Spatialities, Social Media and Sentiment Analysis: Exploring the Potential of the Detection Tool SentiStrength

Christina Reithmeier, Karoline Buschbaum and Detlef Kanwischer

Goethe University Frankfurt/Main, Germany

Abstract

Social media such as Twitter or Facebook not only create new spaces of interaction and communication, they also influence the way we perceive things and lead to changes in our self-perception and our own worldview. Online data occur in various forms and can contain opinions or expressions of feeling. In this article, we explore the potential of SentiStrength, a tool for sentiment analysis in geographic research. We analyse posts on Twitter containing hashtags for possible constructions of spaces in Ostend, a neighbourhood in Frankfurt, Germany. We collected tweets via the Twitter API and used the SentiStrength online application to conduct our sentiment analysis. In order to evaluate the results, we also classified our data manually for comparison. Through its lexicon-based classification, the tool was able to identify positive and negative associations of Ostend. However, we were also able to demonstrate the limitations of the tool compared to manual analysis. Although it provides a quick and comprehensive overview of sentiments, SentiStrength reaches its limits when other media such as images are involved. Overall, the tool offers a good low-threshold approach for scientists to work with digital data.

Keywords:

sentiment analysis, SentiStrength, Twitter, hashtags, construction of space

1   Introduction

More and more aspects of our everyday lives are conveyed, expanded, produced and regulated by software-supported technologies. Social media in particular, such as Twitter, appear on the Internet as a new space for interaction and communication. Strohmaier and Zens (2014) even attribute social media with the ability to change public social life structurally. Poorthuis et al. (2014) describe this structural change as a 'data revolution', in which digital social data could become key elements of social interaction. The information disseminated by such media represents spatial data that has been provided by the users, either voluntarily by linking their location, or unintentionally. The postings implicitly reproduce geographical information, for example through geolocation, giving information about the user's place of origin or referring in their content to geographical entities (places), through which the locations can be perceived differently and acquire new meanings

(Stefanidis et al., 2013). Whether in social networks, on television or as graffiti on house walls, we often encounter so-called hashtags (# symbols, such as #metoo or #BlackLivesMatter). Hashtags emphasize individual terms and bundle certain topics (Bruns & Stieglitz, 2012). The occurrence of hashtags can be very limited in time and still provide a large amount of up-to-date data (Willis & Fecteau, 2016). Since two-thirds of all social media users provide information about their location in a number of ways – whether explicitly via geolocation or via hashtags – those data can be used to investigate spatial perceptions in social media (Stefanidis et al., 2013). Such material with information about geographical location as well as socio-scientifically relevant data can provide new insights. In particular, social media platforms provide an unprecedented opportunity to collect and analyse fine-grained data about socio-spatial actions. Combined with other sources of information and methods, the approach offers great potential for opening up a new field of research: understanding the (geo)web as a socially produced space (Crampton et al., 2013). The Internet, with its wealth of data, offers a supposedly fruitful research base. However, traditional rules and research methods cannot or can only partially be applied to internet-based data, which require new methodological approaches.

In our explorative study, we focus on sentiment analysis as a method for detecting opinions and sentiment in social media on certain topics (Thelwall et al., 2011). Sentiment analysis can potentially give an overview of opinions and perceptions of certain places and help us to understand constructions of those spaces. These new computer-aided approaches usually require basic computing or programming skills, but social scientists often do not have this kind of knowledge and rely on low-threshold programs and tools. One of these tools is SentiStrength, which can be used as a desktop version or online. For the purposes of our study, we used the online version because it is the access point with the lowest threshold and no advanced computing skills are needed. Therefore, our underlying research questions are: How useful is a sentiment analysis using SentiStrength regarding opinions about a neighbourhood? What problems arise in the exploration of spatial constructions from social media?

This article starts with an outline of sentiment analysis in current non-spatial and geographic research, and gives a short overview of various tools for sentiment analysis. We present our case study: a brief explorative analysis of tweets that tagged Frankfurt's Ostend neighbourhood, in which we trace and interpret opinions in social media through sentiment analysis using SentiStrength. The paper closes with a discussion of whether SentiStrength is a helpful tool for scientists lacking programming skills, and an outlook on how sentiment analysis can successfully be used in geographic research into questions of spatialities.

2   Sentiment analysis in current research

Sentiment analysis is a specialized method for identifying, classifying and measuring emotions, opinions or attitudes in textual form (Thelwall et al., 2011; Xia et al., 2016). It was originally designed to automatically extract customer opinions about certain products or brands (Thelwall et al., 2011). Sentiment analysis can be performed using computer-aided methods such as lexicon-based approaches or supervised machine learning, or a combination of both (Dhaoui et al., 2017). Lexicon-based approaches rely on a dictionary of opinion

words and classify sentiment as positive or negative (Dhaoui et al., 2017). Machine learning approaches can be conducted in pre-known or unknown categories via fully automated clustering, LDA or computer-assisted clustering (Amplayo & Song, 2017; Colace et al., 2015; Habernal et al., 2015).

With the proliferation of social networks and the digitization of research methods, sentiment analysis is increasingly applied to content from social media such as Twitter or Facebook (Amplayo & Song, 2017; Habernal et al., 2015; Xia et al., 2016). Habernal et al.'s sentiment analysis looked at Facebook posts written in Czech that contained opinions on particular brands. Lim et al. (2017) used sentiment analysis of data from social media to identify latent infectious diseases. They assumed that outbreaks could be traced through social media and suggest a clustering method to capture unknown categories. Öztürk and Ayvaz (2018) used a lexicon-based approach for investigating public opinions and sentiments towards Syrian refugees expressed on Twitter. They analysed tweets in Turkish and English and discovered that Turkish-language tweets carried more positive sentiment towards Syrian refugees than the English-language ones.

At present, there is little research which combines sentiment analysis and geospatial research. The majority of studies deal with the spatial distribution and location of tweets containing opinions or sentiments about certain topics (Coletto et al., 2016; Cooper, 2017; Daniulaityte et al., 2017; Lu et al., 2015; Mirani & Sasi, 2016; Zhu & Newsam, 2016). For example, Zhu and Newsam (2016) analysed geotagged photos with a deep learning-based classifier that predicts the sentiment based on the image. They concluded that different emotions showed a variety of spatial distributions. Lu et al. (2015) analysed tweets to explore underlying trends in positive or negative geographically-related sentiment with respect to disasters, aiming to reveal interesting patterns in disaster scenarios. They propose a visual analytics framework in order to observe the distribution of tweets, and study sentiment predictions generated by different models.

However, these studies rely on georeferenced posts and do not go beyond the spatial location and distribution of opinions. Therefore, this kind of research does not address possible implications for space or the social processes behind the geotags. Sentiment analysis is often conducted by computer scientists, and little research has been done that focuses on spatial or geographic research questions. We found only two studies which deal with possible spatial ramifications and applications of tweets and other postings that contain opinions and sentiment. You and Tuncer's (2016) study aimed to develop a crowd-calibrated geo-sentiment analysis mechanism that would study public sentiment expressed in social media with regard to place design, the outcome of which could aid local authorities, urban designers or city planners. Zhang and Feick (2016) demonstrate how spatially referenced tweets can shed light on citizens' transportation and planning concerns. They suggest that researchers can use sentiment expressed in geosocial media to identify geographies of public perceptions concerning public facilities and services.

Tools for sentiment analysis

There are a variety of tools for performing a sentiment analysis in scientific research. The tools range from commercial social monitoring programs to free open source applications.

Commercial tools are expensive and often do not meet researchers' requirements, leaving open source applications as a good alternative. The Java-based Stanford Core Natural Language Processing Toolkit provides a set of tools for analysing human language. The sentiment analysis tool is based on deep learning and uses decision trees (Manning et al., 2014). Another tool is SentiWordNet, which is based on the lexical dataset WordNet (Baccianella et al., 2010). The Weka 3 software is a Java-based collection of machine learning algorithms for data mining and includes tools for data preprocessing, classification, regression, clustering, association rules and visualization (University of Waikato, 2017). Another option is GATE. Its software is designed for solving text-processing problems and offers a variety of plugins (University of Sheffield, 2018). All of these tools and applications are open source and can be downloaded; using them requires basic computing skills and knowledge. They also include a variety of languages.

In contrast, Sentiment viz is an online application for sentiment analysis. It offers a variety of features such as topic clustering and tag clouds. However, it is based on the Twitter API and therefore only displays tweets from the last seven days. SentiStrength is available as a downloadable Windows-based version, as a commercial Java-based version, and as an online application. It is based on a lexicon, divides sentences into words, and is available in multiple languages (Thelwall, 2017).

We took a closer look at Weka, GATE and the SentiStrength desktop version but failed to get Weka and GATE to work properly. Weka produced an error message when we tried to implement our dataset. GATE was more user-friendly thanks to its interface, but still required a plugin for sentiment analysis, which was not installable. The SentiStrength desktop version was the most-user friendly tool. Implementing our dataset as a text file did not cause any problems. However, it only counted the number of words a tweet contained rather than giving results on the sentiment. We also tried to edit the lexicon file, but failed. Consequently, we used the online version, which listed more complete results. SentiStrength is the tool with the lowest threshold and thus seems to be a promising application for scientists with limited computing skills.

3   Case Study: tweets containing #ostend #frankfurt

In our current research project, 'Viral #constructions of space in cultural educational processes', our aim is to examine teenagers' perceptions of the Frankfurt neighbourhood Ostend in the age of digitalization, and to evaluate the implications for cultural education. Ostend is a neighbourhood in transition and has a young population. Teenagers are now growing up in a world mediated by social media, which shape their views and perceptions. Thus research on such new media should deliver new insights into spatial and geographic questions.

Over a six-month period, we will be collecting posts on Twitter, Facebook and Instagram that include hashtags referring to Ostend in order to survey the various constructions of space of this neighbourhood that emerge from social media in general. Furthermore, a sentiment analysis will be conducted to identify opinions concerning the neighbourhood.

This article presents a short exploratory study to test the applicability of a sentiment analysis for socio-spatial research purposes, using opinions on Ostend as a test case. Based on hashtags like #ostend #frankfurt, we evaluated the feasibility of using SentiStrength for sentiment analysis to detect spatial constructions on Twitter. Our underlying research questions are: How useful is a sentiment analysis using SentiStrength? What problems arise in the exploration of spatial constructions from social media?

Sentiment analysis poses a challenge for scientists since it requires certain skills. While every step in a sentiment analysis – from data collection to preprocessing to analysis – is largely based on programming skills, few social scientists have these skills. For those without programming skills, data collection via the Twitter API and the classification of the tweets using the SentiStrength online application are good options. Moreover, SentiStrength does not require the preprocessing step, which results in a reduction of data cleansing time. We are aware that sentiment analysis usually requires a certain topic, and that the collection will include all the posts containing that topic. In our case, the 'topic' is the tagged neighbourhood Ostend. Consequently, the tweets collected will include a range of topics relating to the neighbourhood.

Data collection and analysis

First, we investigated tweets from October to December 2017 in which Ostend Frankfurt was marked with a hashtag. The intention was to create a small dataset, since we also wanted to be able to analyse our data manually. We used the Twitter API's search function to collect the tweets. However, in its basic version this search function usually provides only a selection of posts. In addition, our own preselection of hashtags excludes tweets without these hashtags. Furthermore, online data is transitory.

In order to capture as many relevant tweets as possible, we synchronized a glossary of search terms such as #Ostend #ffm with the database. We searched for the 26 hashtags in our glossary individually, before copying the results into an Excel spreadsheet. Here, we added information on user, date and URL, and additional content such as pictures, videos and links. The data was then cleansed. Since an area known as Ostend can be found in multiple cities, we rejected those tweets that referred to anywhere other than Frankfurt's Ostend. Numbering 184, these Ostend tweets were far more numerous than the Frankfurt-related ones. We manually collected a total of 89 Frankfurt-related tweets written by 69 different users. Six further posts had to be excluded, since they were written in Dutch and Indonesian. This left 48 tweets in German and 35 in English. We then classified them into those that referred directly to the actual neighbourhood and those that did not. Among the latter, for example, were tweets marked #ecb (European Central Bank), which is located in Ostend, but the tweets were aimed at EU policies. We kept the tweets that did not refer directly to the neighbourhood in our database in order to help compare the automatic and the manual analyses.

For the automated analysis, the online version of SentiStrength, which classifies sentences based on individual words, was used to classify the collected data. If words occur which the program recognizes as positive, they are given a positive value, whilst those that are recognized as negative are given a negative value. The scale ranges from plus 5 for very

positive scores to minus 5 for very negative ones. The value zero represents a neutral attribution. SentiStrength offers different classification forms: dual, binary, trinary and scale. We decided to use the dual scale, since it classifies the text from -5 to +5 (strong negativity to strong positivity) (Thelwall, 2017). In a third step, for comparison, we classified the posts manually into positive, neutral and negative. During this step, we took a closer look at the topics of our collected tweets but did not create categories.

4 Results

Spatial construction of Frankfurt's Ostend in social media through sentiment analysis

Overall, the SentiStrength online tool classified 79% of the tweets as 'neutral' (0) (66 tweets), 16% as 'not negative' (1) (13 tweets) and 5% as 'not positive' (-1) (4 tweets) (see Figure 1). For our categories 'referring to the neighbourhood' and 'not directly referring to the neighbourhood', SentiStrength classified 41 tweets referring to the neighbourhood as 'neutral' (77%), 9 as 'not negative' (17%), and 3 as 'not positive' (6%). Tweets not referring directly to the neighbourhood were classified as follows: 25 'neutral' (83.3%), 4 'not negative' (13.3%), and 1 'not positive' (3.3%).

By comparison, we classified 48% of all the tweets as 'neutral' (40 tweets), 33% of the tweets as 'positive' (27 tweets), and 19% as 'negative' (16 tweets) (see Figure 1). We identified 18 tweets referring to the neighbourhood as 'neutral' (34%), 26 as 'positive' (49%), and 9 as 'negative' (17%). Finally, we categorized 22 (73.3%) tweets not directly referring to the neighbourhood as 'neutral', 1 as 'positive' (3.3%), and 7 as 'negative' (23.3%). In sum, of all 83 tweets, we classified 38 in the same way as SentiStrength, which presents 46% agreement.
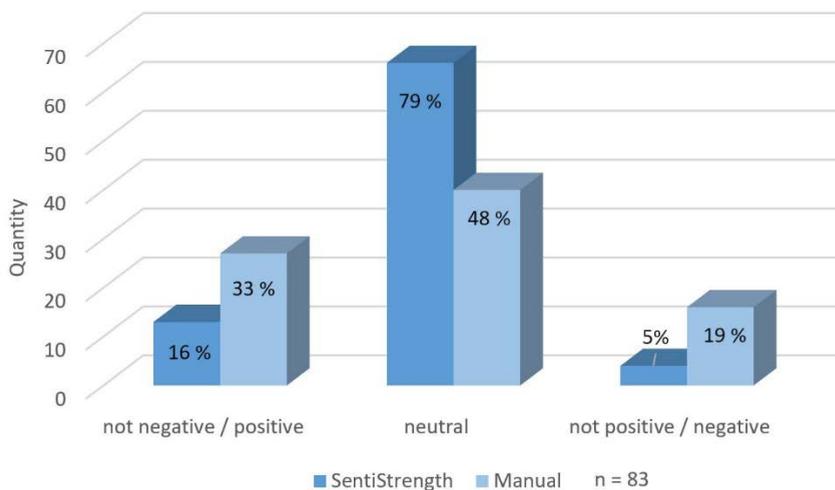

Figure 1: SentiStrength's classification of all tweets referring to Ostend Frankfurt in comparison to manual classification (Source: Own Illustration, 2018)

With 41 tweets classified as 'neutral', 9 as 'not negative' and 3 as 'not positive', SentiStrength leaves us with a few meaningful constructions of space. A tweet containing a reference to a free art exhibition at an adult education centre suggests that Ostend is a place of culture. Another tweet refers to improvement of kerbs in the neighbourhood, suggesting that transport issues are being dealt with. Tweets in which users expressed their joy about the skyline give the impression that Ostend is perceived as a beautiful place.

The classifications 'not positive' and 'neutral' gave us an insight into some of the weaknesses of the online applications. As Thelwell (2017) has already pointed out, SentiStrength has difficulties detecting sarcastic or ironic content, and political or controversial comments. Consequently, the tool will have a lower accuracy when applied to datasets that include a large number of such comments. Some of the tweets we collected fall into this category (see Figure 2).

Figure 2: Ironic tweet with photo attached (Source: https://twitter.com/_blickwinkel_/status/919633264083636224)

The tweet translates as 'How can anyone live in a big city, so ugly', which given the emoticons used and the photos attached is clearly an ironic statement. SentiStrength classified this tweet as 'not positive'. Another problem we faced was the evaluation of nouns. SentiStrength did not rate any of the nouns, but nouns can also convey sentiment or point towards a certain opinion. For example, tweets containing the nouns 'Tötungsdelikt' (homicide offence) or 'Überfall' (robbery) were rated 'neutral'. We also had two cases in which SentiStrength rated '#Germany' and '#GermanDAX' as negative (-2) for no comprehensible reason, because the content of the tweets was not negative in any way. Dialect or colloquial language also posed a problem. A tweet containing the word 'Morsche', meaning 'good morning' in Hessian but 'rotten' or 'brittle' in German, was rated 'not positive'. In another example, SentiStrength could not identify '#dopeshit' as a positive sentiment due to it being colloquial (see Figure 3).

> Dopeness this kid
> #proud #skate #ffm #hafen #osthafen
> #dopeshit #luckyskate #skate #lucky @...
> instagram.com/p/BaO56yEhnGb/
>
> 8:08 AM - 14 Oct 2017 from Frankfurt on the Main, Germany

Figure 3: Tweet containing colloquial language (Source: https://twitter.com/DjMenelik/status/919218251971772416)

The fact that the overwhelming majority of tweets were neutral can be ascribed to the fact that they contained a picture or a link to a picture on Instagram, a photo-sharing application. Although positive sentiment is often linked to a certain place in the picture, SentiStrength is not able to analyse pictures or videos. For example, a tweet containing a picture tagged with '#frankfurtliebe' (#FrankfurtLove) was rated as neutral, as was a tweet containing a picture and the words 'nett da' (nice here).

Our own manual analysis of the material results in more positive constructions of space, as we identified a significantly higher number of positive tweets than SentiStrength. We also classified the tweet praising the improved kerbs, the art exhibition and the skyline as positive. There are tweets in which deliberately positive vocabulary is used to praise food and restaurants, suggesting that the neighbourhood is a good place to eat. Even though positive descriptive adjectives were used here, SentiStrength rated these tweets as neutral. A high number of tweets contain an image, which establishes the positive sentiment towards a certain place within Ostend or the Frankfurt skyline. By contrast, some tweets contain negative sentiment. For example, regarding the police asking for assistance in the investigation of crimes, a negative feeling arises through associations such as danger and violence. In another tweet, a user expresses his complaints about the redesign of Ostend. Again, there is a negative feeling through the context in which the word 'Gentrifizierung' (gentrification) is used, the context implying that the author disagrees with the transformation of the neighbourhood.

While a rather positive impression of Ostend emerges after the first manual analysis of the data, the automated sentiment analysis presents a surprisingly neutral picture.

5   Discussion: Potentials of SentiStrength in geographic research

The automatic and manual sentiment analysis of our data gave a first impression of the neighbourhood Ostend. By classifying the tweets into neutral, positive and negative, SentiStrength helped us identify a variety of constructions of space. Nonetheless, we encountered some difficulties and weaknesses with the online application. The desktop version of SentiStrength offers an adaptable lexicon for the English and German languages and will probably solve some of the issues involved, such as how to judge nouns, exclamation marks, colloquial language and certain adjectives. This would involve a comprehensive initial adaption of the program. By modifying the lexicon, the evaluation of

nouns and the weighting of adjectives or punctuation could be improved. The possibility of adapting the lexicon used for the online version could also lead to an improvement. However, issues with sarcastic or ironic content and tweets consisting of images or videos will persist (Canhoto & Padmanabhan, 2015; Thelwall, 2017). It should be noted that different tools make different classifications, as Lu et al. (2015) have pointed out. Pang and Lee (2008) also note that sentiment and subjectivity are always context-sensitive and domain-dependent. Whether something is labelled positive or negative is not consistent across domains. A further problem is the limit of 100 characters in the online version, which requires splitting up tweets, since Twitter allows up to 280 characters. An extension of the character limit could lead to a more accurate classification.

Evaluation tools such as SentiStrength, which can be used without programming skills, have already been developed for sentiment analysis. However, when using such tools, researchers are forced to accept the presets and algorithms that are unknown to them (Canhoto & Padmanabhan, 2015). Our manual analysis of the classified data material also reveals spatial constructions. It is clear that Ostend is currently undergoing a transformation, which is changing the district not only structurally but also according to its demographics. This is evident in both positive and negative contributions. On the one hand, technological progress is evidenced by the fact that some companies based there use online platforms for self-promotion. Of course, these contributions are deliberately positive since they are intended for advertising purposes. On the other hand, this same technological progress makes it easier for individuals to disseminate their political opinions to a broad section of the population, thus furthering the stigmatization of other individuals.

Sentiment analysis itself could be used for different scenarios. With this method, we can identify opinions and feelings about tagged neighbourhoods, but posts and tweets on specific topics tagged with the hashtag of the district name can also be analysed, and positive as well as negative aspects of a neighbourhood can be identified. Such an analysis can serve as an overview on how the neighbourhood is perceived and produced in social media. Canhoto and Padmanabhan (2015) suggest that sentiment analysis needs also to consider the social context within which a conversation takes place. As Zhang and Feick (2016) show, the results of such an analysis have the potential to boost citizens' contributions to society. This leads to the concept of spatial citizenship, in which citizens are able to access, read, interpret and critically reflect on spatial information, and express location-specific opinions using geo-media (Gryl & Jekel, 2012).

Data collection without programming skills is only possible with the platform-specific search tools (APIs) of the various social networks, or commercial social monitoring programs such as talkwalker and twitcident. A good alternative is R, which requires only minimal training and knowledge, and enables social researchers to collect tweets automatically. All of these programs can help to collect the desired datasets, but they reach their limits when it comes to distinguishing between different places with the same name. This necessitates an initial visual identification and manual selection of possible hashtags with reference to certain places. Searching for spatial hashtags and not for a certain topic, one will collect tweets containing different topics. In the next step, these tagged tweets can be categorized by topic. It is also worth pointing out that a larger dataset would probably require an automatic content analysis (Reithmeier et al., 2016).

6   Outlook

So far, sentiment analysis has been used either within computer science research or for the purpose of showing the spatial distribution of postings. In these contexts, it has proved to be a useful instrument for obtaining new knowledge about spatial distribution patterns. However, as soon as qualitative research questions are posed, there are certain aspects that must be taken into account, as this exploratory study has shown. Research questions that allow qualitative results and statements thus always refer to an evaluation of known topics and spatial limitations. Although the hashtag 'Ostend' represents a spatial limitation, no thematic categorization was conducted, making it difficult to relate spatial constructions to specific locations. It would therefore be necessary to ask about specific events or locations. For Ostend, one could, for example, specifically ask about a favourite place among young people, which would make prior identification of such places indispensable. Furthermore, sentiment analysis could focus on tweets on topics like gentrification and give a long-term overview of ongoing changes in a neighbourhood. This could also highlight spatio-temporal changes in perception.

Even in a larger study area, different spatial constructions can be identified using this method, which offers multiple potentials for geographers to explore new approaches. SentiStrength online provides access to computer-based research methods for social scientists lacking basic programming skills, but it is not applicable to every research purpose. This is due to the preset algorithms, which most social scientists would be unable to change, because at least a basic knowledge of programming is necessary. In order to achieve ideal results, the preset dictionary would also have to be vastly extended and adapted for particular research topics. This leaves many researchers with a limited number of functions, and the full potential of the tool remains unused. Further research in this area would have to adapt to the preset algorithms and dictionary extensions or develop its own classification methods. Training in the utilization of different tools for sentiment analysis is absolutely necessary for conducting an extensive comparison between the different tools and methods. Possible future research should also go beyond the posts and analyse the participating users, looking at their behaviour and intentions.

References

Amplayo, R. K. & Song, M. (2017). An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews. *Data & Knowledge Engineering*, 110, 54-67.

Baccianella, S., Esuli, A. & Sebastiani, F. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010*, 2200-2204.

Bruns, A. & Stieglitz, S. (2012). Quantitative Approaches to Comparing Communication Patterns on Twitter. *Journal of Technology in Human Services*, 30(3-4), 160-185.

Canhoto, A.I. & Padmanabhan, Y. (2015). 'We (don't) know how you feel' – a comparative study of automated vs. manual analysis of social media conversations. *Journal of Marketing Management*, 31(9-10), 1141-1157.

Colace, F., Casaburi, L., De Santo, M. & Greco, L. (2015). Sentiment detection in social networks and in collaborative learning environments. *Computers in Human Behavior*, 51, Part B, 1061-1067.

Coletto, M., Esuli, A., Lucchese, C., Muntean, C.I., Nardini, F.M., Perego, R. & Renso, C. (2016). Sentiment-enhanced Multidimensional Analysis of Online Social Networks: Perception of the Mediterranean Refugees Crisis. *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining ASONAM 2016,* 1270-1277.

Cooper, A-P. (2017). Assessing the Possible Relationship between the Sentiment of Church-related Tweets and Church Growth. *Studies in Religion-Sciences Religieuses*, 46(1), 37-49.

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W. & Zook, M. (2013). Beyond the Geotag: situating 'big data' and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*, 40(2), 130-139.

Daniulaityte, R., Lamy, F.R., Smith, G.A., Nahhas, R.W.), Carlson, R.G., Thirunarayan, K., Martins, S.S., Boyer, E.W. & Sheth, A. (2017). 'Retweet to Pass the Blunt': Analyzing Geographic and Content Features of Cannabis-Related Tweeting Across the United States. *Journal of Studies on Alcohol and Drugs*, 78(6), 910-915.

Dhaoui, C., Webster, C.M, & Tan, L.P. (2017). Social media sentiment analysis: lexicon versus machine learning. *Journal of Consumer Marketing*, 34(6), 480-488.

Gryl, I. & Jekel, T. (2012). Re-centering GI in secondary education: Towards a spatial citizenship approach. *Cartographica: the international journal for geographic information and geovisualization*, 47 (1), 18-29.

Habernal, I., Ptáček, T. & Steinberger, J. (2015). Reprint of 'Supervised sentiment analysis in Czech social media'. *Information Processing and Management*, 51(4), 532-546.

Lim, S., Tucker, C. S. & Kumara, S. (2017). An unsupervised machine learning model for discovering latent infectious diseases using social media data. *Journal of Biomedical Informatics*, 66, 82-94.

Lu, Y., Hu, X., Wang, F., Kumar, S., Liu, H. & Maciejewski, R. (2015). Visualizing Social Media Sentiment in Disaster Scenarios. *WWW'15 Companion: Proceedings of the 24th International Conference on World Wide Web*, 1211-1215.

Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J. & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55-60.

Mirani, T.B., & Sasi, S. (2016). Sentiment analysis of ISIS related Tweets using Absolute location. *2016 International Conference on Computational Science & Computational Intelligence (CSCI)*, 1140-1145.

Öztürk, N. & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Information*, 35(1), 136-147.

Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.

Poorthuis, A., Zook, M., Shelton, T., Graham, M. & Stephens, M. (2014). Using Geotagged Digital Social Data in Geographic Research. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2513938 (December 2017).

Reithmeier C., Buschbaum, K., Blitz, A. & Kanwischer, D. (2016). 'Heaven. #shopping #Frankfurt #weekend #joy' – Hashtags, Constructions of Space, and Geography Education. *GI_Forum 2016, Vol. 1*, 282-294.

Stefanidis, A., Crooks, A. & Radzikowski, J. (2013). Harvesting ambient geospatial information from social media feeds. *GeoJournal*, 78, 319-338.

Strohmaier, M. & Zens, M. (2014). Analyse Sozialer Medien an der Schnittstelle zwischen Informatik und Sozialwissenschaften. In König, C., Stahl, M. & Wiegand, E. (Eds.). Soziale Medien: Gegenstand und Instrument der Forschung. *Schriftenreihe der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute (ASI)*. Wiesbaden: Springer, 73-95.

Thelwall, M. (2017). Heart and Soul: Sentiment Strength Detection in the Social Web with SentiStrength. In Holyst, J. A. (Eds.). *Cyberemotions: Collective emotions in cyberspace*. Cham: Springer International, 119-134.

Thelwall, M., Buckley; K. & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, 62(2), 406-418.

University of Sheffield (2018): GATE – general architecture for text engineering. Retrieved from https://gate.ac.uk/ (April 2018).

University of Waikato (2017). Weka 3: Data Mining Software in Java. Retrieved from https://www.cs.waikato.ac.nz/ml/weka/index.html (December 2017).

Willis, E. R. & Fecteau, A. (2016). Humor and Identity on Twitter: #muslimcandyheartrejects as a Digital Space for Identity Construction. *Journal of Muslim Minority Affairs*, 36(1), 32-45.

Xia, R., Xu, F., Yu, J., Qi, Y. & Cambria, E. (2016). Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis. *Information Processing and Management*, 52(1), 36-45.

You, L. & Tuncer, B. (2016). Exploring public sentiments for livable places based on a crowd-calibrated sentiment analysis mechanism. *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining ASONAM 2016, 693-700.*

Zhang, S. & Feick, R. (2016). Understanding Public Opinions from Geosocial Media. *ISPRS International Journal of Geo-Information*, 5(6), Nr. 74, 1-20.

Zhu, Y. & Newsam, S. (2016). Spatio-Temporal Sentiment Hotspot Detection Using Geotagged Photos. *24th ACM Sigspatial International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2016).*