 23 June 2016. Its result was 51.9% in favour of Leave against 48.1% voting for Remain, with a profound impact on global politics. We collected and archived tweets in real-time through the Twitter Streaming Application Programming Interface (API) between 24 May and 23 June 2016. Only tweets that were open to the public were collected, and they contain any of the seven hashtags, i.e., "#Referendum", "#VoteLeave", "#VoteIn", "#EUref", "#VoteOut", "#VoteStay", and "#Brexit". The choice of hashtag was made based on our analysis of the use of hashtags in a small batch of tweets collected manually and our related observation on Twitter at the start of data collection. We acknowledge that our data may be influenced by our hashtag choice, and our data are not representing nor containing all data on this topic generated on Twitter one month prior to the referendum. After several rounds of cleaning, there are 12,644,199 tweets in the dataset used for analysis.

#### 1.4.2. Research Procedure

We wanted to understand the nature of social networks captured in the data. However, given that we had little knowledge of the data before actually analysing it, at the beginning of the research, we did not want to decide our focus and research questions. Therefore, we wanted to keep the options open and wanted to determine our focus and develop and complete the research design in the process of exploring and observing the data.

In stage 1, in our initial exploration of the whole dataset, we used the combination of Elasticsearch[2] and Kibana[3] to outline the overall trends and the most popular users. The number of tweets—including the number of original tweets published by an account and that of retweets its tweets had received in the dataset—determines the level of popularity of that account.

When we moved to Stage 2, after observing users appearing in the top 1000 most popular users, we decided to focus on the Twitter handles of nine British politicians and the thirteen Twitter accounts of British news outlets (Table 1). This is because they were the most prominent, popular users in terms of receiving retweets and likes. The nine UK politicians were important members of their political parties and had different attitudes towards the referendum (Table 2). By doing so, our attention was drawn to the networks surrounding these politicians and news media.

---

[1] A detailed report of this case study can be seen in our forthcoming book: *The Brexit Referendum on Twitter: A mixed-method, computational analysis*, which will be published by Emerald Publishing Limited in 2021. Some content in this section, particularly that under "1.4.3. Findings of the Case Study", will be included in the book.

[2] See "What is Elasticsearch" at https://www.elastic.co/what-is/elasticsearch (accessed on 7 July 2020).

[3] See "What is Kibana used for" at https://www.elastic.co/what-is/kibana (accessed on 7 July 2020).

**Table 1.** Twitter handles of 9 key British politicians and their political affiliations.

| Names | Political Parties | Twitter Handles |
|-------|-------------------|-----------------|
| Boris Johnson | Conservative | @BorisJohnson |
| Caroline Lucas | Green | @CarolineLucas |
| David Cameron | Conservative | @David_Cameron |
| George Osborne | Conservative | @George_Osborne |
| Jeremy Corbyn | Labour | @jeremycorbyn |
| Nick Clegg | Liberal Democrats (LibDem) | @nick_clegg |
| Nicola Sturgeon | Scottish National Party (SNP) | @NicolaSturgeon |
| Nigel Farage | United Kingdom Independence Party (UKIP) | @Nigel_Farage |
| Sadiq Khan | Labour | @SadiqKhan |

**Table 2.** Thirteen Twitter handles of key British news media.

| News Media | Twitter Handles |
|------------|-----------------|
| BBC | @BBCNews |
| Channel 4 | @Channel4News |
| ITV | @itvnews |
| Sky | @SkyNews |
| Daily Express | @Daily_Express |
| Daily Mail | @DailyMailUK |
| Daily Mirror | @DailyMirror |
| Guardian | @guardian |
| Independent | @Independent |
| Daily Telegraph | @Telegraph |
| Sun | @TheSun |
| Sunday Times | @thesundaytimes |
| Times | @thetimes |

In stage 3, we then did a network analysis on Gephi[4] based on the retweeting relationship between these users. Gephi is an open-source Java application that has been used in analysing social networks on social media such as Twitter (Bruns 2011; Bingham-Hall and Law 2015; Burgess and Matamoros-Fernández 2016). We observed the statistic figures, such as those of the weight of nodes, out-degree and in-degree. Take the out-degree scores for example. If we did not use out-degree[5] to focus on the core networks, the networks were very complicated and intricate. Although we can see some hub-and-spoke clusters, the networks overall are too complex to understand (see Figure 1). We repeated the tests again and again by adjusting different out-degree scores. For example, if we set the out-degree score threshold at 30, we had the core networks between users with out-degree ≥30 (see Figure 2). Compared with Figure 1, Figure 2 is more accessible and easier for us to make sense of the relationship between these users. Figure 2 shows that, except for @BBCnews and @SkyNews, none of the other news media's Twitter accounts played a key role in these networks. Most of them were even isolated or excluded from the core networks, the backbone of the networks surrounding these politicians and news media. Three politicians, @NicolaSturgeon, @David_Cameron and @Nigel_Farage, stayed in the core networks. However, when we further increased the out-degree score threshold, we found that only @David_Cameron and @Nigel_Farage were left on the core networks.

---

[4] https://gephi.org/.

[5] The value of out-degree is the amount of times/frequencies of users' retweeting of the messages sent by the Twitter account under examination.
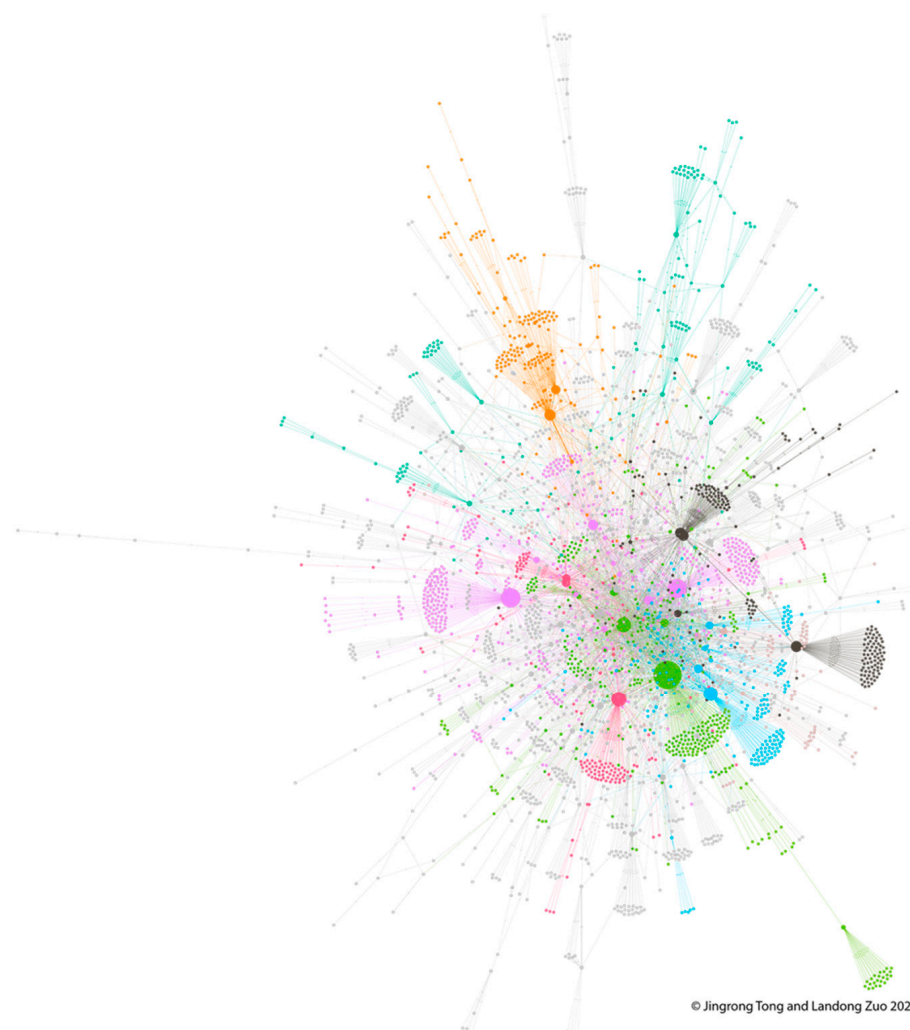
**Figure 1.** Networks surrounding 13 news media handles and 9 politician handles.

In stage 4, we needed to choose our focus. After consulting the related literature on political communication on the Internet, we decided to focus on analysing the networks surrounding @David_Cameron and @Nigel_Farage. We tried to find out how Twitter users retweeted them, in particular, whether and to what extent the two politicians were surrounded by supportive networks on Twitter. Our decision was partly based on our observation (gained at the previous stage) of the core networks surrounding the Twitter handles of these news media and politicians, which suggests a clear opposition between the core networks of Nigel Farage and those of David Cameron. In other terms, Nigel Farage and David Cameron were the two most influential nodes in the networks surrounding the news media and politicians. The fact that the two politicians held opposite attitudes towards the Brexit referendum was another reason for our focus on them.

Then, in stage 5, we analysed the core networks (out-degree ≥30) surrounding @David_Cameron and @Nigel_Farage (see Figure 3). We coded the attitudes of the users, identified in the core networks, toward the referendum. We, thus, had the attitudinal graph of core networks of the two politicians' Twitter handles. The graph clearly shows us that most users in the core networks were pro-Brexit users. At last, in order to understand the nature of the networks surrounding the two politicians on Twitter, we conducted a thematic analysis of the tweets published by the key users in the networks identified. The coding and analysis process proposed by Braun and Clarke (2006) was used in our thematic analysis carried out in Nvivo. The tweets were coded with a focus on the users' attitudes toward and comments on the politicians and the referendum. When we deter-

mined their referendum stances and attitudes towards the two politicians, we also took into consideration other features of their tweeting activities such as the use of hashtags.
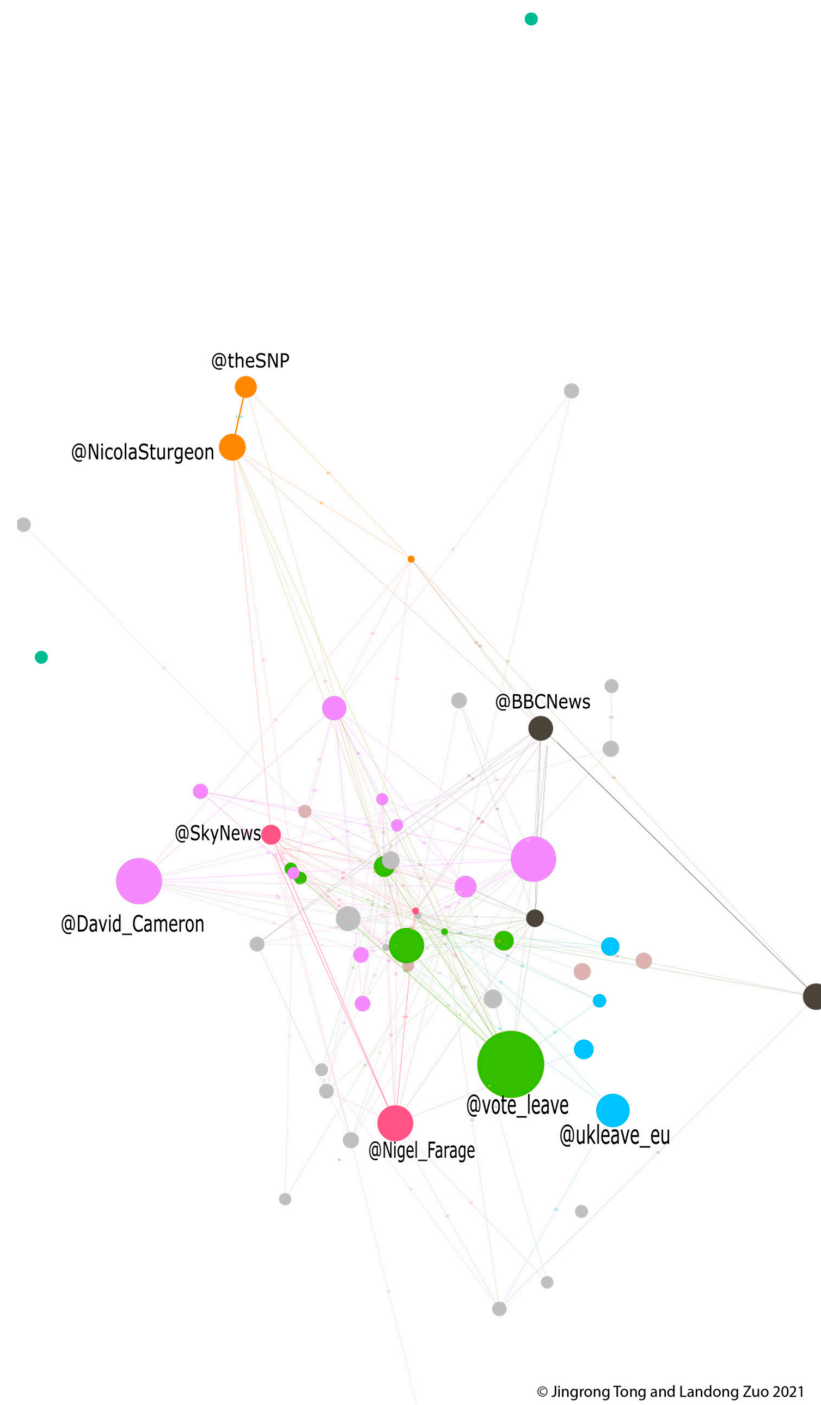


© Jingrong Tong and Landong Zuo 2021

**Figure 2.** Core networks surrounding 13 news media handles and 9 politician handles (out-degree ≥30).
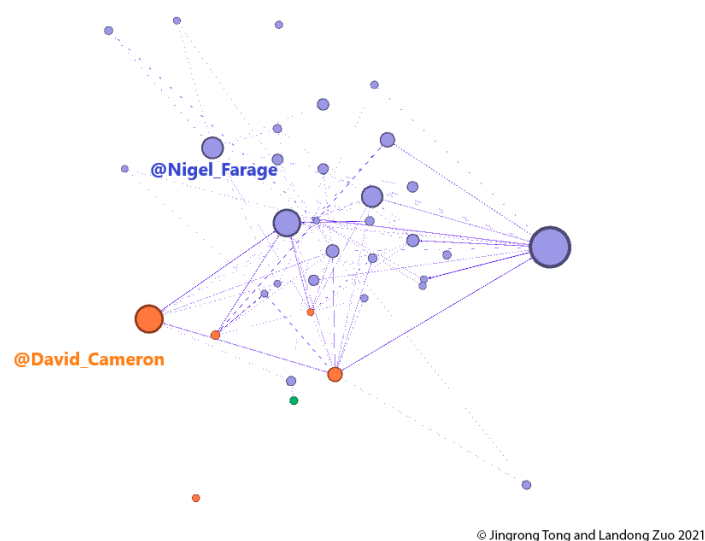
© Jingrong Tong and Landong Zuo 2021

**Figure 3.** The attitudinal graph of the core networks surrounding @David_Cameron and @Nigel_Farage (purple: pro-Brexit; orange: pro-Remain; green: neutral).

### 1.4.3. Findings of the Case Study

An established argument in the literature about social media communication is that politically alike people tend to connect with one another (Gerber et al. 2013; Huber and Malhotra 2017; McPherson et al. 2001). The findings of this study partly resonate with this argument. The findings, however, also point to the importance of political disagreement among Twitter users.

In @David_Cameron's core networks, most of the key users (seven out of eight users) were pro-Brexit users, who criticised and delegitimised David Cameron (see Figure 3). The language of these users was (negatively) emotional with inadequate reasonable arguments. Some tweets even contained personal attacks against David Cameron. The qualitative thematic analysis of their tweets reveals seven themes (see Table 3), which suggest their evident ideological confrontation with David Cameron and his Remain arguments. What is interesting is, while seeing David Cameron as a failure and loser, these key users, who frequently retweeted @David_Cameron, instead hailed Nigel Farage as a hero and winner, as exemplified in this tweet: Michael Gove made @David_Cameron suffer this week. @Nigel_Farage will finish you next week! YUMMY!! #InOrOut #VoteLeave (rephrased).

**Table 3.** Themes in the tweets published by (key) Leave users in @David_Cameron's core networks.

| Theme Numbers | Themes | Example Tweets (All Rephrased Except the First Tweet) |
|:---:|:---|:---|
| 1 | Self-declaring to be Brexiteers; calling to leave and take back control to "restore democracy" | No @David_Cameron Britain doesn't give up. We are determined to bring back democracy to this country #VoteLeave (by @Vote_LeaveMedia) (7 June 2016) |
| 2 | Labelling David Cameron as a traitor, liar and scaremonger | what a liar David Cameron is (22 June 2016) |
| 3 | The failure of David Cameron and the EU | @David_Cameron finally failed to reform the EU which could not be fixed #InOrOut #VoteLeave (2 June 2016) |
| 4 | Poor media performance of David Cameron | @David_Cameron is suffering and cracked on Friday #BBCdebate #IndependenceDay #VoteLeave (21 June 2016) |
| 5 | Only David Cameron and rich people wanted to remain | @David_Cameron Your policy is to let the rich benefit from Remain, but leave the poor to suffer austerity. I distrust you. I #voteleave! (17 June 2016) |
| 6 | David Cameron failed to keep immigration under control and lied about immigration | @David_Cameron is a liar: four years ago, (he) knew he would not meet the immigration target (22 June 2016) |
| 7 | David Cameron helped Turkey join the EU | @David_Cameron @Conservatives You did not spend our money on our NHS; instead you used it to help Turkey join the EU #VoteLeave (9 June 2016) |

In @David_Cameron's core networks, only one key user (@StrongerIn) supported Remain and David Cameron and condemned the Leave Campaign. For example, it applauded David Cameron's performance in TV debates for making "a passionate case for why we're better off IN Europe" (30 May 2016). By contrast, this account criticised the Leave campaign as they "'just don't know' what happens after if we leave" (30 May 2016).

By contrast, in the networks surrounding @Nigel_Farage, 7 out of 8 key users supported Nigel Farage and his Brexit stance (see Figure 3). He was thanked and praised for saving democracy and bringing hope to the UK. These key users were his ardent supporters, and some were members of the UKIP. Their tweets, highly praising Farage, contain six main themes (see Table 4).

**Table 4.** Themes in the tweets published by (key) Leave users in @Nigel_Farage's core networks.

| Theme Numbers | Themes | Example Tweets (All Rephrased) |
|:---:|:---|:---|
| 1 | Self-declaring to be Brexiteers (and even UKIP members) and calling to vote leave | Nigel, well done for being a real patriot! Being a member of the UKIP makes me feel proud. Thank you (10 June 2016) |
| 2 | Excellent performance of Farage in debates and his successful campaign | @Nigel_Farage outperformed David Cameron. In contrast to his open, honest responses, those of David Cameron and George Osborne were aggressive, suggesting them seeing (us) inferior to them #VoteLeave (12 June 2016) |
| 3 | Farage is a hero who tells the truth | @Nigel_Farage, a star, always tells the truth. This is the way of setting us free from the control of the EU #VoteLeave (10 June 2016) |
| 4 | Remainers want to betray our country and give it away | @Nigel_Farage Remainers are unfaithful to this country and nation, which we built, defended and sacrificed our lives for. I vote #Brexit (23 June 2016) |
| 5 | The advantages of Brexit | @Nigel_Farage YES, after Brexit, the exports will be cheaper #BREXIT (12 June 2016) |
| 6 | Condemning immigrants, immigration, and Muslims | #Brexit #UKIP #VoteLeave @Nigel_Farage Muslims pose a threat, like what we have found in India, Leave #EU (6 June 2016) |

The core networks of @Nigel_Farage only had one pro-Remain account: @StrongerIn-Press, which criticised Nigel Farage for ignoring the ramifications of Brexit for "working families", tariffs, "the single market", "the economy" and "sterling" and for lacking knowledge and being unable to answer questions appropriately.

Our analysis reveals that, in this case of profound political polarisation, political disagreement led to an ideological confrontation in users' responses, particularly those of pro-Brexit users, to the two politicians. Hailing @Nigel_Farage as a hero, pro-Brexit users put @David_Cameron under siege and denounced him as a "liar". The technological affordances of Twitter enabled them to actively declare stance, legitimising Nigel Farage, but delegitimising David Cameron. They frequently retweeted and mentioned @David_Cameron in their tweets because they disagreed with him. The strategic Twitter practice of pro-Brexit users gives rise to questions about the social media platform's role in the referendum.

## 2. Discussion

In this case study, we took an observational, narrowing-down approach to identifying the influential actors and their connections to other users in the dataset; the understanding obtained in this process helped us develop our research design. With the use of computational tools such as Elasticsearch, Kibana and Gephi, we brought inductive reasoning and judgement into the statistical calculation and modelling process of data analysis. The computational statistical analysis, modelling, and visualisation of the data enabled us to observe and intuitively understand users' relationships and interactions. Observation in our study was an approach in which we could combine the strengths of both computational statistical analysis and inductive inquiries through interacting with data.

Elasticsearch, Kibana and Gephi were the "ethnographic field" where the observation took place. The combination of Elasticsearch and Kibana, which are usually used for data

analysis in the information industry, provides us with a useful, user-friendly platform for exploring and observing the overall trends of the data as the starting point of data explorations. Gephi, which is designed for exploratory data analysis, is a good platform through which to observe the networks revealed in data. It enables researchers to interact with the visualisation of networks.[6] It allows researchers to explore different features of social networks by iteratively running different algorithms and metrics in Gephi and examining the visualisation of the outcomes. Important algorithms and metrics in Gephi include ranking (ranking can be done by various measures, e.g., degree, betweenness centrality and weight), attributes of nodes and edges, layouts, statistics, and filters. Computational data exploration in Gephi can be complemented and supported by manual data analysis such as profile categorisation.

In data exploration and analysis, the use of different algorithms and metrics, complemented by manual data analysis, contributes to observation in two ways. First, visualising the outcome of every data analysis query (in Kibana and Gephi) can construct and model the relationships among nodes (users) by highlighting specific features of the relationships. The visualisation thus presents the different attributes of the relationships so that researchers can intuitively observe how nodes are connected and gain in-depth insights of these relationships.

Second, in the observation process, exploring Twitter networks by using different algorithms and metrics can direct researchers' attention to key actors in the networks. It can narrow large-scale data down to "small" data. Due to the inclusion of too many nodes and edges, the social networks of large-scale social data can be incredibly complicated and difficult to examine and understand. For a social network analysis of large-scale social data, it is thus crucial to identify "small" but "key" data as an important means of data reduction. In this specific context, "small" refers to small pieces of facets that are extracted, derived and reasoned from the repository of large-scale data. By "key", we mean such facets can provide meaningful data insights for researchers. For example, in the case study, @David_Cameron's and @Nigel_Farage's core networks were mapped from the "small" data extracted from the large-scale referendum dataset, and this "small" data was considered to be "key". The visualisation of these core networks was vital for understanding the networks surrounding them in the dataset. Although we cannot regard the extracted "small" data as representative of the whole dataset, it can offer us an insightful understanding of the dataset. The judgement of which "small" data to extract is gained from the previous observations of the patterns emerging in the data analysis, as getting this "small" data was a result of the process of narrowing down data through the six stages.

These two contributions of the observational approach make it possible to design and conduct an inductive exploration of large-scale social data. The patterns presented through visualisations and the explicit representation of the relationships between key social actors in the core networks need to be interpreted by contextualising them. Partly this is because users' use of Twitter is socially constructed: for example it is influenced by the UK's existing politics. Partly it is because users' social capital and their social positions affect the level of attention they can receive from other users in terms of how many times their tweets can be retweeted, forming the retweeting connections among users. The observational approach allows researchers to find the patterns in the data, contextually interpret them and feedback the interpretations into the next round of data mining and analysis, looking for more patterns. This whole process shapes the research design.

The observational, narrowing-down approach used in the present study comes in a context where "the computational turn" in social science research coexists with its critiques due to the rising recognition of the need for thick social data for social science research (Hand 2014). We agree that it is essential to bring in researchers' inductive reasoning to grasp the cultural meanings created in the context where the data are produced and negotiated. Thus, while computational applications and methods are necessary for large-

---

6 See detailed introduction to Gephi and its features in https://gephi.org/features/.

scale social data analysis, their use should not exclude the involvement of human intuitive reasoning and judgement. The insights into an observed object should come from the combination of applying both. We acknowledge the value of observation and contend that it is essential to turn the data exploration and analysis process into an observational one where human intuitive reasoning can be involved along with the use of computational tools. It is crucial to immerse researchers in data by interacting with the data via computational applications and observing the results from every single data query, as if we were observing human subjects in ethnographic fieldwork.

Our study's findings echo the arguments of previous studies about the usefulness of employing both computational and qualitative techniques in analysing media and social media content (see, for example, Lewis et al. 2013; Starbird et al. 2019; Bruns et al. 2020; Burgess and Matamoros-Fernández 2016). Besides, our study takes these arguments further. It contends that the combination of computational and qualitative methods enables social science researchers to dialogue with data and reduce data to a level in which researchers can conduct their inductive inquiry of large-scale data and design research to find a meaningful story from the data.

Our findings also confirm the importance of "computational reflexivity" and inter-disciplinary collaboration (Ophir et al. 2020). In Ophir and his collaborators' study, this importance is shown in the computational analysis process, particularly in terms of bringing in the insights of social scientists—the ethnographer—in interpreting the results generated by the computer algorithms. Unlike their study, our research also stresses the role of the joint use of human intuitive reasoning and computational algorithms in the research design and data-reduction process. This is a process of narrowing down large-scale data and directing to meaningful subsets of the data that can be analysed in detail to answer research questions developed as the analysis goes along. Identifying small batches of data is thus a must-have for inductively analysing large-scale data. With the assistance of computational tools and algorithms, social science researchers' observations make this achievable. At this point, their capability of using computational tools and methods is crucial. Ideally, social science researchers need to be able to explore data themselves rather than mainly relying on computer or data scientists to feed them with the findings from the analysis.

However, this observational, narrowing-down approach to social network analysis has two major limitations. Firstly, it may be limited by computational ability. For example, in our study, we only mapped the core networks—rather than the entire networks—of users in the whole dataset. The networks presented in our analysis were, therefore, merely part of the whole networks. While whether the rest of the networks are important and worth researching is debatable, one thing for sure is that it would require extensive time and infrastructure resources to map the full networks. Secondly, although based on the statistic calculations and modelling which is assumed to be objective, researchers' insights depend on how the data are explored, the design and choices of algorithms and metrics, and the interpretations of the patterns found in the analysis. In this approach, using computational algorithms and metrics has at least one thing in common with using any other social science methods such as interviewing: the interpretive nature of the research. In interviews, we interview people trying to understand what they think about or do and the reasons behind their thoughts and actions. Likewise, in social data analysis, we use computational algorithms and metrics with an attempt to understand the studied object and to find a story in the data. The gained understanding, however, is influenced by human choices of algorithms and metrics and contextual interpretations. That is to say, the knowledge discovered in the process is not "the truth" but just one version of "the truth".

Through a case study of Twitter communication, this article has demonstrated how this observational, narrowing-down approach works. When it comes to other social media platforms, in terms of the importance of involving inductive reasoning in the analysis process and research design, the principles of observation and data reduction should still apply. However, the technical structures of social media platforms such as Facebook and Instagram are slightly different from that of Twitter. For example, Facebook requires users

to confirm request connections. Therefore, its networks may more likely reflect users' relationships in their offline lives than those of Twitter, which does not have this connection requirement (Bossetta 2018). Accompanying these structural differences are variations in relation to the privacy levels of different social media platforms, users' interactions and purpose of using a particular social platform. These variations may influence social science researchers' research focus and criteria for meaningful findings. In each of the six stages the actual research tasks may need to be adjusted to suit the research aims. Therefore, the question of whether, and if so, the extent to which, the approach can be applied to other social media environments awaits to be tested in future studies. Other questions arising from the present study are how meaningful computational queries can be and how much human experience and inductive reasoning are needed in social media research in particular and in general big data analysis. The need to explore these questions comes not only from the patchy nature of (social media/big) data but also from the urgency of understanding how to make the best of big data and create new knowledge for society. These questions and problems however are beyond the discussion scope of this article and thus require future studies to explore.

## References

Antonakaki, Despoina, Iasonas Polakis, Elias Athanasopoulos, Sotiris Ioannidis, and Paraskevi Fragopoulou. 2016. Exploiting abused trending topics to identify spam campaigns in Twitter. *Social Network Analysis and Mining* 6. [CrossRef]

Berry, David M. 2011. The computational turn: Thinking about the digital humanities. *Culture Machine* 12.

Berry, David M. 2012. Introduction: Understanding the digital humanities. In *Understanding Digital Humanities*. Edited by David M. Berry. London: Palgrave Macmillan, pp. 1–20.

Bingham-Hall, John, and Stephen Law. 2015. Connected or informed?: Local Twitter networking in a London neighbourhood. *Big Data & Society* 2. [CrossRef]

Bonini, Tiziano, Alessandro Caliandro, and Alessandra Massarelli. 2016. Understanding the value of networked publics in radio: Employing digital methods and social network analysis to understand the Twitter publics of two Italian national radio stations. *Information, Communication & Society* 19: 40–58. [CrossRef]

Bossetta, Michael. 2018. The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election. *Journalism & Mass Communication Quarterly* 95: 471–96.

Braun, Virginia, and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3: 77–101. [CrossRef]

Bruns, Axel. 2011. How long is a tweet? Mapping dynamic conversation networks on Twitter using Gawk and Gephi. *Information, Communication & Society* 15: 1323–51.

Bruns, Axel, and Stefan Stieglitz. 2013. Towards more systematic Twitter analysis: Metrics for tweeting activities. *International Journal of Social Research Methodology* 16: 91–108. [CrossRef]

Bruns, Axel, Stephen Harrington, and Edward Hurcombe. 2020. Corona? 5G? or both?': The dynamics of COVID-19/5G conspiracy theories on Facebook. *Media International Australia* 177: 12–29. [CrossRef]

Bryman, Alan. 2012. *Social Research Methods*. Oxford: Oxford University of Press.

Burgess, Jean, and Axel Bruns. 2015. Easy data, hard data: The politics and pragmatics of Twitter research after the computational turn. In *Compromised Data: From Social Media to Big Data*. Edited by Greg Elmer, Ganaele Langlois and Joanna Redden. London, Oxford, New York, New Delhi and Sydney: Bloomsbury, pp. 93–111.

Burgess, Jean, and Ariadna Matamoros-Fernández. 2016. Mapping sociocultural controversies across digitalmedia platforms: One week of #gamergate onTwitter, YouTube, and Tumblr. *Communication Research and Practice* 2: 79–96. [CrossRef]

Chatfield, Akemi Takeoka, Christopher G. Reddick, and Uuf Brajawidagda. 2015. Government surveillance disclosures, bilateral trust and Indonesia-Australia cross-border security cooperation: Social network analysis of Twitter data. *Government Information Quarterly* 32: 118–28. [CrossRef]

Cherven, Ken. 2015. *Mastering Gephi Network Visualization*. Birmingham-Mumbai: PACKT Publishing.

Dubois, Elizabeth, and Devin Gaffney. 2014. The Multiple Facets of Influence Identifying Political Influentials and Opinion Leaders on Twitter. *American Behavioral Scientist* 58: 1260–77. [CrossRef]

Dugue, Nicolas, and Anthony Perez. 2014. Social capitalists on Twitter: Detection, evolution and behavioral analysis. *Social Network Analysis and Mining* 4: 178. [CrossRef]

Eriksson, Katie, and Unni Å. Lindström. 1997. Abduction—A way to deeper understanding of the world of caring. *Scandinavian Journal of Caring Sciences* 11: 195–98. [CrossRef] [PubMed]

Fortunato, Santo. 2010. Community detection in graphs. *Physics Reports* 485: 75–174. [CrossRef]

Gama, Marcia, and João Gama. 2012. An overview of social network analysis. *WIREs Data Mining Knowl Discovery* 2: 99–115. [CrossRef]

Gerber, Elisabeth R., Adam Douglas Henry, and Mark Lubell. 2013. Political Homophily and Collaboration in Regional Planning Networks. *American Journal of Political Science* 57: 598–610. [CrossRef]

Grandjean, Martin. 2016. A social network analysis of Twitter: Mapping the digital humanities community. *Cogent Arts & Humanities* 3: 171458. [CrossRef]

Hand, Martin. 2014. From Cyberspace to the Dataverse: Tranjectories in digital social research. In *Big Data? Qualitative Approaches to Digital Research*. Edited by Martin Hand and Sam Hillyard. London: Emerald, vol. 13, pp. 1–30.

Hansen, Derek, Ben Shneiderman, and Marc A. Smith, eds. 2011. *Analysing Social Media Networks with NodeXL: Insights from a Connected World*. Burlington: Morgan Kaufmann.

Himelboim, Itai, Marc A. Smith, Lee Rainie, Ben Shneiderman, and Camila Espina. 2017. Classifying Twitter Topic-Networks Using Social Network Analysis. *Social Media + Society* 3: 1–13. [CrossRef]

Hine, Christine. 2000. *Virtual Ethnography*. London: Sage Publications Ltd.

Huang, Ronggui, and Xiaoyi Sun. 2014. Weibo network, information diffusion and implications for collective action in China. *Information, Communication & Society* 17: 86–104.

Huber, Gregory A., and Neil Malhotra. 2017. Political Homophily in Social Relationships: Evidence from Online Dating Behavior. *The Journal of Politics* 79: 269–83. [CrossRef]

Jörgens, Helge, Nina Kolleck, and Barbara Saerbeck. 2016. Exploring the hidden influence of international treaty secretariats: Using social network analysis to analyse the Twitter debate on the 'Lima Work Programme on Gender. *Journal of European Public Policy* 23: 979–98. [CrossRef]

Kozinets, Robert V. 2010. *Netnography: Doing Ethnographic Research Online*. London: SAGE Publications.

Lazer, David, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, and Nicholas Christakis. 2009. Life in the network: The coming age of computational social science. *Science* 323: 721. [CrossRef] [PubMed]

Lewis, Seth C., Rodrigo Zamith, and Alfred Hermida. 2013. Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods. *Journal of Broadcasting & Electronic Media* 57: 34–52.

Lin, Yu-Ru, Yun Chi, Shenghuo Zhu, Hari Sundaram, and Belle L. Tseng. 2008. Facetnet: A framework for analysing communities and their evolutions in dynamic networks. Paper Presented at the 17th International Conference on World Wide Web, Beijing, China, April 21–25.

McEwan, Bree, and Miriam Sobre-Denton. 2011. Virtual Cosmopolitanism: Constructing Third Cultures and Transmitting Social and Cultural Capital Through Social Media. *Journal of International and Intercultural Communication* 4: 252–58. [CrossRef]

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27: 415–44. [CrossRef]

Murthy, Dhiraj, and Scott A. Longwell. 2013. Twitter and Disasters: The uses of Twitter during the 2010 Pakistan floods. *Information, Communication & Society* 16: 837–55.

Ogan, Christine, and Onur Varol. 2017. What is gained and what is left to be done when content analysis is added to network analysis in the study of a social movement:Twitter use during Gezi Park. *Information, Communication & Society* 20: 1220–38. [CrossRef]

Ophir, Yotam, Dror Walter, and Eleanor R. Marchant. 2020. A Collaborative Way of Knowing: BridgingComputational Communication Researchand Grounded Theory Ethnography. *Journal of Communication Inquiry* 70: 447–72. [CrossRef]

Paccagnella, Luciano. 1997. Getting the seats of your pants dirty: Strategies for ethnographic research on virtual communities. *Journal of Computer-Mediated Communication* 3: JCMC314. [CrossRef]

Postill, John, and Sarah Pink. 2012. Social media ethnography: The digital researcher in a messy web. *Media International Australia* 145: 123–34. [CrossRef]

Quinn, James F., and Arthur E. Dunham. 1983. On hypothesis testing in ecology and evolution. *The American Naturalist* 122: 602–17. [CrossRef]

Samuel-Azran, Tal, and Tsahi Hayat. 2017. Counter-hegemonic contra-flow and the Al Jazeera America fiasco: A social network analysis of Al Jazeera America's Twitter users. *Global Media and Communication* 13: 267–82. [CrossRef]

Scott, John. 2017. *Social Network Analysis: A Handbook*. London: SAGE.

Starbird, Kate, Ahmer Arif, and Tom Wilson. 2019. Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations. Paper Presented at the PACMHCI. Available online: https://dl.acm.org/doi/pdf/10.1145/3359229 (accessed on 1 January 2021).

Takahashi, Bruno, Edson C. Tandoc, and Christine Carmichael. 2015. Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines. *Computers in Human Behavior* 50: 392–98. [CrossRef]

Thomas, David R. 2006. A General Inductive Approach for AnalyzingQualitative Evaluation Data. *American Journal of Evaluation* 27: 237–46. [CrossRef]

Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. New York and Cambridge: Cambridge University Press.

Wukich, Clayton, and Alan Steinberg. 2013. Nonprofit and Public Sector Participation in Self-Organizing Information Networks: Twitter Hashtag and Trending Topic Use During Disasters. *Risk, Hazards, & Crisis in Public Policy* 4: 83–109.

Xu, Weiai Wayne, Yoonmo Sang, Stacy Blasiola, and Han Woo Park. 2014. Predicting Opinion Leaders in Twitter Activism Networks: The Case of the Wisconsin Recall Election. *American Behavioral Scientist* 58: 1278–93. [CrossRef]

Yoon, Ho Young, and Han Woo Park. 2014. Strategies affecting Twitter-based networking pattern of South Korean politicians: Social network analysis and exponential random graph model. *Quality & Quantity* 48: 409–23.