This chapter is the pre-press, uncorrected version of a chapter accepted for publication in Longbottom, D. and Lawson, A. (eds) (2017) Alternative Market Research Methods: Market Sensing. Routledge.

Citation information for the manuscript is:

Hanlon, A. (2017). In Longbottom, D. and Lawson, A. (eds) Alternative Market Research Methods: Market Sensing. Abingdon, Oxon, Routledge, pp. 149-181.

CHAPTER 7

Social Media Networks: rich online data sources

Author: Annmarie Hanlon

Text
@grahammacca1 Thanks, Graham. If you've DMd your details, someone will be in touch. ^DaveT
@Julesdcc We are working on managing the hand luggage, Julian. We hope you'll see improvements soon! ^Lauren
@Bbflat6B We're sorry you're disappointed, Barry. We'll be in touch as soon as we have more information. ^DaveT
@nikeshashar Sorry to hear this, Nikesh. Have you tried deleting your cookies or using a different browser? ^Lauren
@KentGerman Glad to welcome you on board. Enjoy the flight and give our love to Düsseldorf! ^DaveT
@RobMcG22 Sorry to hear this, Robert. What do you mean by contaminated? ^Lauren
@carolainsolera What a great picture, Carola. Where are we taking you all today? ^DaveT
@Crustyfur Hi Chris. Sorry about the delay. If you DM us your booking reference and email we can set up a case for you? ^DaveT
@mehinix We're sorry you've been delayed, Radu. We'll have you on your way as soon as we can. ^Lauren
@elvismush Hi Elvis, your Bronze benefits will apply to your family if they are on the same booking. ^Gareth
@rebekahkjw Hi Rebekah. Do you have a case number and email address you can DM us please? We'll look into it for you here. ^DaveT
@ykdesignsonline Thanks for the information. We'll feed your comments back internally. We're sorry your expectations weren't met. ^Gareth
@grahammacca1 We understand your frustration, Graham. We'll get back to you as soon as we have more information. ^Lauren
@ZeroMoment1 No worries! We'll ensure your feedback is shared with the whole team. Enjoy the rest of your time in Hong Kong. ^Gareth

British 7.1: Twitter data (Source: author's own data extraction)

Purpose: *This chapter illustrates how social media networks can be harnessed for research to highlight feelings, behaviour and opinions of customers. This is a new area of research and will include discussions on* ***data mining*** *and* ***thematic analysis****.*

Context: *Social media networks were initially perceived to be a fad. Their ability to attract large volumes of customers has altered this perception and marketing managers have since recognised their place within the marketing mix, to:* ***inform new product***

development; deliver customer services; provide business development *and* ***facilitate brand management****. One advantage of social media networks is the ability to conduct research quickly and adapt promotional offers, customer service messages and include customers within the product development process.*

Learning outcomes: *At the end of this chapter you will understand the research opportunities available within social media networks and be better able to plan the management of research via social media.*

THEORY BOX
Philosophy: • ***Research studies using social media may come in different guises depending on either the personal preferences of the researcher or perhaps from the nature and objectives of the study.*** • ***A positivist philosophy involves the pre-determining of a theory (or hypothesis); here the researcher develops a hypothesis from previous knowledge (for example drawn from a literature review). The researcher then sets about searching and testing data which will either prove or disprove the hypothesis. This approach is also sometimes called a scientific approach, and often involves applying measurements and statistics. In this chapter the author presents the research process based on the positivist perspective.*** • ***An alternative might be an interpretive philosophy. Here the researcher has no pre-determined hypothesis as such but rather enters the field of research in order to discover and interpret new data.*** • *Ontology (positivist perspective):* **There is an objective reality and we can understand it through the laws by which it is governed** • *Epistemology (positivist perspective):* **Employs a scientific analysis and measures derived from the epistemologies of positivism and realism**
Approach: **Experimental and deduction. The hypothesis is seeking to prove evidence of positive or negative content.**
Strategy: **Thematic analysis based on data mining.**
Design: **Data capture via social media platform based on specific search terms (organisation or product names, descriptive words, keywords).**
Analysis: • **By theme: topics, words and phrases; data reduction is theme by thematic analysis.**
Presentation: • **By theme: words and phrases presented by theme as tree graphs and diagrams.**

Introduction, Background and Context

In this chapter we will be exploring how social media can be used to provide information to researchers and businesses. Our focus is content created by consumers, rather than the numerical aspect of social media network visit duration and / or least and most successful posts and updates.

Marketing is moving at a faster pace than ever before. In our 24/7 always-on world, customers seek responses to questions within minutes, they share feedback instantly which can go viral, sometimes before businesses have had the opportunity to respond. Consequently positive or negative comments online can build or break businesses. Much of this communication takes place online via Social Media Networks (SMNs) which have been present since 1997 with the earliest social media network recognised as the now defunct SixDegrees (boyd and Ellison, 2007). The most dominant current social media network in the USA and Europe is recognised as Facebook, launched in 2004 and today comprises over one billion active users.

As noted by boyd and Ellison (2007, p11) *“Social media network sites provide rich sources of naturalistic behavioral data.”* This is especially evident within Facebook. Its format has evolved to encourage users to share great amounts of significant personal data, which includes:

PERSONAL IDENTIFIERS	**Names, date of birth, place of birth, home town**
RELATIONSHIP MATERIAL	**Relationship status, linked relationships, family members, friendship groups, significant dates**
WORK AND EDUCATION RECORDS	**Places worked, where studied, education levels**
INTERESTS	**Religious affiliations, political views, hobbies, preferred music, films watched, favourite brands**
BEHAVIOUR	**Pages liked, comments added, downloads performed, purchases made, actions taken**

Figure 7.2: Data available from Facebook

This data is available anonymously to advertisers to more closely tailor and target their offers. In a social media context, anonymously means without names and addresses. It is also available to Facebook partners who develop applications (known as apps) where the primary purpose is to extract data from Facebook to incorporate the information into the business’s own customer relationship management databases and subsequently deliver customised marketing offers. This is one of the primary reasons for companies to develop ‘apps’; to discover more detailed information about their fans or friends, to understand how and when they engage online and to facilitate marketing opportunities.

Other social media networks have started to follow Facebook’s lead as they start to gather more user data. As an example, Twitter is trying to enrich its user data by obtaining dates of birth, not directly, but through the use of a celebratory hashtag #HBD inviting users to share their birth date.

Adding a hashtag symbol (#) ‘tags’ a word, making it easier to index and find through search engines such as Google. This was initially started outside Twitter, by a technology expert keen to group content through Twitter and needing a common search prefix.

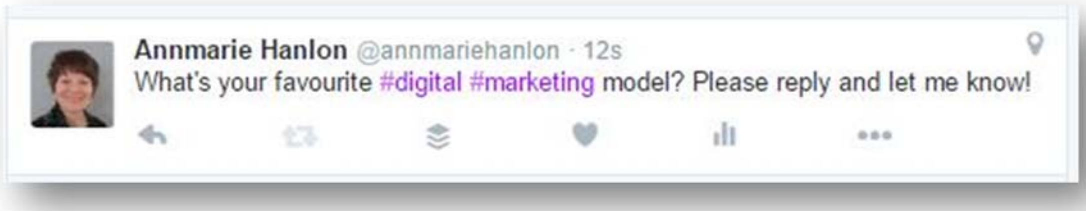


Figure 7.3: Example of hashtag usage in Twitter (Source: author’s own)

User data from the social media networks such as Facebook and Twitter is available to its advertisers and partners, as well as to researchers. Additionally, as social media networks have advanced and become more sophisticated, there is a move towards capturing more than the user profile (gender, age, location) and widening the data sets to include the user content; the updates, posts, tweets and other information which has been shared publically.

For example, Twitter is blazing a trail in gathering extant public posts and has indexed and made searchable every public tweet since the microblogging platform's launch in 2006. At one time it provided free access to researchers! This has since ended. Although data is available via third party suppliers for a fee and with restrictions.

The social media networks therefore provide a rich source of data; from personal opinions to relationship circumstances; location data to individual behaviours; employment records to political persuasion, as well as content created.

What can you research via social media?

Social media offers researchers real insights, often in real-time, for many aspects of data, including:

- **Users – Behaviour and opinions**
- **Networks – Size, scale, topics, connections, tie strength**
- **Content – Comments, images, video, hashtags**
- **Companies – Brand sentiment, product launches, product testing, customer services, feedback.**

Furthermore, data collection and research via social media networks afford additional advantages including: less administration; reduced costs; and rapid response rates (Laskey and Wilson, 2003). Unsurprisingly, there are also some disadvantages.

Disadvantages of conducting research via social media

Challenges with collecting data from social media networks include:

- **The required target population may not use social media**
- **The data is anonymised therefore demographic details are limited**
- **It can be difficult to access the data (see Ellison and boyd, 2013; Morstatter et al., 2013)**
- **The messages available are public not private which can limit the study (see Hong, Convertino and Chi, 2011)**
- **The messages tend to be shorter and contain less detail than other sources.**

This mini case provides an example of how social media has been used to test behaviour.

MINI-CASE INSIGHT 1
Use of Twitter to test behavioural responses to advertising messages

As an example of how Twitter user behaviour can inform social media advertising, Jilin Chen (Chen *et al.*, 2015) and others are working on targeting ads across social networks based on personality type.

They created a Twitter account @TravelersLikeMe and focused on Twitter users visiting New York City (NYC), because they found NYC among the most popular destinations mentioned on Twitter. Where they found Twitter users who said they were planning to visit New York in the near future, the @TravelersLikeMe account sent a reply tweet recommending various activities and encouraging a sign-up to a web link. If the user followed them back, they sent them a direct message.

The research methodology involved surveys and field studies and showed that this specific targeting has had an impact on the open rates.

The concept is that organisations could profile Twitter users and start conversations, based on their personality type.

The use of social media research in main stream marketing research literature

The literature regarding the use of social media as a research tool is an emerging area. Although the history of social media has been well-documented (O'Reilly, 2005; boyd and Ellison, 2007; Kaplan and Haenlein, 2010) and the concepts of social media networks and social media have been widely reviewed in the literature (Kozinets, 2002; Muñiz, Jr. & Schau, 2007; Adjei, Noble, & Noble, 2009; Kaplan & Haenlein, 2010; Mangold & Faulds, 2009; Stephen & Toubia, 2010), special issue papers and research agenda have concluded unsurprisingly that social media requires further research (Leeflang, 2011; Kietzmann *et al.*, 2012; Kane, Labianca and Borgatti, 2014).

As a newer marketing discipline, which was initially perceived as 'simply another channel' a factor noted by scholars (Rowley, 2004; Weinberg and Pehlivan, 2011), interest in social media networks is starting to grow.

One of the challenges within the research area of social media networks is that this domain crosses several research areas beyond marketing and often extends into technology; healthcare and education. The references used in this chapter also span several disciplines as the domain of marketing does not contain all the material needed.

User research

User behaviour as shown in mini case 1, as well as opinions can be researched in social media. This section looks at two popular areas: electronic word of mouth and sentiment.

Electronic word of mouth (eWoM)

Word of mouth marketing is possibly the oldest form of marketing communication and has been studied in traditional marketing communications for many years (see for example, Dichter, 1966; de Matos and Rossi, 2008) and is accepted as an effective method of marketing for business *"research generally supports the claim that WOM is more influential on behaviour than other marketer-controlled sources"* (Buttle, 1998). Social media marketing has been named a form of word-of-mouth marketing (WOMM) using professional techniques to influence consumer behaviour (Berthon *et al.*, 1998; Kozinets *et al.*, 2010; Abrantes *et al.*, 2013). eWOM communication can be described as *"any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet"* (Hennig-Thurau *et al.*, 2004) (Hennig-Thurau et al.,2004, p. 39).

Positive word of mouth can create sales and negative word of mouth can be costly for companies *"positive reviews have the potential to convert a consumer from 'not purchasing' to 'purchasing' by reassuring him/her that the product is of good quality and/or the company is reputable; uncertainty is thereby reduced. Conversely, negative reviews can squelch the 'buy'"* (Mangold and Smith, 2012) and unsurprisingly there is much research taking place into managing negative reviews (Brunner and Ullrich, 2014; Williams and Buttle, 2014). EWoM influences consumer buying behaviour and perceptions of brand (Amblee and Bui, 2011; Abrantes *et al.*, 2013) and is therefore an essential element within market sensing.

The concept of eWOM has evolved and researchers Canhoto and Clark (2013) refer to "brand-related online conversations – also called electronic word of mouth (eWoM)" (p522). One of their research projects adopted a snowball sampling approach to gathering online data, by posting invitations across social media networks and inviting respondents to share the invitation with others. They conducted analysis manually which was possible with a population of 44.

Chu and Kim (2011) suggest there are three aspects to eWoM “opinion seeking, opinion giving and opinion passing” (p50). All can be explored within the data, in terms of the content provided, as well as the sentiment within the content which can impact brands positively and negatively. The next section considers part of this, known as sentiment analysis.

Sentiment analysis

Another aspect of eWoM is *Sentiment Analysis*, which is also referred to as opinion mining. This is an aspect of *Natural Language Processing* and measures polarity classified in basic terms as positive, negative, neutral and mixed valence. It enables organisations to see their customers’ comments and thoughts, which could be used to inform new product development; improve customer services; identify business development opportunities and facilitate brand management.

As Lima, de Castro and Corchado (2015) discuss “Polarity determination can be made at different levels: document, sentence, word, or attribute” (p757).

There are two primary techniques used to conduct sentiment analysis, both based on data mining.

- **Lexicon-based** which considers the polarity of terms which may be extracted from an online system such as SentiWordNet.
- **Machine-based** using software for predictive modelling.

Several researchers (see for example: Prabowo and Thelwall, 2009; Chamlertwat and Bhattarakosol, 2012; Cotelo et al., 2015) are proposing combinations of these techniques.

Challenges with sentiment analysis

Mining data for sentiment analysis is not without challenges. The first is the volume and velocity of data. As an example, in the summer of 2015 Twitter was recording over 6,000 tweets per second, or 350,000 posts per minute. This nature of the data can benefit from automated processing although this has gained mixed responses for various reasons, including the way Twitter permits data collection (as an example, see Pew Research Center, 2013).

Secondly, the lexicon used may incorrectly ascribe a positive comment as a negative and vice versa. Schweidel and Moe (2014) discuss this in their work and also provide examples. There are recognised issues where natural language processing does not recognise sarcasm, irony and humour (see for example: Chamlertwat and Bhattarakosol, 2012).

As an example, consider the tweets shown in Figure 7.4. Based on the content of the language, would they be machine-coded negatively or positively?

Figure 7.4: Example of hashtag usage in Twitter (Source: author's own)

Thirdly, there can be duplication in the data downloaded. If downloading tweets and retweets, the same core content may be retrieved several times.

Finally, the rules on data collection from the social media networks can change. Twitter initially provided open access to data for researchers although this has now been limited.

More information

For a comprehensive guide to research in microblogging, see Cheong and Ray (2011), Marc Cheong has researched Twitter extensively and his work in this area is worthy of review. Okazaki et al. (2014) provide very useful procedural guidelines on opinion mining whilst Huang and Xu (2014) have written an insightful paper on exploring social data.

Researchers using social media can also download their own personal data from most social media networks. This can provide useful insights to understand what is available.

For more in this area, there are academic papers by several scholars, including: *"Word-of-mouth communications in marketing: A meta-analytic review of the antecedents and moderators"* by de Matos and Rossi (2008) and *"What We Know and Don't Know About Online Word-of-Mouth: A Review and Synthesis of the Literature*" by King, Racherla and Bush (2014).

Study Exercise

Go to your Facebook page:

- Select Settings
- Select General Account settings
- Download a copy of your Facebook Data.

Or go to your Twitter account:

- Select Settings
- Select Your Twitter data
- Scroll to the bottom of the page and select Twitter archive
- Download
- Request your archive
- You will receive an email when the archive is ready
- Click to download and follow the instructions.

- Was it easy to find the information?
- How long did the process take?
- What did you discover?

Network research

Understanding the size and scale of networks as well as the topics, connections and tie strength are areas of interest to both academics and practitioners. Since Granovetter's seminal work on the *Strength of Weak Ties* in 1973, which heralded the need for a social media network like LinkedIn, tie strength has evolved and comprises new areas such as blogger outreach programmes; enabling organisations to identify opinion leaders who will share their stories. This is where understanding the network, its size, scale and scope, can better inform practitioners as well as researchers.

Technical applications can calculate the tie strength of the social media networks. The Facebook application programming interface (API) facilitates data access and advises users about the data which they are sharing. Spiliotopoulos and Oakley (2014) provide a useful description of this process and Groeger and Buttle (2014) visually illustrate network ties in their research.

An API (application programming interface) provides the building blocks for computer programmers and developers to access software systems, following their guidelines. Most social media networks share their API with developers.

Content research

Content research explores a variety of formats; from words and images, to video and symbols like hashtags. The benefits of researching content allow organisations to shape their customer messaging, as well as managing issues and developing more effective advertising programmes.

To assess future content potential, in the form of predictions, Suman Kalyan Maity and colleagues are exploring a social media question and answer site (Quora) and have analysed the prediction of question topics (Maity *et al.*, 2015). They gathered data over four years using web-based crawling techniques to understand topic dynamics and their popularity.

Business research

Businesses conduct marketing research into many areas, such as new product development and customer services, which we will discuss here.

New product development

Historically new product development could take years to generate ideas. Starting with idea generation, creation of prototypes to gaining initial customer feedback and finally bringing the products to market. With social media, this entire process can be significantly reduced, with the target audience being involved in the process of co-creating the product or service. There is value in involving customer in the process of new product development as shown in research by Fang, Palmatier and Evans (2008, p322) *"customer participation affects new product value creation by improving the effectiveness of the new product development process by enhancing information sharing and customer–supplier coordination and by increasing the level of customer and supplier specific investments in the product development effort."*

Several companies have used social media networks, in particular Facebook, as a new product development research platform. This is a phenomenon known as *crowdsourcing*.

Crowdsourcing can be defined as harnessing the skills of many to deliver a solution. The solution ranges from ideas and suggestions, to finance and practical help (Surowiecki, 2004).

MINI-CASE INSIGHT 2
Use of Facebook to develop new products

Walkers crisps uses Facebook as an element of its new product development.

Their research process for new product development takes place via a social media network. They use crowdsourcing and ask their fans on Facebook to recommend new product ideas. The process is usually followed by short-listing, mass voting and subsequently a winner is selected. The entire process takes place and is shared on Facebook.

Customer services

Market sensing works in a dyadic way; customers seek information from companies and companies seek information from customers. As customers moved online and communicated with brands across a 24/7 environment, several companies had to move their customer services into the social media space. In some cases there was no prior research to advise that social media was a customer services environment, simply a sense of need and urgency, created by customers.

Developing a customer services offer online has long been identified (Walsh and Godfrey, 2000) and using social media networks such as Twitter, has been embraced by consumers who have realised that comments in public generate faster responses (Canhoto and Clark, 2013).

One of the management issues for companies is the lack of control; over content, timing and frequency of information which generates its own challenges *"(firms) are struggling to navigate the emerging complex, consumer- empowered environment"* (Gallaugher and Ransbotham, 2010, p197).

MINI-CASE INSIGHT 3
Use of Twitter as a customer services tool

Social media networks act as research tools in a dyadic format; for both organisations and customers. A public example of a company which was forced, by customers, to harness Twitter as a customer service tool, is British Airways.

British Airways did not consider Twitter as a research platform and was not listening to its customers' comments online. When British Airways joined Twitter in 2008, it was a monadic redirection system. Its function was to signpost customers to the official website. This use of Twitter for British Airways changed in 2013, when customers demanded a dyadic approach.

One customer (known as @HVSVN) flying with British Airways was unhappy when his luggage was lost and could not achieve the desired response from the company. To gain redress he publically described the poor customer service. In each comment (tweet) he referenced British Airways' Twitter name. When no formal acknowledgement was forthcoming, he spent $1,000 on Twitter advertising and promoted his comments across the social media network (British Airways, 2014). The comments gained widespread attention and raised fundamental questions about the use and management of social media for businesses.
Several companies, including British Airways, now use social media as part of their integrated marketing research process.

MINI-CASE INSIGHT 4
Use of Twitter as a customer services tool

Using social media networks for ongoing customer research can enable businesses to manage specific processes. A case example of good practice in marketing through social media networks includes a UK train company, London Midland.

Their Twitter home page explicitly demonstrates their adoption of the purpose of the social media network, as evidenced by their biography *"Here to help from 7am (8am weekends). We aim to reply to all tweets, but pls try to be polite if things have gone wrong – we're real people just trying to help!"* which shows leadership by stating their rules of engagement and enabling a consistent and authentic voice, where staff share real names and add personality to updates.

For examples of how corporations use Twitter as an engagement tool, read Mamic and Almaraz (2014).

Study Exercise

- Identify one example of best practice of customer service via a social media network.
- Does the organisation state its 'rules of engagement?'
- How quickly does the company respond to negative feedback?
- Is the brand voice corporate (like British Airways) or personal (like London Midland)?

ETHICS BOX

The social networks have eyes and ears as well as a host of data that you can extract. At this stage it should be noted that there are debates about whether it is ethical to use online data.

There are two major ethical considerations in data mining. Firstly, permission to use the data and secondly, permission to gather the data.

Ethics - Permission to use data

Individuals may tweet publically, but may be unaware that their posts may be used after publication. From an ethical stance it is unclear where posts on social media networks are public or private behaviours. Tweets may be made publically but considered private content to be shared within a network.

Another issue is how the researcher has access to the network. Some researchers are part of specific networks and could potentially use the data for their own research. For example, internet message boards by mothers sharing ideas and support, such as Mumsnet in the UK and Babycentre and Essential Baby in Australia, are available to any expectant mother who could also be a researcher.

Simply following an organisation on Twitter enables the researcher to see all their tweets, as well as comments made about the organisation.

At the same time, membership of all social media networks is optional. Organisations and individuals can select whether or not to join a special network. Organisations and individuals can decide whether their content is private (locked) or public.

Ethics – Permission to gather data
The social media platforms allow researchers the ability to gather data, but have rules about the volume of data captured and the timescales. See for example https://support.twitter.com/articles/160385-twitter-api-limits and https://www.facebook.com/terms.php before embarking on studies.

Some platforms also have dedicated data resellers, such as Gnip which sells Facebook and Twitter data.

The 'gold standard' of ethics policies is widely accepted as that used by the British Psychological Society (see www.bps.org.uk/what-we-do/ethics-standards/ethics-standards for the full list of options).

Taking you through the process stage by stage

In this section we will go over the main stages in conducting this type of research.

Process Box: main stages and activities for the research process

STAGE	ACTIVITIES (*and key issues*)
1	Selection of platform and topics • *Justify your selection* • *Agree your topic focus*
2	Sampling • *To test the dataset availability*
3	Selection of data mining method • *By machine, manually or third-part intervention* • *Cover ethical issues*
4	Conduct the data mining • *Data extraction* • *Data processing*

5	Thematic analysis • *6 step process*
6	Presenting your data

Stage 1: Selection of platform and topics

The methodological approach we are presenting is based on a ***positivist philosophy***:

- the data source is based on specific platforms,
- the data is selected based on pre-determined topics, words or organisations,
- the researcher is analysing by thematic analysis,
- the interested reader of the research is interpreting the outcomes for relevance to their own situations which could change business practice.

The data source is based on social media networks which facilitate data access and allow researchers to capture data. This currently includes the world's major social media networks, such as: Facebook; Twitter; Tencent Weibo; Sinar Weibo; Tumblr; Google+; YouTube; and Instagram.

The selection of data source will be dependent upon:

- the research objectives,
- the target audience,
- the research topics,
- the skills of the researcher in terms of technical ability and languages spoken.

The target audience

Different individuals and organisations use different social media networks. The latest social media research from the United States advises that Facebook is used by over 70% of adult internet users, Twitter is popular with those aged under 50 years old and 54% of its users live in households earning more than $50,000. Instagram is popular with those aged 18 to 29 and Pinterest is mainly used by women (Duggan *et al.*, 2015). The key is understanding the target audience and identifying the most relevant social media network.

The research topics

The data to be analysed emanates from user-generated content which has been publicly shared online. This comprises, for example:

- posts, tweets or comments which users have added to an *organisation's* Facebook, Twitter or other social media network page,
- posts or comments about an organisation, added to the *user's* social media network page.

The first step is deciding the type of information required or 'unit of analysis' (Okazaki *et al.*, 2014). Several researchers have classified different categories of online content, with examples shown here. Their papers discuss the classifications in detail:

Categories of tweets (Bruns and Stieglitz, 2013)	**Categories of tweets (Sriram *et al.*, 2010)**	**Categories of influence (Cha *et al.*, 2010)**
Original tweets	Neutral News	Indegree influence
@mentions	Personal News	Retweet influence
Genuine @replies	Opinionated News	Mention influence
Retweets	Opinions	
Unedited retweets	Deals	
Edited retweets	Events	
Tweets containing urls	Private Messages	

Secondly, the researcher has to decide whether the topic is about an organisation, a subject as denoted by a hashtag or other specific words, including product names. This is for several reasons. As one example, the volume of data can be overwhelming, therefore a focus is required. Secondly, when using machine methods (see stage 3) data can only be extracted when specific terms are provided. It is difficult to search for 'feelings about brand X' whereas 'negative feelings (including the words, dislike, unhappy, terrible) about brand X', can be obtained, as the search terms have been elucidated. This positivistic approach has been adopted by companies such as DataSift (see http://datasift.com) who have developed 'Curated Stream Definition Language'. Effectively they have built (as have others) their own lexicon. This is a paid-for service and outside the financial scope of most researchers.

The skills of the researcher

Skills at this stage are required in terms of technical ability and languages spoken. Some of the larger social media networks, such as Tencent Weibo and Sinar Weibo which are widely used instead of Twitter in China, are in simplified Chinese. Xing mainly occurs in German and Japanese users of Twitter, unsurprising, predominately communicate via Kanji. This means language ability would be required for analysis in some geographical areas, as not all social media networks converse in English.

The additional skill required can be technical. There are options, discussed in stage 3, about data mining methods.

Stage 2: Sampling

To test the potential dataset, sampling is required. This could take place manually (if machine mining will be used for the main survey), to understand if the required data is available.

There are several free options to conduct the sampling. As an example to search for hashtags include, for example, the website Tagboard (see tagboard.com). This allows searching via hashtag and enables any user to perform a simple search, such as the one shown here, using the hashtag #travel.

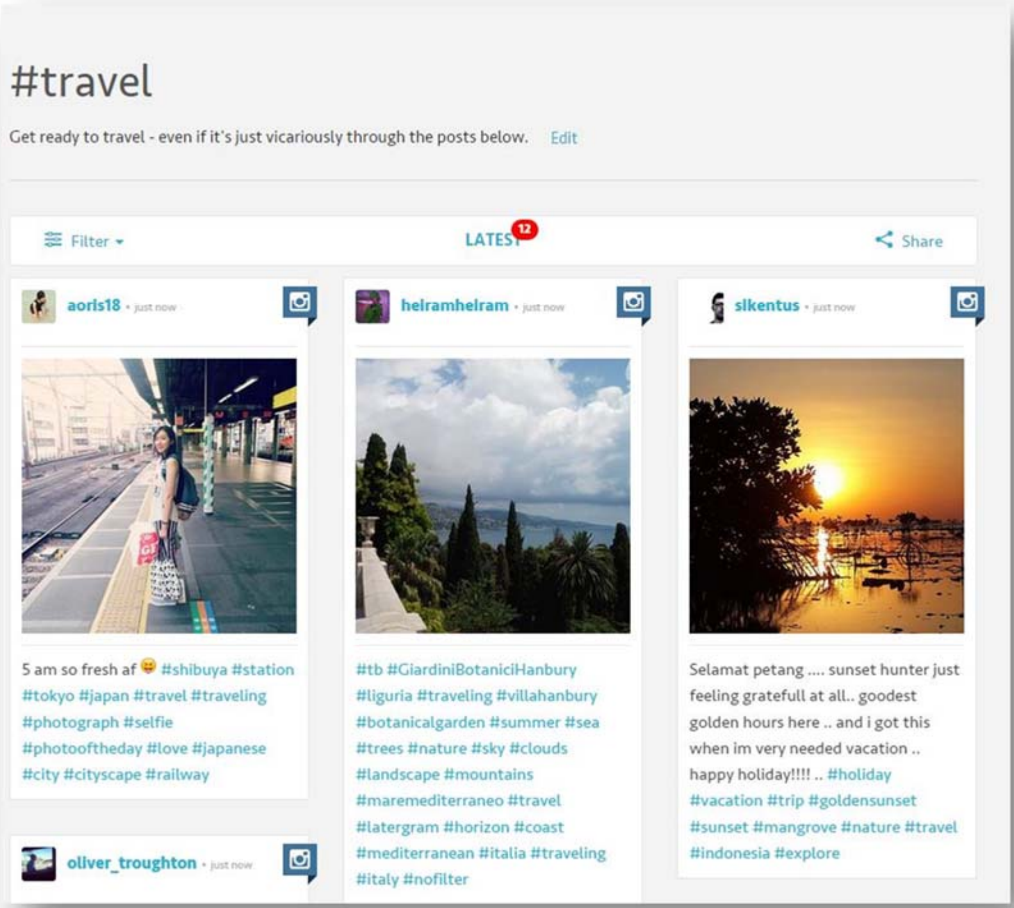


Figure 7.5: Example of data from Tagboard using the hashtag travel in Twitter (Source: author's own)

A similar tool is TweetArchivist (see www.tweetarchivist.com) which displays the results in a different format and enables the viewer to see limited details about the topic.

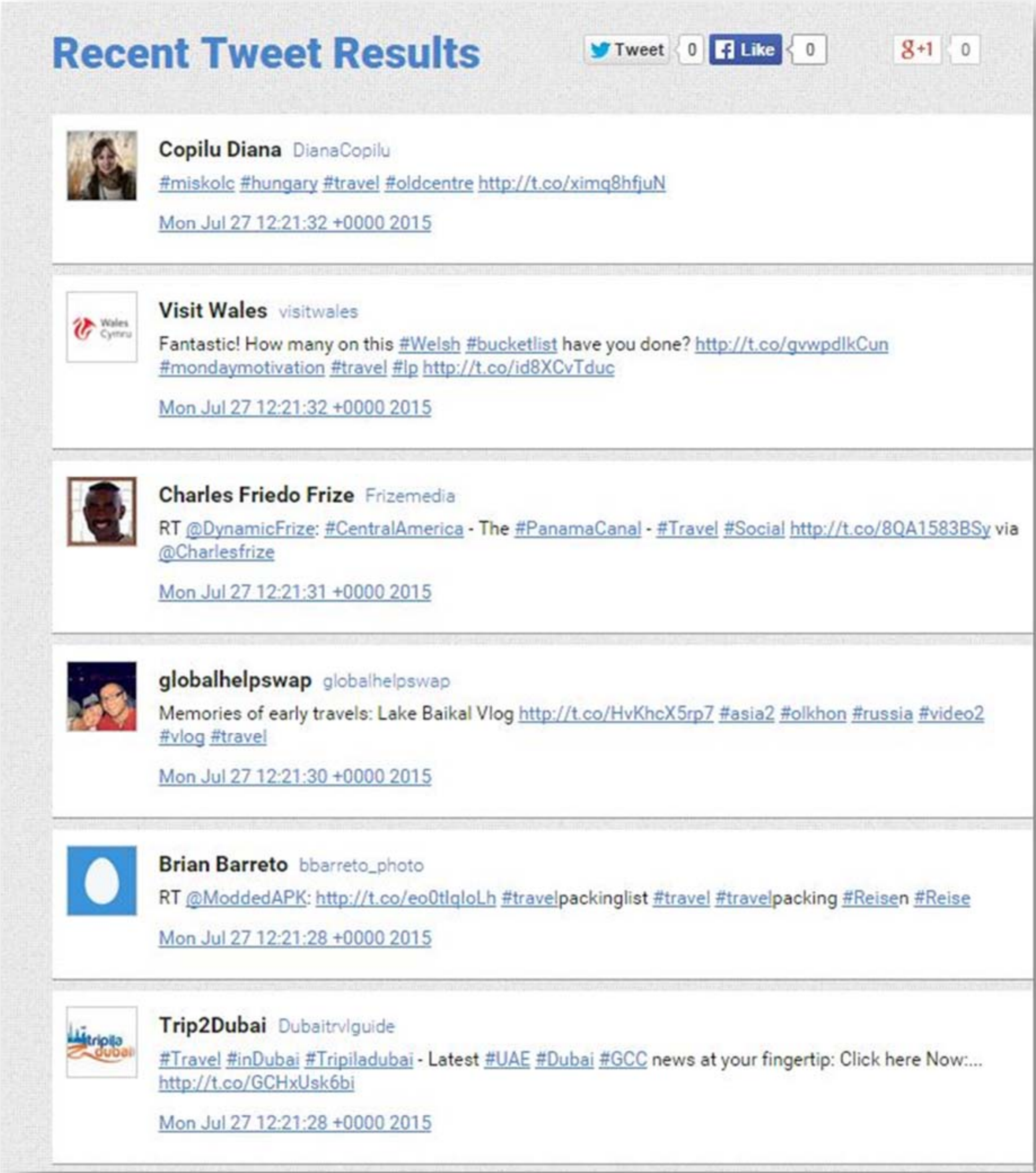


Figure 7.6: Example of data from Tweetarchivist using the hashtag travel in Twitter (Source: author's own)

This is supported by additional data options, such as the top users, most frequently mentioned words associated with the hashtag and the language in which the content was provided.

To obtain more detailed data from TweetArchivist requires a paid account which starts at $14.99 per month.

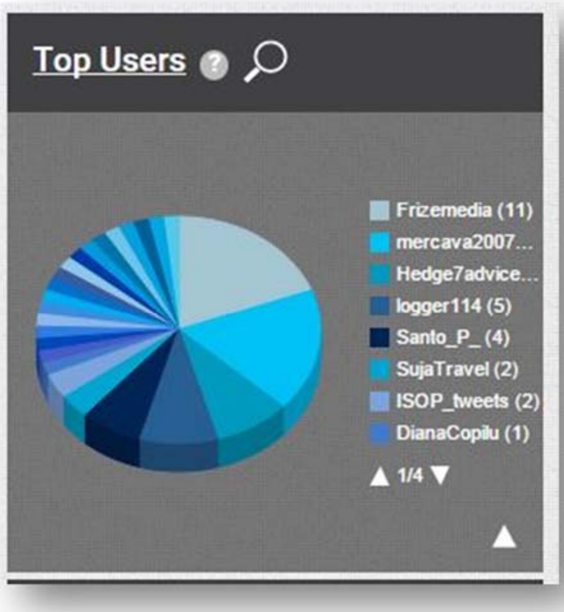


Figure 7.7: Example of data sources from Tweetarchivist using the hashtag travel in Twitter (Source: author's own)

Twitter can also provide samples. There are fees attached to this and some form filling. Initially, an estimate for the data can be obtained via a form http://gnipinc.formstack.com/forms/data_rofr and approval is also required.

Stage 3: Selection of data mining method

There are different methods of data mining and based on the research question, timescales and budget, there are several options available.

Table: Some methods of data mining

Method	Advantages	Disadvantages
Machine mining	• Quick to capture the data • Captures large amounts of data quickly	• Expensive • Can require third-party input (technical help or software) • Demographic details may be missing • Machines cannot see sarcasm or humour!

Manual mining	• Less expensive as it involves the researcher's time • Can interpret nuances such as sarcasm or humour	• Time to capture the data • Time to copy and paste into other software (such as a spreadsheet) for analysis
Third-party intervention	• Other people follow your instructions and carry out the tasks	• Clear instructions needed • Only available in certain countries • Costs can mount up • Questions of data reliability and bias need to be addressed

Machine mining

Machine mining ranges from ready to use paid-for options to building data extraction applications. Twitter provides access to its data via several options, the most popular is via its API (application programming interface) which enables developers to extract raw data. This includes data such as the message content, time of message, geo-location information. Marc Cheong's thesis *'Inferring Social Behavior and Interaction on Twitter by Combining Metadata about Users & Messages'* contains useful background to Twitter data extraction using technology (Cheong, 2013a).

Other machine mining methods include a range of paid-for tools provided through social media network partners such as Gnip, Datasift and others.

A paper by Torgeir Aleti Watne, Marc Cheong and Will Turner is a good case example of using machine mining in Twitter (see Watne, Cheong and Turner, 2014).

QSR NVIVO coding software includes N-Capture which when installed as an extension on your browser, will enable automatic retrieval of tweet, Facebook and LinkedIn data, direct from the page. Some NVIVO training may be required.

Manual mining

Manual extraction of data takes considerable time and patience. This is achieved by identifying a search term, such as #travel and using Twitter's advanced search feature (see twitter.com/search-advanced).

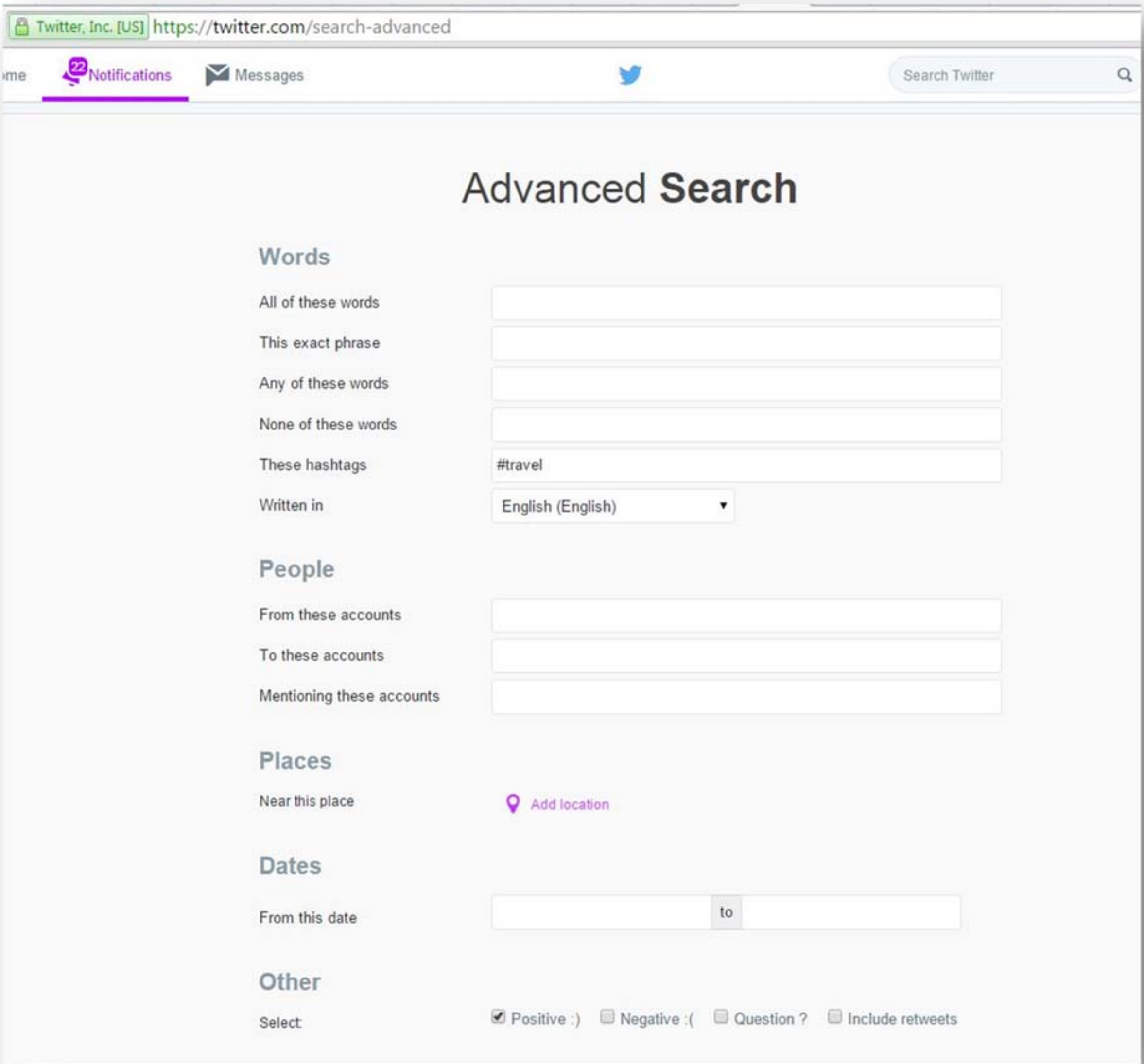


Figure 7.8: Example of Twitter advanced search options (Source: author's own)

This allows the researcher to specify the exact search terms, whether the data is about certain organisations, geographically centred, within certain dates and whether the tweet was positive or negative (determined vaguely with use of emoticons and language, but not failsafe).

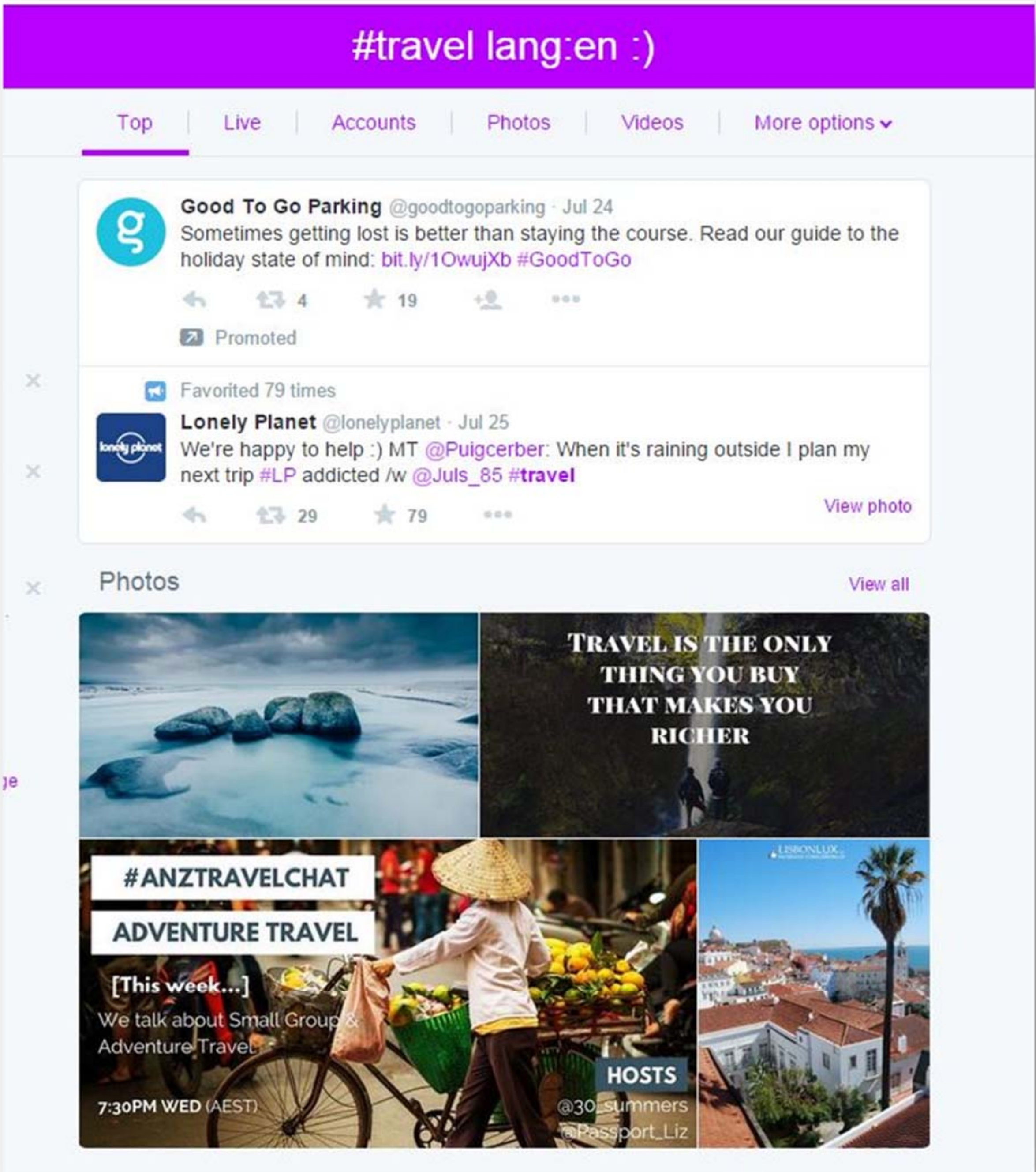


Figure 7.9: Example of Twitter advanced search results (Source: author's own)

The search results are based on the defined terms and the researcher at this stage could manually copy and paste the posts from the screen into an excel sheet. This takes some time, depending upon the required sample size and research question.

Third party options

As a result of the time required for manual data mining, a range of alternative third-party options have been created, known as micro-tasking. This includes the largest in this field, Amazon's Mechanical Turk (MTurk) as well as others such as Microworkers.com and Clickworker.com. Jaime Arguello from the School of Information and Library Science at the University of North Carolina at Chapel Hill, has been researching 'Predicting Speech Acts in MOOC Forum Posts' and explains how he has used MTurk (see Arguello and Shaffer, 2015).

Amazon's Mechanical Turk

Amazon has created a crowdsourcing internet marketplace for businesses that need tasks completing by humans, these are called Human Intelligence Tasks or HITs. This started in 2005 and was intended to provide a facility for companies to subcontract digital tasks where computers failed, such as transcribing, writing and tagging images. (See mturk.com)

Since this time, the Mechanical Turk has been used as a way of gathering widespread opinions and conducting surveys. There is bias attached as the population is already digitally literate. The fee rate for the task is set by the researcher and people decide whether or not to take on the task. Typically rates range from a few cents to a few dollars and there have been debates from the online community as to the ethics of using people as machines.

One task that Mechanical Turk is geared towards is sentiment analysis. Instructions are provided online in a 'ready to go' format. See https://requester.mturk.com/create/sentiment/about

NB At the time of writing, Mechanical Turk does not support researchers from countries outside the United States.

Another third-party option, is Volunteer Science. This is an online semi-gaming environment created by another US university. Researchers can set up surveys, forums and panels and recruit volunteers to participate (see volunteerscience.com).

Stage 4: Conduct the data mining

The methodological approach taken is experimental and deductive (where the researcher has a structure of themes and a set hypothesis), using a structure of themes emerging from the literature.

The data extraction depends on the platform selected. By platform we mean social media network and the main social media networks which facilitate data extraction are Facebook and Twitter.

Data extraction can take place either by human coding or machine coding. Okazaki et al. (2014) argue that machine coding is more reliable than human coding, due to inconsistent coding practices when people are involved, whereas machine coding can focus on an agreed algorithm and is replicable.

Data mining via Twitter is not without challenges. Twitter imposes limits on the volume of data obtained over specific timescales. Marc Cheong (2013) conducted research into specific organisations (micro-breweries) and here describes how he captured data from Twitter:

> "In order to obtain Twitter data from different (organisations), we used a 'best-effort' data collection strategy (Cheong & Lee, 2010) involving the 'Twitter REST API' (Application Programming Interface). The Twitter REST API is a service that obtains raw metadata about a particular handle or user and their messages, allowing researchers to 'harvest' tweets from particular users. We harvested tweets using a collection of 'scripts' written in the Perl programming language by first using the Search API (a subset of the REST API). The output of the scripts contain more than just plain message text as it also consists of metadata (data about data) that explains the context of a given user and its postings."

REST stands for REpresentational State Transfer and can be described as a software format. REST API is like a key which provides access to read Twitter data.

This illustrates some of the technical skills required, using machine mining.

Once the data has been extracted, the next step is data processing. If the data is extracted manually, via machine or using mixed methods, it is likely the data will be added to a spreadsheet for further analysis.

Truth and Trolls! Beware the authenticity trap

One consideration is the concept of self-presentation of identity (see for example Marwick, 2005), which is an issue across all social media networks. Is the researcher reviewing content from real users? There are caveats of which researchers need to be aware.

- Facebook is keen for its users to use their real names and for a time introduced its 'real names policy'. The issue was with people whose real names were creative and Facebook thought they were fake profiles, threatening to close their accounts if they were unable to provide proof of their identity.
- Twitter users are restricted to 15 characters to create their identity, which means it can be difficult to obtain your real name on the social media network.
- Some people behave badly across social media networks. This anti-social behaviour is often characterised by trolls and users who are banned for offending the community. Some of these users create fake profiles which may need to be excluded from the work – unless it is about anti-social behaviour! Read 'Antisocial Behavior in Online Discussion Communities' by Cheng, Danescu-Niculescu-Mizil and Leskovec (2015) which describes this behaviour.

Stage 5: Analysis options

Within social media data whilst there are different techniques for data analysis including thematic, sentiment and brand analysis, these techniques are derived from Content Analysis and can be used for quantitative and qualitative analysis, as well as mixed methods. For more on content analysis we have found two useful texts *'Content analysis: an introduction to its methodology'* by Klaus Krippendorff (2004) and *'The content analysis guidebook'* by Kimberley A Neuendorf (2002).

One of the advantages of market sensing via social media is that the content is usually limited in terms of the number of characters. Twitter posts are required to comprise of fewer than 140 characters (including punctuation and spaces) and posts on Facebook are usually short, although the limit is around 60,000 characters.

Themes identify the essence of a piece of content and enable researchers to understand ideas and concepts related to the research question. Thematic analysis has been well-defined by Braun and Clarke (2006) in their paper which proposes clear guidelines for undertaking this method. There are two decisions to be taken before the thematic analysis starts:

- is the work at a *semantic* level, where the themes are explicitly identified in the data, or
- is the work at a *latent* level which identifies the “underlying ideas” (Braun and Clarke, 2006, p84).

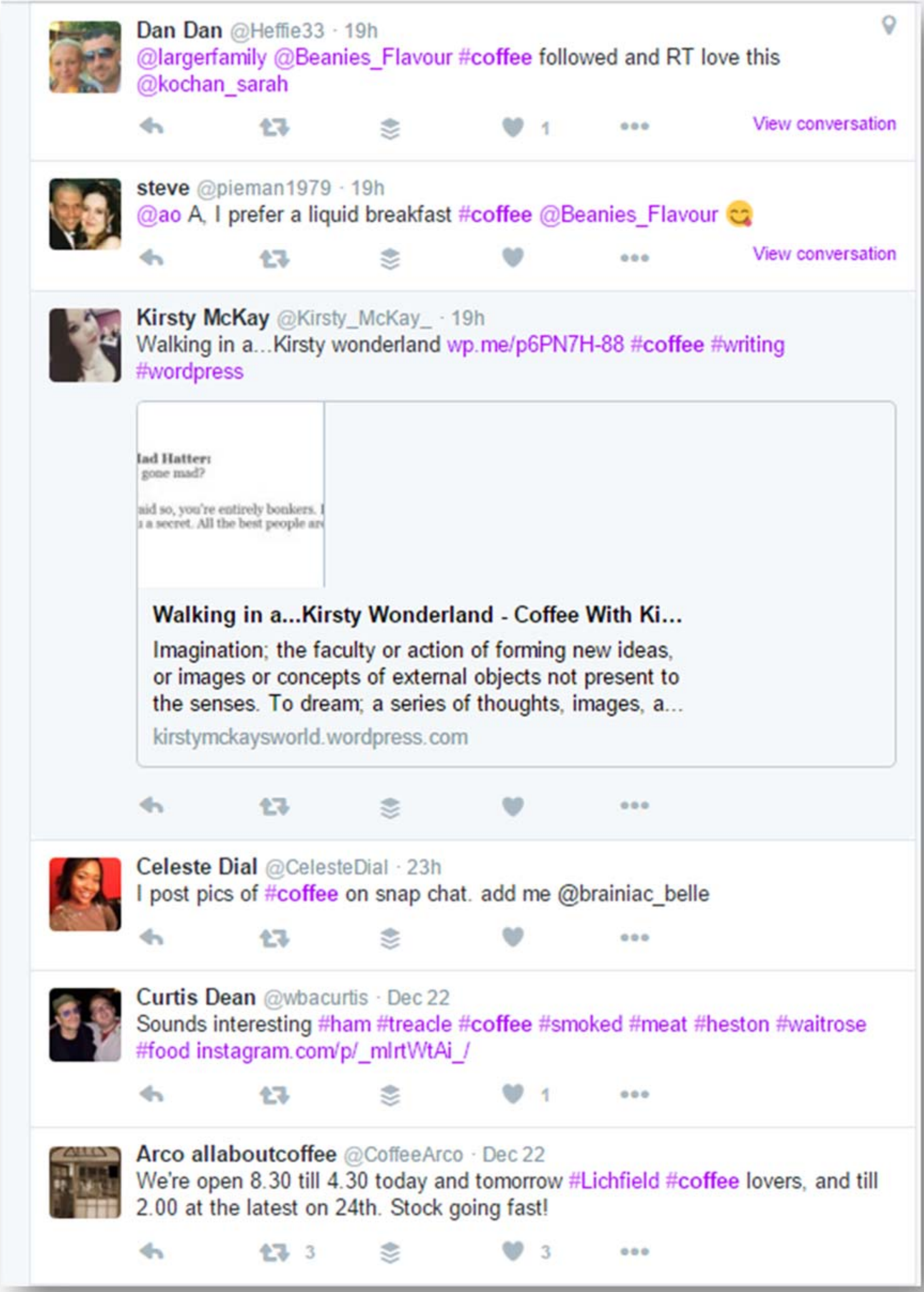
Regardless of whether the analysis is semantic or latent, the best process for undertaking thematic analysis that we have found, is that proposed by Braun and Clarke (ibid.) which follows six steps:

1. Familiarizing yourself with your data
2. Generating initial codes
3. Searching for themes
4. Reviewing themes
5. Defining and naming themes
6. Producing the report.

1. Familiarizing yourself with your data

This is about ensuring you understand what your data is. If you have followed this process so far, you will already be familiar with your data, be it brand names or specific hashtags used in a social media network between specific dates.

Here is an example of Twitter data, using Twitter advanced search, seeking tweets based on the hashtag coffee (#coffee).


Dan Dan @Heffie33 · 19h
@largerfamily @Beanies_Flavour #coffee followed and RT love this @kochan_sarah
1
View conversation
steve @pieman1979 · 19h
@ao A, I prefer a liquid breakfast #coffee @Beanies_Flavour
View conversation
Kirsty McKay @Kirsty_McKay_ · 19h
Walking in a...Kirsty wonderland wp.me/p6PN7H-88 #coffee #writing #wordpress
Walking in a...Kirsty Wonderland - Coffee With Ki...
Imagination; the faculty or action of forming new ideas, or images or concepts of external objects not present to the senses. To dream; a series of thoughts, images, a...
kirstymckaysworld.wordpress.com
Celeste Dial @CelesteDial · 23h
I post pics of #coffee on snap chat. add me @brainiac_belle
Curtis Dean @wbacurtis · Dec 22
Sounds interesting #ham #treacle #coffee #smoked #meat #heston #waitrose #food instagram.com/p/_mlrtWtAi_/
1
Arco allaboutcoffee @CoffeeArco · Dec 22
We're open 8.30 till 4.30 today and tomorrow #Lichfield #coffee lovers, and till 2.00 at the latest on 24th. Stock going fast!
3
3

2. Generating initial codes

When you are familiar with your data, it is time to generate the initial codes. Depending on your objectives, you might start to code at the top level, either based on brand name, product or sentiment.

3. Searching for themes

This phase only starts when all data has been initially coded and a long list emerges. The aim is to reduce the long list of codes and start to apply themes. Braun and Clarke suggest different approaches to the search for themes including using mind maps and other visual tools, to bring the themes to the fore.

4. Reviewing themes

Reviewing themes comprises reducing the themes and may involve merging. At this stage additional researchers may be involved to check the codes, test the validity level using an agreed mechanism to check intercoder reliability and to ensure the research can be replicated.

5. Defining and naming themes

This stage seems repetitive of the previous stage as it involves further refining and defining themes. Braun and Clarke suggest (p92):

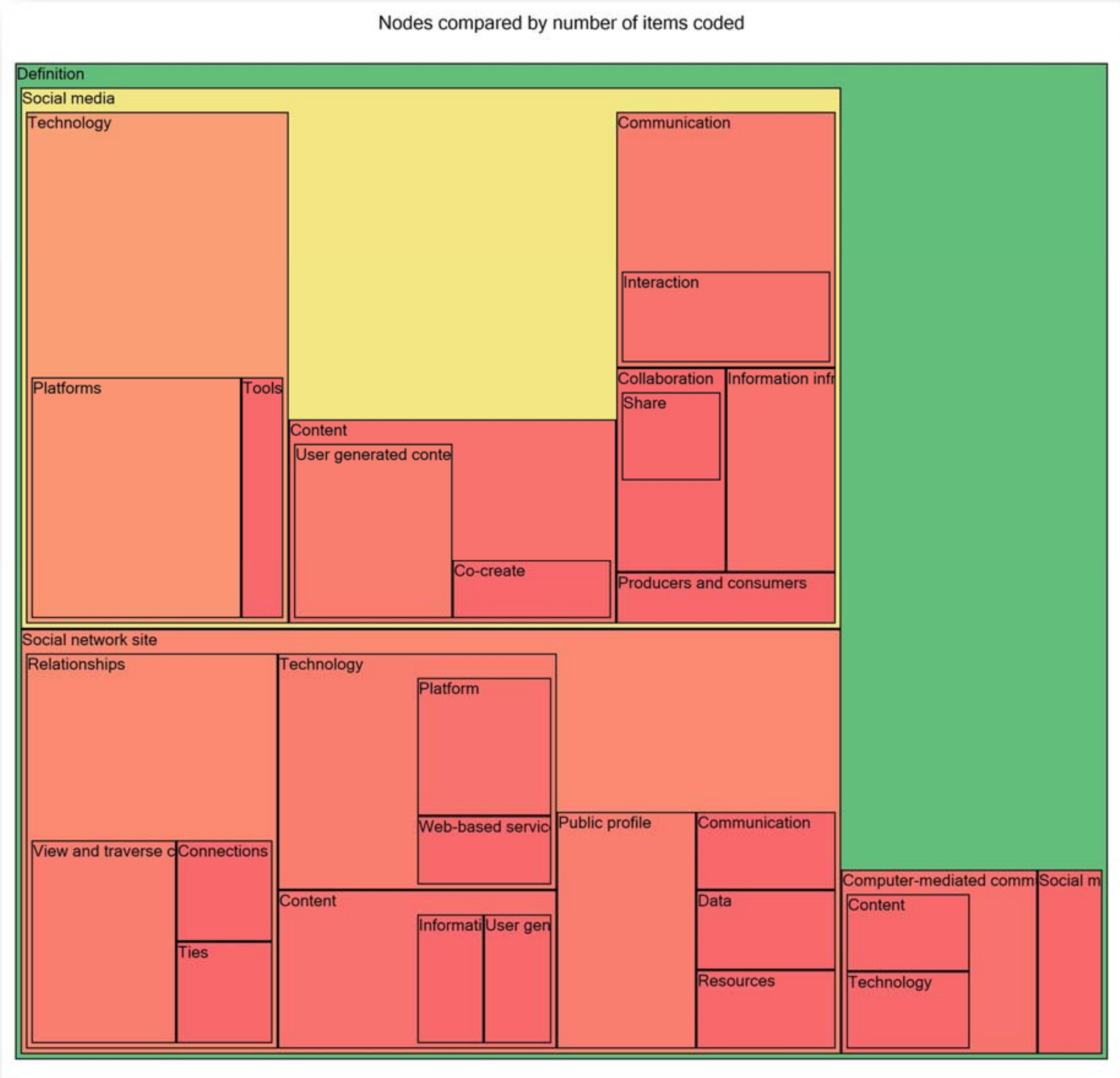
> *"For each individual theme, you need to conduct and write a detailed analysis. As well as identifying the 'story' that each theme tells, it is important to consider how it fits into the broader overall 'story' that you are telling about your data, in relation to the research question or questions, to ensure there is not too much overlap between themes."*

6. Producing the report

The last phase is also about presenting the data, which we consider after this section.

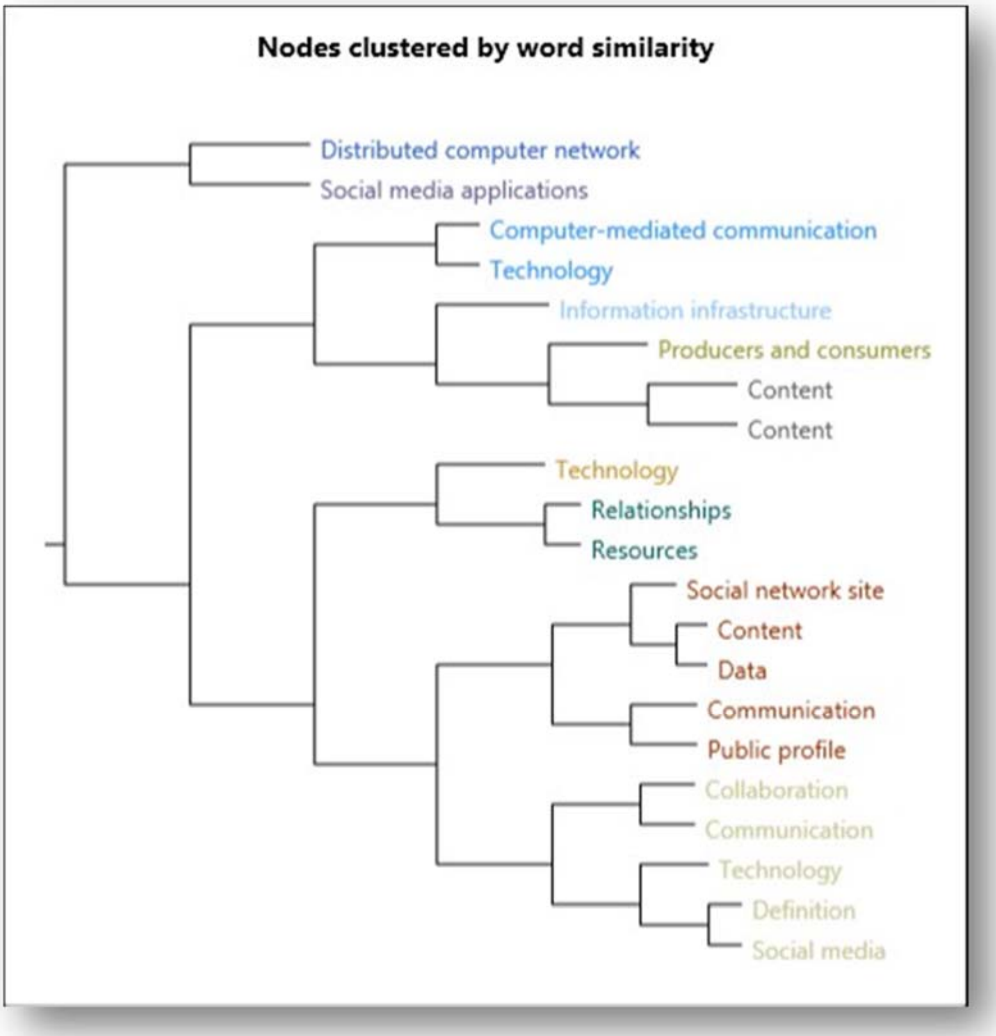
Stage 6: Presenting your data

The style of data presentation may become obviously as the thematic analysis evolves. Thematic analysis can involve using visual tools and for the definitions of social media and social media networks, the software being used (Nvivo 10) provided several visual options including a tree graph, as shown here:



The tree graph succinctly compares the sub-themes (nodes) within each theme. It highlights the degree of use (the larger the box, the more used the term).

The same data can also be displayed as a horizontal dendrogram which shows the cluster analysis.



Conclusion

Using social media for market sensing is a new domain. Tools, technology and platforms are evolving, facilitating the process and making the data collection process easier for researchers. Whilst this approach could be considered as largely positivistic, seeking to prove evidence of positive or negative content, this depends upon the approach taken to data mining. A machine-led approach warrants such an approach as it is founded on the selection of specific words and phrases, without room for an interpretive approach.

However it is possible that the researcher using social media may choose instead to follow a more interpretive philosophy (as we discussed in the theory box at the start of this chapter). Studies involving manual mining may be better suited to the interpretive philosophy.

One major consideration with all market sensing via social media is the need for rigour in collecting the data. With 6,000 tweets a second, if the original data is not prudently collected, it may not be possible to reacquire the same information at a later time.

Social media affords both a rich source of material and unparalleled access to behavioural data. Whilst there have been many discussions about users sharing too much content and some users moving away from various networks, regardless of how the social media platforms evolve and potentially become more segmented, users will continue to share personal data across many networks, providing a prolific and continuing source of research material.

End of Chapter Summary

This chapter has demonstrated how data from social media can be extracted to reveal feelings, behaviour and opinions of customers, in a way which was not previously possible.

This type of research may help marketers in areas of:

- **New product development**: providing insights into customers' needs and preferences.
- **Customer services**: enabling organisations to better manage customers' opinion giving and opinion passing, and to ensure that major issues are addressed sooner rather than later.
- **Business development**: sharing insights into opportunities for areas of business development.
- **Brand management**: enabling marketers to understand and gather customers' perceptions of brand offers and brand values.

End of Chapter Exercise

A classroom exercise which requires students to have access to computers or laptops (or perhaps in a computer lab). This will provide students an understanding of the data available online and the issues that arise.

This works especially well where students have their own Twitter accounts and can see how their own tweets feature in the results.

The tutor should ascribe a hashtag, for example #coffee or #technology or #travel. The aim is to select a hashtag which is relatively neutral (avoiding politics and religion). This can also form the start of the session where suggested hashtags are proposed by the students.

Once the hashtag has been agreed, those with Twitter accounts should compose a tweet using the hashtag.

Working individually, students should select a method to download Twitter data. This can be achieved using the tools mentioned in stages 2 and 3.

It is likely that this will data need to be exported to or opened in Excel in order to better see and to possibly group the results.

Once the data is downloaded, the students should:

- Identify the type and volume of data – what is included and what is excluded?
- If different data collection tools were used, how did the results vary? Did the students find their own hashtag-based tweets in the results?
- Discuss in pairs how this data could be used.

A group discussion can take place at the end of the session to explore the uses, benefits and challenges with using social media for market sensing.

References

Abrantes, J. L., Seabra, C., Lages, C. R. and Jayawardhena, C. (2013) ‘Drivers of in-group and out-of-group electronic word-of-mouth (eWOM)’, *European Journal of Marketing*, 47(7), pp. 1067–1088. doi: 10.1108/03090561311324219.

Adjei, M. T., Noble, S. M. and Noble, C. H. (2009) ‘The influence of C2C communications in online brand communities on customer purchase behavior’, *Journal of the Academy of Marketing Science*, 38(5), pp. 634–653. doi: 10.1007/s11747-009-0178-5.

Amblee, N. and Bui, T. (2011) ‘Harnessing the Influence of Social Proof in Online Shopping: The Effect of Electronic Word of Mouth on Sales of Digital Microproducts’, *International Journal of Electronic Commerce*, 16(2), pp. 91–114. doi: 10.2753/JEC1086-4415160205.

Arguello, J. and Shaffer, K. (2015) ‘Predicting Speech Acts in MOOC Forum Posts’, in *Ninth International AAAI Conference on Web and Social Media*. Oxford, UK. Available at: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10526.

Berthon, P., Lane, N., Pitt, L. and Watson, R. T. (1998) ‘The World Wide Web as an Industrial Marketing Communication Tool: Models for the Identification and Assessment of Opportunities’, *Journal of Marketing Management*, 14(7), pp. 691–704. doi: 10.1362/026725798784867635.

Boyd, danah m. and Ellison, N. B. (2007) ‘Social network sites: Definition, history, and scholarship’, *Journal of Computer-Mediated Communication*, 13(1), pp. 210–230. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00393.x/full/ (Accessed: 23 July 2014).

Braun, V. and Clarke, V. (2006) ‘Using thematic analysis in psychology’, *Qualitative Research in Psychology*, 3(July 2015), pp. 77–101. doi: 10.1191/1478088706qp063oa.

British Airways (2014) *Twitter*, *Twitter page*. Available at: https://twitter.com/British_Airways.

Brunner, C. B. and Ullrich, S. (2014) ‘How to deal with a negative online consumer review : Can different response scenarios from various sources rebuild consumers ’ product purchase intentions?’, in *47th Academy of Marketing Conference*. Bournemouth, 7-10 July 2014: Academy of Marketing.

Bruns, A. and Stieglitz, S. (2013) ‘Towards more systematic *Twitter* analysis: metrics for tweeting activities’, *International Journal of Social Research Methodology*, 16(2), pp. 91–108. doi: 10.1080/13645579.2012.756095.

Buttle, F. a. (1998) ‘Word of mouth: understanding and managing referral marketing’, *Journal of Strategic Marketing*, 6(3), pp. 241–254. doi: 10.1080/096525498346658.

Canhoto, A. I. and Clark, M. (2013) ‘Customer service 140 characters at a time: The users’ perspective’, *Journal of Marketing Management*, 29(5-6), pp. 522–544. doi: 10.1080/0267257X.2013.777355.

Cha, M., Haddai, H., Benevenuto, F. and Gummadi, K. P. (2010) ‘Measuring User Influence in Twitter : The Million Follower Fallacy’, in *International AAAI Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence (AAAI), pp. 10–17. doi: 10.1.1.167.192.

Chamlertwat, W. and Bhattarakosol, P. (2012) 'Discovering Consumer Insight from Twitter via Sentiment Analysis.', *Journal of Universal Computer Science*, 18(8), pp. 973–992. Available at: http://jucs.org/jucs_18_8/discovering_consumer_insight_from/jucs_18_08_0973_0992_chamlertwat.pdf (Accessed: 13 September 2015).

Chen, J., Haber, E., Kang, R., Hsieh, G. and Mahmud, J. (2015) 'Making Use of Derived Personality : The Case of Social Media Ad Targeting', in *Proceedings of the Ninth International AAAI Conference on Web and Social Media*. Association for the Advancement of Artificial Intelligence (AAAI), pp. 51–60.

Cheng, J., Danescu-Niculescu-Mizil, C. and Leskovec, J. (2015) 'Antisocial Behavior in Online Discussion Communities', in *International AAAI Conference on Web and Social Media, Ninth International AAAI Conference on Web and Social Media*. Oxford, UK, pp. 61–70.

Cheong, M. C.-Y. (2013a) *Inferring Social Behavior and Interaction on Twitter by Combining Metadata about Users & Messages*. Monash University. Available at: http://arrow.monash.edu.au/vital/access/manager/Repository/monash:120048.

Cheong, M. C.-Y. (2013b) *Inferring Social Behavior and Interaction on Twitter by Combining Metadata about Users & Messages*. Monash University.

Cheong, M. and Ray, S. (2011) 'A literature review of recent microblogging developments', *Victoria, Australia: Clayton School of …*. Available at: http://www.csse.monash.edu.au/publications/2011/tr-2011-263-full.pdf.

Chu, S.-C. and Kim, Y. (2011) 'Determinants of consumer engagement in electronic word-of-mouth (eWOM) in social networking sites', *International Journal of Advertising*, 30(1), p. 47. doi: 10.2501/IJA-30-1-047-075.

Cotelo, J. M., Cruz, F. L., Troyano, J. A. and Ortega, F. J. (2015) 'A modular approach for lexical normalization applied to Spanish tweets', *Expert Systems with Applications*. Elsevier Ltd, 42(10), pp. 4743–4754. doi: 10.1016/j.eswa.2015.02.003.

Dichter, E. (1966) 'How Word-of-Mouth Advertising Works', *Harvard Business Review*, Nov-Dec, pp. 147–166.

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A. and Madden, M. (2015) *Social Media Update 2014*. Available at: http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf.

Ellison, N. B. and Boyd, D. M. (2013) 'Sociality through social network sites', in Dutton, W. H. (ed.) *The Oxford Handbook of Internet Studies*. Pre-press. Oxford: Oxford University Press, pp. 151–172. doi: 10.1093/oxfordhb/9780199589074.001.0001.

Fang, E., Palmatier, R. W. and Evans, K. R. (2008) 'Influence of customer participation on creating and sharing of new product value', *Journal of the Academy of Marketing Science*, 36(3), pp. 322–336. doi: 10.1007/s11747-007-0082-9.

Gallaugher, J. and Ransbotham, S. (2010) 'Social media and customer dialog management at Starbucks', *MIS Quarterly Executive*, 9(4), pp. 197–212. Available at: http://misqe.org/ojs2/index.php/misqe/article/view/301 (Accessed: 25 October 2014).

Granovetter, M. S. (1973) 'The strength of weak ties', *American Journal of Sociology*, 78(6), pp. 1360–1380. doi: 10.1037/a0018761.

Groeger, L. and Buttle, F. (2014) 'Word-of-mouth marketing influence on offline and online communications: Evidence from case study research', *Journal of Marketing Communications*, 20(1-2), pp. 21–41. doi: 10.1080/13527266.2013.797736.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G. and Gremler, D. D. (2004) 'Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet?', *Journal of Interactive Marketing*, 18(1), pp. 38–52. doi: 10.1002/dir.10073.

Hong, L., Convertino, G. and Chi, E. H. (2011) 'Language Matters in Twitter : A Large Scale Study Characterizing the Top Languages in Twitter Characterizing Differences across Languages Including URLs and Hashtags', in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pp. 518–521.

Huang, Q. and Xu, C. (2014) 'A data-driven framework for archiving and exploring social media data', *Annals of GIS*, 20(4), pp. 265–277. doi: 10.1080/19475683.2014.942697.

Kane, G., Labianca, G. and Borgatti, S. P. (2014) 'What's Different About Social Media Networks? A Framework and Research Agenda', *MIS Quarterly*, 38(1), pp. 275–304. Available at: http://papers.ssrn.com/sol3/Delivery.cfm?abstractid=2239249 (Accessed: 12 October 2014).

Kaplan, A. M. and Haenlein, M. (2010) 'Users of the world, unite! The challenges and opportunities of Social Media', *Business Horizons*, 53(1), pp. 59–68. doi: 10.1016/j.bushor.2009.09.003.

Kietzmann, J. H., Silvestre, B. S., Mccarthy, I. P. and Pitt, L. F. (2012) 'Special Issue Paper Unpacking the social media phenomenon: towards a research agenda', *Journal of Public Affairs*, 12(2), pp. 109–119. doi: 10.1002/pa.

King, R. A., Racherla, P. and Bush, V. D. (2014) 'What We Know and Don't Know About Online Word-of-Mouth: A Review and Synthesis of the Literature', *Journal of Interactive Marketing*. Elsevier B.V., 28(3), pp. 167–183. doi: 10.1016/j.intmar.2014.02.001.

Kozinets, R. V, de Valck, K., Wojnicki, A. C. and Wilner, S. J. S. (2010) 'Networked Narratives: Understanding Word-of-Mouth', *Journal of Marketing*, 74(March), pp. 71–89.

Kozinets, R. V. (2002) 'The field behind the screen: using netnography for marketing research in online communities', *Journal of marketing research*, 39(1), pp. 61–72. Available at: http://journals.ama.org/doi/abs/10.1509/jmkr.39.1.61.18935 (Accessed: 2 August 2014).

Krippendorff, K. (2004) *Content analysis: an introduction to its methodology*. Thousand Oaks, CA.: SAGE.

Laskey, N. and Wilson, A. M. (2003) 'Internet based marketing research: a serious alternative to traditional research methods?', *Marketing Intelligence & Planning*, 21(2), pp. 79–84. doi: 10.1108/02634500310465380.

Leeflang, P. (2011) 'Paving the way for "distinguished marketing"', *International Journal of Research in Marketing*, 28(2), pp. 76–88. doi: 10.1016/j.ijresmar.2011.02.004.

Lima, A. C. E. S., de Castro, L. N. and Corchado, J. M. (2015) 'A polarity analysis framework for Twitter messages', *Applied Mathematics and Computation*. Elsevier Ltd., 270, pp. 756–767. doi: 10.1016/j.amc.2015.08.059.

Maity, S. K., Sarup, J., Sahni, S. and Mukherjee, A. (2015) 'Analysis and Prediction of Question Topic Popularity in Community Q & A Sites : A Case Study of Quora', in *Proceedings of the Ninth International AAAI Conference on Web and Social Media*, pp. 238–247.

Mamic, L. I. and Almaraz, I. A. (2014) 'How the larger corporations engage with stakeholders through Twitter', *International Journal of Market Research*, 55(6), pp. 851–872. doi: 10.2501/IJMR-2013-070.

Mangold, W. G. and Faulds, D. J. (2009) 'Social media: The new hybrid element of the promotion mix', *Business Horizons*, 52(4), pp. 357–365. doi: 10.1016/j.bushor.2009.03.002.

Mangold, W. G. and Smith, K. T. (2012) 'Selling to Millennials with online reviews', *Business Horizons*. 'Kelley School of Business, Indiana University', 55(2), pp. 141–153. doi: 10.1016/j.bushor.2011.11.001.

Marwick, A. (2005) 'I'm More Than Just a Friendster Profile: Identity, Authenticity, and Power in Social Networking Services', *Association for Internet Researchers 6.0*, pp. 1–26. Available at: http://ssrn.com/abstract=1884356.

de Matos, C. A. and Rossi, C. A. V. (2008) 'Word-of-mouth communications in marketing: a meta-analytic review of the antecedents and moderators', *Journal of the Academy of Marketing Science*, 36(4), pp. 578–596. doi: 10.1007/s11747-008-0121-1.

Morstatter, F., Pfeffer, J., Liu, H. and Carley, K. (2013) 'Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose', *Proceedings of ICWSM*, pp. 400–408. doi: 10.1007/978-3-319-05579-4_10.

Muñiz, Jr., A. M. and Schau, H. J. (2007) 'Vigilante Marketing and Consumer-Created Communications', *Journal of Advertising*, 36(3), pp. 35–50. doi: 10.2753/JOA0091-3367360303.

Neuendorf, K. A. (2002) *The content analysis guidebook*. Thousand Oaks, CA.: Sage.

O'Reilly, T. (2005) *What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software.*, *O'Reilly Blog*. Available at: http://oreilly.com/web2/archive/what-is-web-20.html (Accessed: 18 October 2014).

Okazaki, S., Diaz-Martin, A. M., Rozano, M. and Menendez-Benito, H. (2014) 'How to mine brand Tweets Procedural guidelines and pretest', *International Journal of Market Research*, 56(4), pp. 467–489.

Pew Research Center (2013) *How we analyzed Twitter social media networks with NodeXL*.

Prabowo, R. and Thelwall, M. (2009) 'Sentiment analysis: A combined approach', *Journal of Informetrics*, 3(2), pp. 143–157. doi: 10.1016/j.joi.2009.01.003.

Rowley, J. (2004) 'Just another channel? Marketing communications in e-business', *Marketing Intelligence & Planning*, 22(1), pp. 24–41. Available at: http://www.emeraldinsight.com/journals.htm?articleid=854660&show=abstract (Accessed: 18 October 2014).

Schweidel, D. A. and Moe, W. W. (2014) 'Listening In on Social Media: A Joint Model of Sentiment and Venue Format Choice', *Journal of Marketing Research*, LI(August), pp.

387–402. doi: 10.1509/jmr.12.0424.

Spiliotopoulos, T. and Oakley, I. (2014) ‘Predicting Tie Strength with the Facebook API’, in *Proceedings of the 18th Panhellenic Conference on Informatics*. New York, NY, USA, pp. 1–5.

Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H. and Demirbas, M. (2010) ‘Short text classification in twitter to improve information filtering’, *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR ’10*, p. 841. doi: 10.1145/1835449.1835643.

Stephen, A. T. and Toubia, O. (2010) ‘Deriving value from social commerce networks’, *Journal of marketing research*, XLVII(April), pp. 215–228. Available at: http://journals.ama.org/doi/abs/10.1509/jmkr.47.2.215 (Accessed: 21 April 2014).

Surowiecki, J. (2004) *The Wisdom of Crowds*, *How Collective Wisdom Shapes Business Economies Societies and Nations*. New York: Anchor Books. doi: 10.3174/ajnr.A3417.

Walsh, J. and Godfrey, S. (2000) ‘The internet: a new era in customer service’, *European Management Journal*, 18(1), pp. 85–92. doi: 10.1016/S0263-2373(99)00071-7.

Watne, T. A., Cheong, M. and Turner, W. (2014) ‘#Brand or @User ? How Australian “Mass Brewers” and “Craft Brewers” Communicate with Consumers through Twitter’, in *47th Academy of Marketing Conference*. Bournemouth, 7-10 July 2014: Academy of Marketing.

Weinberg, B. D. and Pehlivan, E. (2011) ‘Social spending: Managing the social media mix’, *Business Horizons*. ‘Kelley School of Business, Indiana University’, 54(3), pp. 275–282. doi: 10.1016/j.bushor.2011.01.008.

Williams, M. and Buttle, F. (2014) ‘Managing negative word-of-mouth: an exploratory study’, *Journal of Marketing Management*, (July), pp. 1–25. doi: 10.1080/0267257X.2014.933864.