ORIGINAL RESEARCH

Exploring the relation between festivals and host cities on Twitter: a study on the impacts of Lucca Comics & Games

Yesim Tonga Uriarte[1] · Marinella Petrocchi[2] · Maria Luisa Catoni[1] · Stefano Cresci[2] · Rocco De Nicola[1] · Maurizio Tesconi[2] · Rafael Brundo Uriarte[3]

Abstract

This article investigates the relation between festivals and their destination cities. We examine what festival attendees think about the festival experience and the host city, and how the online image of the host city is shaped through the dissemination of these opinions on Twitter. We conduct a case study and analyze the opinions about a big scale event, Lucca Comics & Games and its host city, Lucca. Our results indicate that the festival significantly reinforces the online image of the host city on social media and not only promotes the existing strong characteristics of the city, such as historic monuments and architectural elements, but also enriches the portrayal of the city by relating the existing legacy to the fantasy culture with the festival experience. This study sheds light on the social aspects of emergent digital information practices and platforms within the context of festivals. Furthermore, our methodological framework serves as a model for prospective research on the relationship between events and cities and contributes to the development of a common methodological framework for the empirical assessment of this relation.

Keywords Twitter · Social media · Festivals · Online destination image · Digital identity · Comic-con · Lucca Comics & Games

Yesim Tonga Uriarte and Marinella Petrocchi contributed equally to this manuscript.

✉ Yesim Tonga Uriarte
  yesim.tonga@imtlucca.it

1   IMT School for Advanced Studies Lucca, Piazza San Francesco 19, 55100 Lucca, Italy

2   Institute of Informatics and Telematics (IIT), National Research Council, Via Giuseppe Moruzzi 1, 56124 Pisa, Italy

3   Kunsthistorisches Institut in Florenz–Max-Planck-Institut, Via Giuseppe Giusti, 44, 50121 Firenze, Italy

 Springer

1 Introduction

'With the emergence of computational social science, there is an opportunity to consider large-scale, and sometimes "messy," detailed datasets of social interaction' (Murthy et al. 2016). Computer-mediated communication, particularly social media, which are a powerful tool for sharing information, opinions, experiences and sentiments (Litvin et al. 2008; Xiang and Gretzel 2010), and emerging smart event experiences (Bustard et al. 2019), offer fruitful strands for tourism research. One of the topics that can be explored with such data is the relation between festivals and cities.

The increasing phenomenon of the creative city as a brand and the "festivalization" of culture (Bennett et al. 2014) resulted in a growing body of literature on the subject. The scholarly interest on the effects of events on the image of destination cities started to rise in the 1990s. Majority of the previous studies either employ attribute checklists and category-based approaches (which condition and limit the respondents with a list of predefined attributes) or rely on unstructured and interpretive qualitative data (with limited samples of respondents due to exhaustiveness of methods used for analyzing collected qualitative data). On the other hand, social media—providing detailed datasets of social interactions—and big data analysis have a big potential to contribute to on-going research on the relationship between events and host cities. For instance, 'the rapid diffusion of "microblogging" services such as Twitter is ushering in a new era of possibilities for organizations to communicate with and to engage their core stakeholders and the general public' (Lovejoy and Saxton 2012: 337). Festival related communication on social media not only helps host cities to reach a wider public, including people who did not attend the festival, but also provides a rich source of data, which would allow researchers to get an overview of the relation between events and their host cities from diverse perspectives. Nevertheless, despite the prominence of social media in the communication of host cities before, during and after the festival period, little scholarly attention has been dedicated to this topic. Likewise, 'little is known about how social media influences emotions and attachments to brands, and whether social media-based relationships lead to desired outcomes such as positive word of mouth' (Hudson et al. 2015: 68). Thus, Twitter appears as an ideal social media platform to investigate the relation between festivals and cities.

We, therefore, aim to study how opinions about festivals and their host cities are shared on Twitter, 'one of the most popular microblogging platforms worldwide' (Ozdikis et al. 2017: 291). We extract information from tweets and analyze diverse relations among the social media users and the content of their posts through a case study, Lucca Comics & Games (LC&G). The latter is a festival dedicated to fantasy culture and one of the biggest comic-cons in the world.[1] Considering the festival's wide thematic scope—which includes comics, games, movies, cosplay, music, Japan

---

[1] For a detailed analysis of the history and management of Lucca Comics & Games, see Tonga Uriarte et al. (2019a).

and junior,[2] its large audience—estimated around 500,000 attendees,[3] and its wide-scale tourism impact (Tonga Uriarte et al. 2019b), such a case study represents an ideal test-bed. We study whether LC&G influences the frequency of tweets about the city, and how the polarity and the content of tweets reflect the festival experience. To answer these research questions, we rely on a quantitative approach based on four main steps:

- *dataset filtering*: we created five different datasets in order to grasp specific characteristics of different groups of users in relation to the LC&G crowd and the destination city;
- *city and comic-con classification*: we classified collected data in relation to the festival and the city. This step allowed us to conduct further analysis on collected data, thus revealing different characteristics between city- and festival-related tweets;
- *polarity detection*: we conducted sentiment analysis to infer the polarity of collected tweets in Italian and English, indicating positive, negative or neutral content. Furthermore, we investigated the paradigm of information diffusion on social networks, pointing out how so-called influencers are able to spread their viewpoints among their followers (Guille et al. 2013);
- *content analysis*: we focused on city-related tweets in order to investigate users' preferences and opinions about the city.

Furthermore, we explored the amount of people who tweeted about LC&G, even without attending the festival. Attendance has been tested by sending a tweet to those people who were tweeting with the hashtag #luccacomics2016 during the comic-con dates, asking whether they were attending the event. We obtained a benchmark for understanding the scale of the festival and Twitter's role in promoting the city to a wide range of audiences, inclusive of virtual communities who did not attend the festival.

Overall, our results indicate that the festival significantly contributes to the online image of the destination city on Twitter. It not only promotes the existing distinctive features of the city, such as monuments and historic elements but also enriches the city portrayal by adding a dynamic layer, by relating the existing urban legacy to the fantasy culture of the festival experience.

The contribution of this paper is threefold. First, we provide a rich empirical base regarding the impact of a festival on the online image of the host city through one of the most popular microblogging platforms, Twitter. Considering the type of the festival that we use as a case study, our study is encouraging further comparative

---

[2] These are the thematic areas where there are dedicated booths and activities in line with the theme. For instance, in the junior area, there are publishers of illustrated children books and comics as well as activities for children and families.

[3] According to the most recent estimates, the last edition of the festival, in 2019, has been confirmed as the largest and most important cross-media show in the West. About 270,000 ticketing admissions, plus hundreds of thousands of extra tickets'. Source (in Italian): https://www.luccacomicsandgames.com/it/2019/news/grazie-a-tutti/.

analysis. Secondly, our conclusions provide fruitful insights for theoretical inquiries regarding the ways in which opinions about festivals and their destination cities spread through social media. Finally, we leverage a methodological framework that combines a variety of state-of-the-art methods, in different steps. This may serve as a model for prospective research on the relationship between events and cities and may contribute to the development of a common methodological framework for the empirical assessment of this relation.

The rest of the paper is organized as follows. Section 2 gives a comprehensive overview of related previous research, states the problem and provides preliminary notions. Section 3 introduces the case study. Section 4 illustrates the dataset and the methodological framework. Section 5 presents and discusses the results. Section 6 analyses the challenges and benefits of our approach draws conclusions and hints at future research directions.

2 Literature review

2.1 Events and Twitter

Twitter is not only a medium for posting personal updates or status information. According to Risse et al. (2014), Twitter has become a channel for sharing and distributing information of all kinds, a highly effective event news medium, which is much faster than the traditional news channels, and a valuable source of rich information on contemporary society—including sentiments, opinions and news about happenings. 'Tweets give valuable insights into individuals, groups, and organizations, and enable an understanding of the public perception of events, people, products, or companies, including the flow of information' (Risse et al. 2014: 208). Twitter reached 126 million daily active users in early 2019[4] and, thanks to the wide user base, it provides a rich source of data for analyzing and revealing cultural preferences, opinions, perceptions and sentiments, with 'the possibility of a live documentation of our society'. As a result, the scholarly interest in the influence of Twitter and its relation with events has been increasing.

Previous studies on Twitter have a wide scope considering, for instance, its history, transformations and influences (Akcora and Demirbas 2010), users' behaviours and motivations (Kwak et al. 2011; Murthy et al. 2016), its marketing power (Burton and Soboleva 2011; Leung et al. 2015), and, from different disciplinary perspectives, its users from a psychological viewpoint (Golbeck et al. 2011; Golbeck 2019).

Regarding the relation between social media and events, social media has become an increasingly important data source for learning about breaking events (i.e., for event detection) and for following the developments of unfolding or past events (Zubiaga 2019). Event detection involves tracking the stream of updates from social media to identify new events. In Dou et al. (2012), two approaches for event

---

detection are outlined: a retrospective one and an online one. In the retrospective approach, new data collected from social media are checked against a historical data-base of events to determine if the new data are likely associated with a new event (Atefeh and Khreich 2015). In the online (i.e., streaming) approach, new data are clustered as they are collected from the social stream. Then, a new event is detected when enough data are collected that do not cluster with previous data. Such data are then associated to a new cluster, related to the newly detected event (Petrović et al. 2010). In addition to journalism, another favourable application scenario of event detection is emergency management. Within this context, social media-based systems for emergency event detection (e.g., systems for detecting floods, wildfires and traffic jams) are used to provide rapid alarms for places where traditional sensor networks are unavailable, or simply too costly to deploy (Sakaki et al. 2010; D'Andrea et al. 2015).

In the case of scheduled events (e.g., elections), predictive analyses were carried out to estimate the number of participants (de Lira et al. 2019), or to forecast the likely outcome (Tsakalidis et al. 2015). These studies are typically based on the analysis of the textual content of posted messages that are related to the event under investigation, for instance by leveraging techniques of natural language processing and machine learning or, more recently, approaches based on deep learning. The same analyses can also be carried out while the event is unfolding, a task that is typically referred to as "nowcasting" (Lampos and Cristianini 2012; Cresci et al. 2018).

About intrinsically unpredictable events (e.g., natural disasters such as earthquakes) (Avvenuti et al. 2016), the majority of scholarly efforts were devoted to analysing the events as they unfold. In the specific context of natural disasters—the most studied type of unpredictable events—this task is dubbed as the automatic acquisition of situational awareness (Imran et al. 2015). To extract information about the event, scholars again exploited the textual and multimedia content of posted messages. Textual mentions of places/locations are often complemented with the geographic coordinates of such places, via a process called geoparsing (Avvenuti et al. 2018b). Results of these analyses can then be visualized on interactive maps, designed to help to track the unfolding event (Middleton et al. 2013; Avvenuti et al. 2018a, b).

Finally, events that are not time-critical can also be investigated through post-hoc analyses. For instance, in the tourism domain, Brilhante et al. (2015) analyse historical data from Wikipedia and Flickr to build a personalized recommender system for sightseeing tours. Similarly, analysing historical social media data of touristic paths, Muntean et al. (2015) propose a system for predicting next movements of tourists' current trail. In Cresci et al. (2014), authors collected data about accommodations, restaurants and points of interest from several social media including Facebook, Foursquare, Google Places and Booking, and created a searchable Web catalogue called Tourpedia, which includes also the aggregated sentiment scores of the attractions, computed from the collected reviews.

Previous studies allow us to identify three main streams of research on the relationship between events and Twitter: (i) linking tweets to the Semantic Web for event detection (Atefeh and Khreich 2015), for instance, using support vector machines (Sakaki et al. 2010) or statistical clustering methods like a Naive Bayes classifier

and TF-IDF (Sankaranarayanan et al. 2009) or a machine learning approach (Rowe and Stankovic 2012; Xu et al. 2017); (ii) using location estimation methods for events detected in Twitter (Ozdikis et al. 2017; Abdelhaq et al. 2017), relying on GPS geotags, place attachment, tweet texts, and the user profile; and (iii) analyzing the impacts of Twitter on political campaigns and elections, as well as, on civil movements and uprisings; considering for example elections (Jungherr et al. 2016; Anstead and O'Loughlin 2015), the Arab Spring (Mourtada and Salem 2011; Rane and Salem 2011; Mourtada et al. 2011) or Gezi demonstrations (Chrona and Bee 2017).

Within this context, research on *organized* events is an emerging field. For instance, Reinhardt et al. (2009) investigate the reason and the added value of tweeting in conferences, through online surveys. The results indicate that microblogging is not only relevant for the conference participants to share resources, follow parallel sessions and improve visibility, but it also serves as an additional way of discussing related topics and exchanging supplementary information, which goes beyond the physically-limited conference space and provides the possibility to engage anyone in the virtual space. Correspondingly, Wilkinson et al. (2015) investigate the ways social media change conference experiences, through empirical evidence from eight international urology conferences. They argue that Twitter is a very powerful tool that amplifies the content of scientific meetings, by expanding the conference experience to a wider audience.

Moreover, 'sporting events feature among the most popular topics covered on Twitter, both in terms of volume and frequency of updates (Twitter 2010, 2011), with spectators using social media as a backchannel to post their commentary while watching live events' (Highfield 2013). Thus, also such events received great scholarly attention (Filo et al. 2015). For instance, Frederick et al. (2015) examine the agenda-setting on Twitter during the 2012 London Olympics by performing content and frequency distributions analysis and chi-square tests on tweets by the @London2012 account and tweets containing the #London2012 hashtag. The authors, by considering the differences between the tweets sent by @London2012 and the tweets containing #London2012. conclude that the efforts to centralize media messages through the official account were not successful. Interestingly, the same findings also indicate that 'hashtags provide individuals with an opportunity to break away from the effects of a traditional media agenda, thereby creating and disseminating a unique line of thought and opinion' (Frederick et al. 2015: 328).

2.2 Destination images and Twitter

As another rising stream of research, scholars and private initiatives have examined the relations between destination images and social media. Among the existing applications derived from these studies, the Travel Appeal Dashboard[5] platform 'collects, analyses and measures, in real-time, online feedback and reviews posted

---

[5] https://blog.travelappeal.com/learning-centre/improve-your-destinations-reputation.

on all major review channels and social media sites'. One of the methodologies applied to measure the web reputation of a destination as perceived by social media users is the evaluation of their opinions by means of "sentiment analysis": this gave rise to customised reports such as those on visitors' appreciation of Tuscan wines[6] and on Italian cultural sites.[7]

Neidhardt et al. (2017) apply text mining techniques and sentiment analysis to users' comments in an online travel forum, to determine their emotions regarding forthcoming trips. The results of this study show that forum users are influenced by other people's comments and they can change opinions and feelings about destinations and related online reputation.

Focusing on Twitter, Sevin (2013) conducts a comparative study to examine how Twitter is used by destination marketing projects in Illinois, San Francisco, Idaho, Texas, and Milwaukee. The author analyses the overall trends and usage patterns of microblogging, and the relation of social media ecology and place branding. The analysis is based on codebooks and the results indicate that the investigated destination marketing campaigns do not necessarily benefit from interpersonal communication and networking capabilities of Twitter. The selected destinations mostly use Twitter to distribute information and share news about events, such as festivals, concerts or fairs. Lalicic et al. (2019) investigate the communication of destination brands through social media. They use more than 15,000 tweets and 6000 Facebook posts and analyse the use of emotional adjectives in the collected data. 'The study demonstrates that destinations can improve user reaction rates on social media and facilitate a positive destination brand image through the use of specific emotional brand values' (: 63), particularly those that best align with the user's expectations.

Furthermore, Sevin (2014) argues that 'the literature and practice of place branding have focused on the competition of brands at the expense of exploring the relations between people, symbols, meanings, and physical characteristics of cities' (: 47) and proposes an analytical framework combining social and semantic network analysis methods to evaluate place brands. The framework is then applied to a use case by analyzing a sample dataset of tweets about Boston and New York City. The results illustrate some drawbacks of the framework, such as not considering certain competitive aspects of branding, like tourism, exports and foreign direct investments, or the need for a dynamic mapping approach. Furthermore, the author remarks that 'most importantly, place-relevant associations might be highly influenced by certain high-profile events' (Sevin 2014: 55), as testified by the research finding that, in New York's Twitter network, Justin Bieber's appearance on the Today's Show affected the associations. As exemplified by Sevin (2014), different experiences affect destination images and can even change them (Boo and Busser 2006). Thus, new or different experiences at the same destination can create new and/or enriched images of the destination in the perceiver's mind and it is of crucial importance to understand how these short-term events influence long-term associations with the destination reputation.

---

[6] Travel Appeal customized report: In Wine we Trust! How is wine spoken about online?
[7] Travel appeal customized report: Italian Cultural Sites: Reviews and Visitor Satisfaction.

In summary, from the body of work on the relationship between social media and destination image, like Kim et al. (2017) we can conclude that social media are still considered an emerging information source in tourism destination marketing. Overall, the topic of tourists' destination image formation in social media is still under-investigated, and this holds especially for empirical studies. Indeed, most previous studies (e.g., Tham et al. 2013; Ghazali and Cai 2014) adopt a conceptual and theoretical approach. Besides, the few existing empirical studies provide contrasting results, as in the case of Sevin (2013), Sevin (2014) and Lalicic et al. (2019), thus calling for additional research and experimentation.

Moreover, despite a few previous studies that focus on the relations between social media and destination image,[8] to the best of our knowledge, there has not been much scholarly interest in the relation between temporary organized events and the host city through the social media, which provide valuable insights and a big amount of user-generated content. Correspondingly, Williams et al. (2017) argue that 'since the volume of synchronous destination eWOM [electronic Word-Of-Mouth] may grow in future, the role of events as online animators of destination features may need additional critical examination' (: 95). In this regard, our work aims to address the need for such a critical examination of the relationship between events and destination image and to provide a rich empirical study on the topic.

2.3 Impact of events on destination image, identity and reputation

Much effort has also been devoted to study the impact of events on destination image, identity and reputation. In this broad line of research, social media are not explicitly considered, and scholars mainly focus on the impact of events themselves. Analyses have mainly revolved around measures of economic impact, and only seldomly considered other socio-cultural impact measures.

From the methodological viewpoint, several studies leveraged surveys filled by event participants before and after the event (Ninomiya et al. 2019; Folgado-Fernández et al. 2017; Arnegger and Herz 2016; Li and Kaplanidou 2013; Lee et al. 2005; Richards and Wilson 2004). Scholars were interested in assessing the differences between the answers given before and/or after the event, or the differences in answers given by different groups of users (e.g., participants *vs* non-participants, residents *vs* foreigners) and used the observed differences to estimate the impact of the event. The need to rely on surveys for collecting data represents one of the main limitations of these studies. Indeed, the number of participants in the studies were a maximum of some thousands. In this respect, carrying out similar analyses without explicitly asking contributions to event participants and stakeholders, by analysing, like in the present work, spontaneous social media messages, represents a promising, scalable and economically-feasible opportunity. In fact, many recent studies that assess the impact of events through social media, such as the studies discussed in the previous section, often consider as much as tens/hundreds of thousands of users. Moreover, recent work demonstrated the pivotal

---

[8] Also see Sharif and Mura (2019) for the impact of user-generated content on visiting attitudes and destination perception.

role of social media in the tourism sector, given the information-intensive nature of the related industries and their large use of Internet and Web technologies to distribute products and to communicate with customers (Kim et al. 2017).

The previously described survey-based approaches have been mainly applied to sport events, on which a vast literature exists (Ninomiya et al. 2019; Li and Kaplanidou 2013; Lee et al. 2005). On the contrary, festivals received little attention, also because of the somewhat limited economic and developmental impact they have when compared to mega-events such as Olympic Games and World Cups (Moscardo 2007; Richards and Wilson 2004). Despite these limitations, festivals proved, nonetheless, to have a significant socio-cultural impact in local environments, such as towns (Yeoman et al. 2012). Among the most interesting related studies, Florek and Insch (2011) investigate the fit between event and destination image. Their results underline 'the potential to build and enhance the image of the destination, depending on the synergistic effects of the image attributes of the host and event'. At the same time, they also acknowledge that 'a regretful mismatch between the image of the host and the event could taint the image of either or both, and in the long-term harm their reputations'.

3 Case Study: Lucca Comics & Games

The number of festivals dedicated to fantasy culture has been growing in many countries. Interestingly, these festivals—also known as comic-cons—embody a strong combination of artistic and commercial aspects, along with alternative lifestyle narratives.

Nowadays, Lucca Comics & Games is the biggest cultural and commercial event in Italy (and among the biggest in the world) dedicated to fantasy culture as an experiential mass-phenomenon. It takes place at the end of October in Lucca and turns the whole historic city center into a live scenery for five days. The event attracts around 500,000 attendees in total, including ticketed and un-ticketed participants, which is an impressive number if one considers that Lucca has ca. 90,000 inhabitants. Thus, it can also be considered a big cultural tourism phenomenon.

Correspondingly, the festival has a strong and reciprocal relation with the city. Indeed, since the festival program is spread throughout the whole city on a total area of more than 50,000 square meters, the historic center provides a unique imagery for the festival activities and completely merges with the LC&G atmosphere. Thus, the festival turns the destination into a fantasy city and brings a large number of participants, who do not visit Lucca outside of the festival period.

4 Methodological approach

4.1 Datasets description

The 2016 LC&G edition took place from October 28 to November 1.All the Twitter data used in this study are related to LC&G and/or the city of Lucca (except for one

dataset described at the end of this section) and were collected with a social media crawler, which exploits the Twitter public API.[9] Overall, we obtained five different datasets.

The first dataset was collected using Twitter's Streaming API. The tweets were selected based on the occurrences of specific hashtags, related to LC&G 2016 (e.g., #luccacomics, #LuccaComicsandGames, #Luccacg16). The tweets were collected from October 16 to November 2, 2016. We refer to this dataset as *LC&G 2016 Generic dataset*. This dataset includes not only the core followers but also people that could be indirectly or only potentially interested in the event.

The second dataset contains geo-referenced tweets. More specifically, we specified a geographic bounding box to include the historical walls surrounding the city of Lucca, where the comic-con takes place. Then, we collected all the tweets posted from within the bounding box, between October 26 and November 2, 2016. In this way, we constrained the data collection to those tweets posted from within the city centre. We refer to this second dataset as the *Geo dataset*.

The third dataset contains tweets indicating whether users discussing about LC&G were physically attending the LC&G event, or they were tweeting about it from somewhere else (e.g., just following the event via social media and tweeting from other places). This dataset represents, to the best of our knowledge, a novelty with respect to previous studies on event attendance, and on the impact of events on the online image of cities. To build this dataset, we followed a hybrid crowd-sensing approach, where we automatically sent a reply to each tweet containing the hashtag #luccacomics2016 posted during the days of the comic-con. In these replies, we asked users about their physical presence at the event. This strategy is called `hybrid crowdsensing', because it combines an opportunistic sensing phase, where we collected all messages about a specific topic (i.e., #luccacomics2016), with a participatory sensing phase where users were directly contacted and asked a targeted question (Avvenuti et al. 2017). The questions were sent during the event dates (specifically, starting one day before and terminating one day after the event) to a total of 1354 users. We received 531 replies to our questions, which we subsequently parsed to determine if the replying user was actually attending the event in Lucca. This dataset was collected to estimate the number of people that tweeted about LC&G, without physically attending the event. Notably, this strategy has already been profitably used in other studies concerned with event attendance (Cresci et al. 2018). Henceforth, this dataset is referred to as the *Question dataset*.

The fourth dataset was collected in February 2017, to increase the amount of available data by adding tweets created by specific online communities revolving around the comic-con. We started crawling the tweets of 22 artists and 23 exhibitors that were involved in performances and expositions at LC&G 2016. Then, we collected a subset of tweets produced by the followers of such artists and exhibitors. Specifically, we collected all the artists' and exhibitors' profile information and up to the 3200 most recent tweets posted by these accounts. We also collected the full lists of their followers and friends (also known as followings). Then, for each of

---

9 https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter.

Table 1 Statistics about the datasets used in our study

Dataset	Tweets IT	Tweets EN
LC&G 2016 generic	12,428	4006
Geo	3689	2556
LC&G communities	1,141,235	1,556,213
Contrastive	4,218,570	1,572,160
Question	1355	–

Table 2 Number of tweets after the filtering and the classification process

Dataset	Language	Relevant	City	Comicon
LC&G 2016 generic	IT	11,749	320	11,644
	EN	3019	35	3008
Geo	IT	933	171	787
	EN	659	144	625
LC&G communities	IT	6840	645	4807
	EN	1305	92	795
Contrastive	IT	541	119	439
	EN	4	3	1
Question	IT	532	–	–

the 45 accounts, we: randomly sampled up to 1000 followers; collected their profile information and up to 200 of their most recent tweets. Overall, we collected 69,241 tweets by 45 artists and exhibitors, as well as 3,398,467 tweets by 25,006 of their followers. We refer to this dataset as the *LC&G 2016 Communities dataset*.

Finally, the fifth dataset contains a collectionof 8,824,224 random tweets posted by 2,519,117 distinct users from September 12 to September 23, 2016, obtained via the Twitter Sample API.[10] This Twitter API delivers a random 1% sample of all tweets produced worldwide. As such, unlike the rest of our datasets, this one is not related to LC&G or to the city of Lucca. The rationale of this additional data collection is to have a control dataset. We did not refer to specific hashtags when collecting this fifth dataset. We refer to this dataset as *Contrastive dataset*.

Table 1 shows the overall statistics about our datasets. In particular, Table 2 shows the total number of tweets, both in Italian and in English, the number of relevant tweets (i.e., the tweets related to either the city of Lucca, or the comic-con, or both), and the number of tweets that are related to the city and to the comic-con respectively. The number of tweets in the relevant, city and comic-con columns has been computed as described in Sects. 5.1 and 5.2. It is worth noting that the sum of the values in the last two columns can be higher than those in the relevant column because some tweets refer both to the city and the comic-con.

---

[10] https://developer.twitter.com/en/docs/tweets/sample-realtime/overview/GET_statuse_sample.

By combining several different data collection strategies (e.g., topic-based, location-based, user-based) enacted before, during and after the event, recent state-of-the-art crowdsensing techniques (Avvenuti et al. 2017), and techniques based on established practices in statistical inference (i.e., the construction of a contrastive dataset that we leverage to "weight" results obtained on the other datasets), we built a large, unique, and one of the most comprehensive datasets for studying the impact of festivals on the online image of the host city.

4.2 Dataset filtering

As shown in the previous section, the datasets for our analyses contain tweets retrieved with different methods, like a random sampling of user timelines (LC&G communities), a crawling based on geographical boundary restrictions (Geo), and a hashtag-based crawling, with temporal restrictions (LC&G generic). In the process of dataset building, tweets which were neither related to the city of Lucca nor to the festival might be collected. To remove the irrelevant tweets, we selected only the ones containing a set of commonly used hashtags related to the comic-con and to the city. In this step, we included all the hashtags with the term "lucca" (examples are #lucca, #luccacomics2016, #luccacng16) and the hashtags with the term "lcg".

4.3 City—comic-con classification

The methodology illustrated in this section aims at classifying the content of tweets in the Geo, Generic, and Communities datasets as related to the city or to the festival. The underlying research question is whether we can distinguish the content of tweets by category. The result of this analysis will serve as a building block of our study, as clarified in Sects. 4.4 and 4.5. The methodology is based on a supervised classification approach, in which a training set is first built, several classification algorithms are trained over the training set, and finally, the trained classifiers are applied to the dataset.

The hashtag-based filtering allows us to restrict the analysis to relevant tweets, whose content is related either to the city, or to the festival, or to both. However, it does not allow to distinguish the tweets about Lucca and those about the comic-con. Given the high number of available tweets, we leveraged a machine learning classifier to automatically determine the topic (i.e., city or comic-con) of each tweet. Working in a supervised fashion, the classifier was first trained over a known set of examples. Once trained, it assigned a label to unknown input data, determining their category. We trained two different binary classifiers, one for deciding if a tweet was about the city of Lucca and another for deciding if a tweet was related to the comic-con. The combined results of the two classifiers also opened up the possibility to assess whether a tweet was related to both topics.

In the following, we present the different steps used in our approach—cleaning, feature selection, and classification—for the creation of a supervised model from a set of training data. In the cleaning step, we applied transformations to the text of the tweets, by first removing some data that are not informative for our

Table 3 Optimization process: parameters and their values

Algorithm	Parameter	Possible values
Count vectorizer	N-gram range	(1, 1), (1, 2), (1, 3), (1, 4)
TD-IDF VEctorizer	N-gram range	(1, 1), (1, 2), (1, 3), (1, 4)
Support vector machine	KERNEL	linear, rbf
Support vector machine	Soft margin parameter C	0.1, 0.3, 0.5, 1, 5, 10, 20, 100
Multinomial naive Bayes	Bayes alpha	0.01, 0.05, 0.1, 0.2, 0.3, 0.5, 0.75, 1, 1.5, 2, 5, 10
Bernulli naive Bayes	Bayes alpha	0.01, 0.05, 0.1, 0.2, 0.3, 0.5, 0.75, 1, 1.5, 2, 5, 10

purpose (like URLs and the re-tweet indication) and then applying the Snowball stemming algorithm, as implemented in NLTK (Bird et al. 2009), to reduce the terms to their root form.

In the feature selection step, we considered sequences of n words (called n-grams) as features. For each tweet, we calculated a weight for every single word and for each n-gram, while varying n. The number of words considered for the n-gram is a fitting parameter. For the weight calculation, we used two different algorithms: (i) the count vectorizer that assigns as weight the number of occurrences of the word/sequence of words in the text; and (ii) the term frequency—inverse document frequency (TF-IDF) that assigns a weight directly proportional to the number of occurrences in the document and inversely proportional to the number of occurrences of the same term in the whole collection. The two algorithms are provided by the Scikit-Learn Python library (Pedregosa et al. 2011).

In the classification step, the extracted features were used to assign a category to each tweet. For this step, we considered several classification algorithms provided by Scikit-learn: support vector machines (SVM), multinomial naive Bayes and Bernulli naive Bayes.

Table 3 lists the algorithms considered for features extraction and those used for classification. The table also shows, for each algorithm, the considered parameters and their values. All possible combinations of such algorithms, parameters and values were tested on 1756 tweets in Italian and 1051 tweets in English, randomly selected from the LC&G 2016 Communities dataset. Such tweets were manually labeled, and they constituted the training set of our classifiers.

Specifically, we separated the labeled tweets in two sets, the training set with 9/10 of the tweets and the test set with the remaining tenth. It is well known that trained classifiers can be subject to "overfitting" being too specialized on the training dataset and unable to generalize the classification to new and unseen data (Hawkins 2004). To avoid overfitting, we cross-validated the classification models using a k-fold split with k = 10. The k-fold algorithm first separates the training set in k parts (Witten et al. 2011), then, it uses one part to validate the training performed with the remaining $k - 1$ parts; the operation is repeated k times, each time using a different part to perform the validation.

Table 4 Perfomance results in classifying city- and comicon-related tweets

Label	Language	Classifier	N-gram	Parameters	Precision
Lucca	IT	SVM	(1, 2)	C: 0.1—kernel: linear	0.895
Lucca	EN	Multinomial	(1, 2)	Alpha: 0.01	0.972
Comicon	IT	SVM	(1, 2)	C: 0.3—kernel: linear	0.863
Comicon	EN	Multinomial	(1, 3)	Alpha: 0.05	0.966

The performances of the best performing classifiers are shown in Table 4. Performances are measured in terms of the well-known precision metric, defined as follows:

$$precision = \frac{1}{k} \times \sum_{i=1}^{k} \frac{tp_i}{(fp_i + tp_i)},$$

where *tp* is the number of true positives and *fp* is the number of false positives. In our scenario, the true positives are the tweets correctly classified by the classifier (i.e., the tweets classified as concerned with the city (resp., the comic-con) that are indeed about the city (resp., the comic-con). Analogously, the false positives are the tweets classified as concerned with the city, which are instead concerned with the comic-con, and vice-versa.

Table 4 also shows, for each pair of label/language values, the combination of classifier algorithm, n-gram size and classifier parameters which provided the best results.

4.4 Polarity detection

We introduce the algorithms used throughout the paper to label the polarity of a tweet. The research question we aim at answering with the application of polarity detection techniques to the tweets in our datasets is whether we can evaluate the feeling expressed by Twitter users towards the city (as far as the city-related tweets are concerned) and towards the festival (as far as the comic-con-related tweets are concerned). To answer these questions, we apply state-of-the-art polarity detection algorithms to our datasets.

Following the approach outlined in Sect. 5.2, we classified our tweets as either related to the city or to the comic-con, and then we studied their polarity. Polarity detection is a natural language processing analysis that assigns a label indicating the sentiment expressed by a text (Cambria 2016; Liu 2012). For this study, we used the sentiment analysis algorithms set described in Dell'Orletta et al. (2014) and Cimino and Dell'Orletta (2016), which can be used to analyze tweets in English and Italian. We considered 3 possible outcomes of the analysis, indicating positive, negative or neutral content.

Table 5 Frequency of tweets about the city of Lucca, per dataset

Dataset	Language	Tweet frequency of tweets	Ratio
Geo	IT	0.046354	1643.25
LC&G 2016 generic	IT	0.025748	912.78
LC&G 2016 communities	IT	0.000565	20.03
Contrastive	IT	0.000028	1.00
Geo	EN	0.05663438	29,524.13
LC&G 2016 generic	EN	0.008737	4578.59
LC&G 2016 communities	EN	0.000059	30.98
Contrastive	EN	0.000002	1.00

4.5 Content analysis

A further analysis specifically focused on the tweets related to the city of Lucca was applied to Geo, Communities, and LC&G datasets. In particular, we analyzed the tweets' content, by extracting the hashtags from the text and counting both the number of tweets and the number of users that have used each hashtag to depict users' preferences and opinions about the city. A first outcome is that the most frequent hashtags in the Geo and Generic datasets do not convey any insightful information since most of them were related to geographical location (e.g., #lucca, #tuscany, #italy) or to the comic-con. Therefore, taking a different approach, we analyzed all the words contained in the tweets to identify those words which appeared more frequently in the city related tweets. For each dataset, we performed the following steps:

- Selection of the subset of tweets classified as city-related;
- Removal of URLs, indication to retweets, mentions and non-alphabetical or numerical characters;
- Stemming of each word (except for the hashtags);
- Removal of the stop-words, as defined in NLTK.Snowball (Bird et al. 2009);
- Selection of the most frequent words;
- Manual lemmatization of frequent words to the noun form.

5 Results

5.1 Opinions about the city

In this section, we show the outcome of the analysis carried out over the geo, generic and communities datasets, whose scope was to infer the opinions of specific, LC&G related, communities about the city of Lucca. We consider two

Table 6 Detected polarity for datasets and tweets

Topic	Dataset	Negative	Neutral	Positive
City	Geo	0.0000	0.8403	0.1597
	LC&G generic	0.0286	0.6000	0.3714
	LC&G communities	0.0326	0.6087	0.3587
Comicon	Geo	0.0128	0.7104	0.2768
	LC&G generic	0.014	0.5881	0.3979
	LC&G communities	0.0138	0.6277	0.3585

different features, namely (i) the relative number of tweets exchanged about the city and (ii) the polarity associated with such tweets.

We applied the filtering and classification techniques presented in Sect. 5.1 and 5.2 to identify tweets about the city and tweets about the comic-con.

Table 5 shows (i) the frequency of the tweets about the city, compared with the total number of tweets, per dataset, and (ii) the ratio between the frequency of such tweets for each dataset over the contrastive dataset. Remarkably, each of the comic-con-related datasets (i.e., geo, generic, and communities) shows a considerably higher frequency than the contrastive dataset. This suggests that people interested in the festival are also more interested in the city where the festival is located, with respect to what derives from a non specific Twitter dataset. Also, the highest frequency occurs in the Geo dataset, indicating that those who tweeted from within the city walls during the comic-con, had also a higher inclination to tweet about the city. It is also noteworthy that both the LC&G 2016 Geo and the Generic datasets present a high frequency of tweets concerned with the city, showing that people interested in the comic-con are also interested in the city.

Table 6 presents the results of our study of the polarity in the city- and comic-con-related tweets, conducted using the polarity detection techniques illustrated in Sect. 4.3. Regarding the tweets related to Lucca, the most frequent detected polarity is neutral. Tweets with no strong emotions associated, like news or announcements, fall within this categorization.

The same Table 6 presents the results of the polarity detection for tweets regarding the comic-con. The results appear similar to the previous ones about the city. The only exception is represented by the Geo dataset that presents close to a double share of "positive" tweets for the comic-con related tweets, that is equal to circa 28% in total.

Finally, by manually inspecting the Italian tweets classified as negative in the LC&G communities dataset, we discover that among the reasons for a negative polarity such issues are mentioned as the festival is finishing, visitors have spent all their budget, traffic jams and long queues to enter the exhibitions and expo areas.

5.2 Opinions of communities

Previous literature extensively studied the paradigm of information diffusion on social networks, pointing out how so-called influencers are able to easily spread their

viewpoints among their followers (Guille et al. 2013). In this section, we analyze our datasets to understand whether, and to what extent, the opinions of the influencers about the city match the opinions of their followers. The LC&G 2016 Communities dataset contains tweets of the communities that follow artists and exhibitors, participating in LC&G in 2016, whom we consider as influencers of their followers' communities.[11] In the Communities dataset, we consider the tweets with hashtags #Lucca and #LC&G. For each artist, we calculate the share of the positive, negative and neutral tweets.[12] Then, we select the tweets of their followers, grouped according to the polarity of the message too. As before, only tweets with hashtags #Lucca and #LC&G are considered.

We discover a correlation of 0.40 between the polarity of tweets written by artists and the ones written by their communities, considering both hashtags. We further investigate such correlation, limiting the analysis to comic-con tweets only (#LC&G) and to city tweets only (#Lucca). For the former, we obtain a very similar correlation value (0.41), while for the latter the correlation is 0.34. These results indicate an average similarity of opinions, regarding both the city and the comic-con, between the influencers (i.e., the artists) and their communities. We were not able to perform the same analysis for the exhibitors, because the data crawling took place months after the comic-con, and with the Twitter-imposed limit of 3,200 tweets retrievable from the timeline of an account, we could get only a few tweets of the exhibitors related to the city and/or the comic-con.

5.3 Content analysis

In this section, we analyze the text of the tweets about the city, to retrieve specific terms representing what people talk about. The procedure is presented in Sect. 4.4.

The top-thirty terms in the analyzed datasets present many similarities. First, there are specific words conveying information about geographical locations (e.g., Lucca, Tuscany, Italy). In addition, many words are related to the comic-con, like cosplay and comics. Interestingly, many terms referring to Lucca landmarks are also present, such as *mura*, which is the Italian word for "walls", and the names of popular ancient squares (e.g., Anfiteatro, San Pietro) and churches (e.g., San Michele, San Giovanni).

However, relevant differences regarding the most frequent words emerge between the analyzed datasets. The Geo and LC&G Generic datasets mostly contain words related to the activities of the comic-con and names of city landmarks, where such activities are performed. The LC&G Communities dataset, which contains tweets crawled from the timeline of the users, shows a different set of words, containing both hashtags related to generic tourism (e.g., #lovelucca, #visitlucca) or to other

---

[11] The list of influencers was publicly available, e.g., on the festival website.

[12] We found the negative comments to be a negligible percentage over the whole number of comments by the artists.

Fig. 1 Most frequent words for the investigated datasets

events taking place in the city, like the Lucca Summer Festival.[13] In Fig. 1, we report the outcome related to the 30 most frequent words in the investigated datasets.

5.4 Participation in Lucca Comics & Games

As introduced in Sect. 4, the Question dataset was collected by sending a tweet to people who were tweeting with the hashtag #luccacomics2016 during the comic-con dates. We asked them about their physical participation in the comic-con. Out of the 1354 sent messages, we received 531 replies and 28.06% of these stated that the user was not attending the festival. Albeit rough, this estimation of the number of people who follow the LC&G via Twitter without attending the comic-con, indicates the existence of a vibrant online community that shows a great interest toward Lucca Comics & Games, as to actively following—and talking about—it online, although not physically attending the event.

---

[13] https://summer-festival.com/.

5.5 Discussion

Our results provide insights into the relationships between the festival and the destination city through the lenses of Twitter. In our analysis, we used five different datasets, city—comic-con classification, and conducted a set of analyses related not only to the frequency but also to the content and polarity of tweets. The results lead to four main conclusions for our case study.

1. Primarily, in line with the expectations, there is a considerable correlation between the event and the city, as testified by a higher frequency of tweets related to the city in the datasets collected during the festival period, namely within the Geo, LC&G 2016 Generic and Communities datasets, and by a significantly higher ratio in comparison to the Contrastive dataset. Thus, we can conclude that the online visibility of the city has significantly increased during the festival period thanks to the event, which represents an opportunity to distribute comments and opinions about the city among a large number of people worldwide through Twitter, including those who did not participate in the event.

2. The majority of the tweets related to the city appear as neutral due to the high proportion of official news and announcements related to the event, indicating that, concerning the mention of the city, Twitter is mostly used by official and public actors for practical purposes related to the festival program rather than by users for sharing their experiences. Overall, the polarity analysis yields similar results for both the city-related and the comic-con-related tweets, with a very low negative polarity. Nevertheless, the difference between the polarity of city- and comic-con-related tweets in the Geo dataset implies that people tweeting from the city are more likely to give opinions and express feelings towards the festival than towards the city, during their presence at the event. Notably, the similarity between the city-related and comic-con-related tweets in the LC&G Generic and Communities datasets suggests a close relation between the festival experience and the portrayal of the city. Thus, we can conclude that there is a relation between the attendees' positive experiences with the festival and their positive attitude towards the destination city. Furthermore, the negative polarity within the LC&G Communities comic-con-related tweets dataset is not necessarily related to the bad experiences, like traffic jams or long queues, but it might be related to personal reasons, such as having spent all the budget for the event- or simply to the festival coming to an end (revealing, indeed, a positive experience of the attendee).

3. Contents of the city-related tweets appear to be similar throughout the different datasets; for instance, words related to the location, thematic scope and city landmarks, which are also festival venues, appear commonly. Nevertheless, the LC&G Communities dataset also contains city-related tourism hashtags, such as #lovelucca and #visitlucca, accommodation-related hashtags, such as #bedandbreakfast, as well as other city events, such as #collezionando and #luccasummer. These results demonstrate that, thanks to the event, a distinct relation is built between the LC&G communities and the host city. Furthermore, through the increased communication of the event and the city, conjointly, on Twitter, the city of Lucca not

only incorporates a more dynamic layer into its identity, being perceived among the biggest meeting points of comics and games community, but it also promotes among a wide public this enriched city image, which includes historic elements, such as *mura* (historic walls), *anfiteatro* (a public square in Lucca that has an elliptical shape since it was built on the second century Roman Amphitheater) and *duomo* (cathedral), and is highly linked to festivity, cultural and creative features, such as comics/*fumetto*, games, *mostra* (exhibition) and cosplay.

4. When considering the diffusion of opinions within the communities, in other words the relation between influencers and their followers, we notice a correlation between the polarity of posts by the comics and games artists and their followers' tweets. However, these correlation values are lower than expected. This may be explained by the fact that the followers' interest in the artist is related mainly to the artist herself and her works, and less to her actual presence in a specific event. Nevertheless, these results still hint at the possibilities and opportunities that events offer to destination cities. By organizing a key event for cultural and creative communities in a successful way, Lucca gathers influencers and, through them, attains the opportunity to promote both the city and the landmark event and to establish a distinct relation with a public that would not even have visited the city, were it not for the festival.

6 Conclusions

Festivals serve many goals: supporting creative industries dynamics, empowering local communities, articulating and cultivating local identities and social cohesion (Rutten 2006). They also shape and enrich the city image and communicate the destination city as an attractive ecosystem. The use of social media to understand and illustrate public opinion is starting to enter the mainstream media discourse (Anstead and O'Loughlin 2015) and provides an important source of big data, useful for understanding the relations between events and cities, which is of crucial importance for events and cultural tourism. Correspondingly, in this paper, we leveraged state-of-the-art methodologies and tools, applied to a large and comprehensive Twitter dataset, for studying the impact of festivals on the online image of the host city and for investigating, in particular, how opinions about festivals and their destination cities are diffused as well as whether—and to what extent—they are related to each other.

In the literature review, we have shown how previous studies demonstrate the benefits of acquiring new knowledge from social media in different areas. Correspondingly, this study applies this approach to an under-explored topic. It shows how different analyses of social media platforms, in our case Twitter, can provide an enriched understanding of the ways festival attendees, both physical and virtual, talk about the event and engage with the destination city through the festival experience. Furthermore, our research hints at the relations established between influencers' and communities' opinions.

Overall, our findings provide some evidence that big scale festivals can communicate destination cities as attractive ecosystems among a big public with diverse

profiles (from core communities to indirectly related audiences). In this respect, our empirical study gives insights both for event organizers and for destination managers, and it can be used for further comparative research and theoretical inquiries about event tourism. We also demonstrate the applicability of the chosen methodological approach for this under-explored topic.

Despite the above-mentioned findings, this study has inevitable limitations. The timespan covers only one festival edition and mainly focuses on the festival period. Expanding the investigation along a wider timespan would be a promising future work. Moreover, content and sentiment analysis provide an overview that is limited to capturing the intrinsic and exogenous factors triggering opinions and positive/negative feelings towards the event and the city. Positive sentiments towards the city might be caused by participating in an exhibition organized in a historic monument or by meeting the cosplayers community on the city walls. Furthermore, the approach adopted to perform the sentiment analysis does not quantify the degree of the expressed sentiment (e.g., it does not discriminate between high, medium, and low positive tweets) nor the different aspects that contribute to the formation of such aggregated sentiment. These factors deserve further investigation, which would require to expand the methodological approach over multiple steps. In turn, these additional aspects would open up even more favourable application scenarios of these analyses that can be used for developing decision support tools for urban and touristic management.

References

Abdelhaq H, Gertz M, Armiti A (2017) Efficient online extraction of keywords for localized events in twitter. GeoInformatica 21(2):365–388

Akcora CG, Demirbas M (2010) Twitter: roots, influence, applications. Technical Report, Department of Computing Science and Engineering, SUNY Buffalo

Anstead N, O'Loughlin B (2015) Social media analysis and public opinion: the 2010 UK general election. J Comput-Mediat Commun 20(2):204–220

Arnegger J, Herz M (2016) Economic and destination image impacts of mega-events in emerging tourist destinations. J Destin Mark Manag 5(2):76–85

Atefeh F, Khreich W (2015) A survey of techniques for event detection in twitter. Comput Intell 31(1):132–164

Avvenuti M, Cresci S, Marchetti A, Meletti C, Tesconi M (2016) Predictability or early warning: using social media in modern emergency response. IEEE Internet Comput 20(6):4–6

Avvenuti M, Bellomo S, Cresci S, La Polla MN, Tesconi M (2017) Hybrid crowdsensing: a novel paradigm to combine the strengths of opportunistic and participatory crowdsensing. In: Proceedings of the 26th international conference on World Wide Web companion. ACM, pp 1413–1421

Avvenuti M, Cresci S, Del Vigna F, Fagni T, Tesconi M (2018a) CrisMap: a big data crisis mapping system based on damage detection and geoparsing. Inf Syst Front 20(5):993–1011

Avvenuti M, Cresci S, Nizzoli L, Tesconi M (2018b) GSP (geo-semantic-parsing): geoparsing and geotagging with machine learning on top of linked data. In: European semantic web conference. Springer, Cham, pp 17–32

Bennett A, Taylor J, Woodward I (2014) The festivalization of culture. Ashgate, Burlington

Bird S, Klein E, Loper E (2009) Natural language processing with Python. O'Reilly, Sebastopol

Boo S, Busser JA (2006) The hierarchical influence of visitor characteristics on tourism destination images. J Travel Tour Mark 19(4):55–67

Brilhante IR, Macedo JA, Nardini FM, Perego R, Renso C (2015) On planning sightseeing tours with TripBuilder. Inf Process Manag 51(2):1–15

Burton S, Soboleva A (2011) Interactive or reactive? Marketing with twitter. J Consum Mark 28(7):491–499

Bustard JRT, Bolan P, Devine A, Hutchinson K (2019) The emerging smart event experience: an interpretative phenomenological analysis. Tour Rev 74(1):116–128

Cambria E (2016) Affective computing and sentiment analysis. IEEE Intell Syst 31(2):102–107

Chrona S, Bee C (2017) Right to public space and right to democracy: The role of social media in gezi park. Res Policy Turkey 2(1):49–61

Cimino A, Dell'Orletta F (2016) Tandem lstm-svm approach for sentiment analysis. In: Proceedings of the third Italian conference on computational linguistics (CLiC-it 2016) & fifth evaluation campaign of natural language processing and speech tools for Italian (EVALITA 2016).

Cresci S, D'Errico A, Gazzé D, Duca AL, Marchetti A, Tesconi M (2014) Towards a DBpedia of tourism: the case of Tourpedia. In: International semantic web conference, pp 129–132

Cresci S, Cimino A, Avvenuti M, Tesconi M, Dell'Orletta F (2018) Real-world witness detection in social media via hybrid crowdsensing. In: Twelfth international AAAI conference on web and social media

D'Andrea E, Ducange P, Lazzerini B, Marcelloni F (2015) Real-time detection of traffic from twitter stream analysis. IEEE Trans Intell Transp Syst 16(4):2269–2283

de Lira VM, Macdonald C, Ounis I, Perego R, Renso C, Times VC (2019) Event attendance classification in social media. Inf Process Manag 56(3):687–703

Dell'Orletta F, Cimino A, Cresci S, Tesconi M (2014) Linguistically-motivated and lexicon features for sentiment analysis of Italian tweets. In: Proceedings of the 4th evaluation campaign of natural language processing and speech tools for Italian (EVALITA 2014), pp 81–86

Dou W, Wang X, Ribarsky W, Zhou M (2012) Event detection in social media data. In: IEEE VisWeek workshop on interactive visual text analytics-task driven analytics of social media content, pp 971–980

Filo K, Lock D, Karg A (2015) Sport and social media research: a review. Sport Manag Rev 18(2):166–181

Florek M, Insch A (2011) When fit matters: leveraging destination and event image congruence. J Hosp Mark Manag 20(3–4):265–286

Folgado-Fernández JA, Hernández-Mogollón JM, Duarte P (2017) Destination image and loyalty development: the impact of tourists' food experiences at gastronomic events. Scand J Hosp Tour 17(1):92–110

Frederick EL, Burch LM, Blaszka M (2015) A shift in set: examining the presence of agenda setting on twitter during the 2012 London Olympics. Commun Sport 3(3):312–333

Ghazali RM, Cai L (2014) Social media sites in destination image formation. In: Tourism social media: transformations in identity, community and culture, vol 18

Golbeck J (2019) Dogs good, trump bad: the impact of social media content on sense of well-being. In: The 11th conference on web science. ACM, pp 125–129

Golbeck J, Robles C, Edmondson M, Turner K (2011) Predicting personality from twitter. In: The third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing, pp 149–156

Guille A, Hacid H, Favre C, Zighed DA (2013) Information diffusion in online social networks: a survey. SIGMOD Rec ACM 42(2):17–28

Hawkins DM (2004) The problem of overfitting. J Chem Inf Comput Sci 44(1):1–12

Highfield T (2013) Following the yellow jersey: tweeting the tour de France. In: Weller K, Bruns A, Burgess J, Mahrt M, Puschmann C (eds) Twitter and society. Peter Lang, New York, pp 249–261

Hudson S, Roth MS, Madden TJ, Hudson R (2015) The effects of social media on emotions, brand relationship quality, and word of mouth: an empirical study of music festival attendees. Tour Manag 47:68–76

Imran M, Castillo C, Diaz F, Vieweg S (2015) Processing social media messages in mass emergency: a survey. ACM Comput Surv (CSUR) 47(4):67

Jungherr A, Schoen H, Jürgens P (2016) The mediation of politics through twitter: an analysis of messages posted during the campaign for the German federal election 2013. J Comput Mediat Commun 21(1):50–68

Kim SE, Lee KY, Shin SI, Yang SB (2017) Effects of tourism information quality in social media on destination image formation: the case of Sina Weibo. Inf Manag 54(6):687–702

Kwak H, Chun H, Moon S (2011) Fragile online relationship: a first look at unfollow dynamics in twitter. In: CHI'11 proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 1091–1100

Lalicic L, Huertas A, Moreno A, Jabreel M (2019) Which emotional brand values do my followers want to hear about? An investigation of popular European tourist destinations. Inf Technol Tour 21(1):63–81

Lampos V, Cristianini N (2012) Nowcasting events from the social web with statistical learning. ACM Trans Intell Syst Technol (TIST) 3(4):72

Lee CK, Taylor T, Lee YK, Lee B (2005) The impact of a sport mega-event on destination image: the case of the 2002 FIFA World Cup Korea/Japan. Int J Hosp Tour Adm 6(3):27–45

Leung XY, Bai B, Stahura KA (2015) The marketing effectiveness of social media in the hotel industry: a comparison of Facebook and Twitter. J Hosp Tour Res 39(2):147–169

Li X, Kaplanidou K (2013) The impact of the 2008 Beijing Olympic Games on China's destination brand: a US-based examination. J Hosp Tour Res 37(2):237–261

Litvin SW, Goldsmith RE, Pan B (2008) Electronic word-of-mouth in hospitality and tourism management. Tourism Manag 29(3):458–468

Liu B (2012) Sentiment analysis and opinion mining. Morgan & Claypool Publishers, Traverse City

Lovejoy K, Saxton GD (2012) Information, community, and action: how nonprofit organizations use social media. J Comput Mediat Commun 17(3):337–353

Middleton SE, Middleton L, Modafferi S (2013) Real-time crisis mapping of natural disasters using social media. IEEE Intell Syst 29(2):9–17

Moscardo G (2007) Analyzing the role of festivals and events in regional development. Event Manag 11(1–2):23–32

Mourtada R, Salem F (2011) Civil movements: the impact of Facebook and Twitter. Arab Soc Media Rep 1(2):1–30

Mourtada R, Salem F, Al-Dabbagh M, Gargani G (2011) The role of social media in Arab women's empowerment. Dubai Dubai Sch Gov 1(26):2011

Muntean CI, Nardini FM, Silvestri F, Baraglia R (2015) On learning prediction models for tourists paths. ACM Trans Intell Syst Technol (TIST) 7(1):8

Murthy D, Gross A, Pensavalle A (2016) Urban social media demographics: an exploration of twitter use in major American cities. J Comput Mediat Commun 21(1):33–49

Neidhardt J, Rümmele N, Werthner H (2017) Predicting happiness: user interactions and sentiment analysis in an online travel forum. Inf Technol Tour 17:101. https://doi.org/10.1007/s40558-017-0079-2

Ninomiya H, Kaplanidou K, Hu W, Matsunaga K, Nagazumi J (2019) An examination of the relationship between destination image and marathon participants' behaviours. J Sport Tour 23:225–241

Ozdikis O, Oguztuzun H, Karagoz P (2017) A survey on location estimation techniques for events detected in Twitter. Knowl Inf Syst 52(2):291–339

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikitlearn: machine learning in Python. J Mach Learn Res 12:2825–2830

Petrović S, Osborne M, Lavrenko V (2010) Streaming first story detection with application to twitter. In: Human language technologies: the 2010 annual conference of the North American chapter of the association for computational linguistics. Association for Computational Linguistics, pp 181–189

Rane H, Salem S (2011) Social media, social movements and the diffusion of ideas in the Arab uprisings. J Int Commun 18(1):97–111

Reinhardt W, Ebner M, Beham G, Costa C (2009) How people are using twitter during conferences. In: Creativity and innovation competencies on the web. Proceedings of the 5th EduMedia, pp 145–156

Richards G, Wilson J (2004) The impact of cultural events on city image: Rotterdam, cultural capital of Europe 2001. Urban Stud 41(10):1931–1951

Risse T, Peters W, Senellart P, Maynard D (2014) Documenting contemporary society by preserving relevant information from twitter. In: Weller K, Bruns A, Burgess J, Mahrt M, Puschmann C (eds) Twitter and society. Peter Lang, New York, pp 207–219

Rowe M, Stankovic M (2012) Aligning tweets with events: automation via semantics. Semantic Web 3(2):115–130

Rutten P (2006) Culture and urban regeneration: cultural activities and creative industries. A driving force for urban regeneration. URBACT Culture Network

Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World Wide Web. ACM, pp 851–860

Sankaranarayanan J, Samet H, Teitler BE, Lieberman MD, Sperling J (2009) Twitterstand: news in tweets. In: GIS '09 proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems. ACM, pp 42–51

Sevin E (2013) Places going viral: Twitter usage patterns in destination marketing and place branding. J Place Manag Dev 6(3):227–239

Sevin HE (2014) Understanding cities through city brands: city branding as a social and semantic network. Cities 38:47–56

Sharif SP, Mura P (2019) Narratives on Facebook: the impact of user-generated content on visiting attitudes, visiting intention and perceptions of destination risk. Inf Technol Tour 21:139–163

Tham A, Croy G, Mair J (2013) Social media in destination choice: distinctive electronic word-of-mouth dimensions. J Travel Tour Mark 30(1–2):144–155

Tonga Uriarte Y, DeFillippi RJ, Riccaboni M, Catoni ML (2019a) Projects, institutional logics and institutional work practices: the case of the Lucca Comics & Games Festival. Int J Project Manag 37(2):318–330

Tonga Uriarte Y, Antognozzi T, Catoni ML (2019b) Investigating tourism impacts of festivals: an exploratory case study of a big scale comic-con. Event Manag 23(6):817–833. https://doi.org/10.3727/152599519X15506259855823

Tsakalidis A, Papadopoulos S, Cristea AI, Kompatsiaris Y (2015) Predicting elections for multiple countries using Twitter and polls. IEEE Intell Syst 30(2):10–17

Wilkinson SE, Basto MY, Perovic G, Lawrentschuk N, Murphy DG (2015) The social media revolution is changing the conference experience: analytics and trends from eight international meetings. BJU Int 115(5):839–846

Williams NL, Inversini A, Ferdinand N, Buhalis D (2017) Destination eWOM: a macro and meso network approach? Ann Tour Res 64:87–101

Witten I, Frank E, Hall M (2011) Data mining: practical machine learning tools and techniques. Morgan Kaufmann, Burlington

Xiang Z, Gretzel U (2010) Role of social media in online travel information search. Tourism Manag 31(2):179–188

Xu W, Liu L, Shang W (2017) Leveraging cross-media analytics to detect events and mine opinions for emergency management. Online Inf Rev 41(4):487–506

Yeoman I, Robertson M, Ali-Knight J, Drummond S, McMahon-Beattie U (2012) Festival and events management. Routledge, Abingdon

Zubiaga A (2019) Mining social media for newsgathering: a review. Online Soc Netw Media 13:100049