Analysing Cultural Events on Twitter

Brigitte Juanals, Jean-Luc Minel

Analysing Cultural Events on Twitter

Brigitte Juanals[1] and Jean-Luc Minel[2]

[1] IRSIC, Aix Marseille University, France,
Brigitte.Juanals@univ-amu.fr
[2] MoDyCo, University Paris Nanterre - CNRS, France
jean-luc.minel@u-paris10.fr

Abstract. In this paper, we first present a model to represent message flows and their contents on Twitter, then a model and an instrumented methodology to describe and analyze these flows and their distribution among the various stakeholders. The aim is to explore the engagement and interactions between different types of stakeholders. We apply our methodology and tools to the 12th edition of the cultural event "European Night of Museums" (NDM16).

Keywords: circulation of information, influence, instrumented methodology, social network, Twitter, European Night of Museums

1 Introduction

Since the 2000s social networks owned by American companies (Facebook, Twitter, Instagram, YouTube, etc.) have become very popular in the cultural field. As these social networks attract the general Internet audience, cultural institutions have incorporated into their communication strategies a strong presence in these digital spaces through the dissemination of contents (news, practical or cultural information, representations of works and associated information, etc.) and the development of interaction with their public. These evolutions go hand in hand with the development of cultural marketing. Lastly, the "eventualization" of culture, following the explosion of temporary exhibitions, has amplified a "shift of patrimonial institutions to the logic of working like streaming media". It is a fact that in France and in Europe, over the last decade, temporary cultural events have multiplied, in the form of public events or festivals - such as the "European Heritage Days", "European Night of Museums", "La Nuit de l'Archéologie", "Passeurs d'images". These events have now become recurrent, generating initiatives that contribute to cultural outreach in society as well as to the development of cultural tourism.

Our methodology is based on the analysis of these events and checked on the 12th edition of the cultural event 'European Night of Museums' (NDM16), which took place in the heritage institutions that were partners in the event and was extended to digital media on the website dedicated to the operation as well as on social networks (Twitter, Facebook, Instagram). The outline of this

paper is the following. First, in Section 2, we will present the specificities and the contribution of our approach and our methodology to collect and analyze flows of tweets. In Section 3, we describe the model designed then in Section 4 we discuss the results from quantitative and lexical analysis. Finally, we conclude in Section 5.

2 Related Work

Nowadays, museums and cultural institutions use social media as a means to communicate and promote their cultural activities, as well as to interact and engage with their visitors, the main use of social media by museums remaining information and promotion of activities such as exhibition openings or events [6]. A qualitative study on the ways that museums use Twitter in this perspective shows that they choose to link resources, engage the public with new social media tools and favor a two-way form of communication [14]. Villaespesa Cantalapiedra [17] carried out fieldwork including a series of interviews with museum professionals which showed that the term 'engagement', 'can be interpreted in a variety of ways (...): From fostering inspiration and creativity in the user, originating a change of behavior, increasing the user's knowledge, receiving interaction from the user in the form of a like or a comment,' [17]. In particular, when Langa [10] studied the building of a relationship that forty-eight museums engaged on Twitter with online users, she showed that its primary use was as a marketing tool (public relations, events announcements, fact of the day, etc.) and that it led to a lesser engagement and a low audience participation. As mentioned in [7], tweet analysis has led to a large number of studies in many do-mains such as ideology prediction in Information Sciences [4], natural disaster anticipation in Emergency [15] and tracking epidemic [13] while work in Social Sciences and Digital Humanities has developed tweet classifications [16]. However, few studies aim at classifying tweets according to communication classes. An exception worth mentioning is the work presented in Lovejoy and Saxton [11] in which the authors (Twitter users) analyze the global behavior of nonprofit organizations on Twitter based on three communication classes: Information, Community and Action classes. Recently, several studies on tweet classification have been carried out in NLP [9,1] but to the best of our knowledge, only [3] has classified cultural institutional tweets in communication categories based on NLP techniques.

3 Designing a Methodology

3.1 Designing a Model

So far, scientific studies on the forms of digital communication engaged in by cultural institutions and audiences have analyzed the practices of institutions and audiences as well as their uses of digital platforms. Our contribution aims to design a model for the circulation of message flows on a social network platform, taking the case of Twitter. We chose an inter-institution space that corresponds

both to a social network platform widely used by institutions and to a growing communication situation in the cultural field at the present time. This choice led us to focus on the category of cultural event programmed on a national scale. This enables us to explore how information circulates and is exchanged as well as the communication relations established between the different stakeholders present in an inter-institution space. Based on the analysis of message flows on the scale of the cultural event studied, we attempt to answer questions concerning the communicative practices of the stakeholders, such as for example, what is the current strategic usage of social media conducted by different types of stakeholders (not only museums or cultural institutions) during cultural events? We expect marketing and promotional messages to be present but we also inquire into initiatives fostering audience participation, providing cultural contents and favoring interaction with users. We will specifically examine the communication policies of cultural institutions during a cultural event on Twitter in order to assess whether they are part of their mission to democratize culture for a wide audience on this platform or whether it is rather a marketing campaign to promote a place or an event. A second question concerns the identification of passing accounts (see below). Finally, two periods of time were distinguished. A period before the event and a period during the event. This distinction is an empirical consequence of our studies on Museum Week 2014 et 2015 [3].

Terminology and Attributes. Our analysis focuses on Twitter messages (called tweets) sent by accounts of cultural institutions and by other institutional or non-institutional stakeholders who participated in a cultural event. A first step was to build a terminology to describe the objects studied according to three dimensions: the message, the stakeholders, and the forms of stakeholder participation. Concerning messages, we will call a message sent by a twitter account an "original tweet" and an original message sent by an account different from the issuing account a "retweet". The current Twitter API gives access to the original tweet (and its sending account) of a retweet. The generic term tweet includes "original tweet" and "retweet". Regarding stakeholder qualification, we distinguished Twitter accounts, accounts managed by institutions (called "organizational account", OA), and accounts managed by individuals (called "private account", PA). This distinction is based both on the official list of museums in France provided by the French open data website, and the description on the Twitter account provided by the Twitter API. The analysis of the description field is necessary because non-museum institutions such as the City of Paris (@Paris) participated in the Night of Museums. Analysis of the flows during the MuseumWeek 2014 and 2015, European Night of Museums 2016, Europeans Days Heritage 2016 events, led us to identify six attributes used to qualify accounts according to their modes of participation and one computed score was associated at each attribute (cf. Table 1). We used the terms "participant", "producer", "relayed", "relaying", "mentioned" and "passing". The attribute "participant" was assigned to an account if it produced at least one original tweet or retweet during of of the the two temporal periods of the event (before and during the event). The attribute "producer" was assigned to an account if

it produced at least one original tweet. The attribute "relayed" was assigned to an account if at least one of its tweets was retweeted or quoted . The "relaying" attribute was assigned to an account if the account retweeted or quoted at least one tweet. The "mentioned" attribute was assigned to an account if its Twitter account name was mentioned at least once in a tweet. The attribute "passing" was assigned to an account if it was both "relayed" and "relaying". Our hypothesis is that an account with a high passing score is a key influential user who actively participates at the circulation ofinformation. Consequently, we computed for each account a passing score that is the product between the number of accounts that this account retweeted and the number of accounts that retweeted it. The value of this index is not significant in itself; it simply provides a means of comparing accounts. Note that these attributes were calculated irrespective of the number of followers. From these six attributes, it is possible , to compare the behavior of several accounts (see Section 4.2). Several quantitative

Table 1. Score calculation method

Participant	0 (no participation in the considered period of time) or 1
Producer	Total amount of tweets and retweets
Relayed	Total number of accounts which relayed his/her original tweets
Relaying	Total number of accounts she/he relayed
Mentioned	Total number of mentions of the account in text tweets
Passing	Product of relayed score by relaying score

analyses can be carried out. Quantitative analyses focus on the observation of flows and aim to identify accounts that contribute to the circulation of information during the cultural event. For each attribute (see above) a ranking of the accounts is computed, keeping only the first 10 or 15 accounts in this ranking. This make it possible to order the accounts that produce the largest number of original tweets, which are the most retweeted, and so on. In order to build a clasifier, a classification analysis of the contents of the messages was carried out in three stages. First, a team composed of two linguists and 10 community managers of cultural institutions designed a model, that is to say, determined the classes in which to categorize the tweets, and the features used to assign a tweet to a class. Four classes were identified (Encouraging participation, Interacting with the community, Promoting the event and informing about it, Sharing experience). The features selected were semio-linguistic (mostly lexical, but also including punctuation marks, emoticons), tweet-specific features (for example, the presence / absence of hashtags in tweets) as well as metadata such as the identity of the account. In the second stage, a classifier was built based on a corpus of 1000 tweets annotated by hand by cultural experts according to the categories defined in the previous step [3]. The classifier is based on the Naive Bayes and SVM models, with unanimous vote. In a third stage, the classifier was applied to the corpus of tweets to categorize all the tweets. Results of the thor-

ough evaluation of the quality of the classifier carried out on the MuseumWeek 2015 campaign are detailed in [7]; the F-measure $F_{0.5}$ coefficient is 0.696.

3.2 Implementing the Methodology

We implemented our methodology by building a workflow (cf. Figure 1) based on the one hand, on open access tools like R[3] (statistical), TXM[4] (text mining), Gephi[5] (graphs visualization), Neo4j [6] (graphs mining), Scikit-learn[7] (machine learning), and Sphynx (faceted search engine)[8], and on the other hand, we developed some scripts python. All data are stored in a NoSql database [9], and the scripts are used to query this database and compute specific attributes.

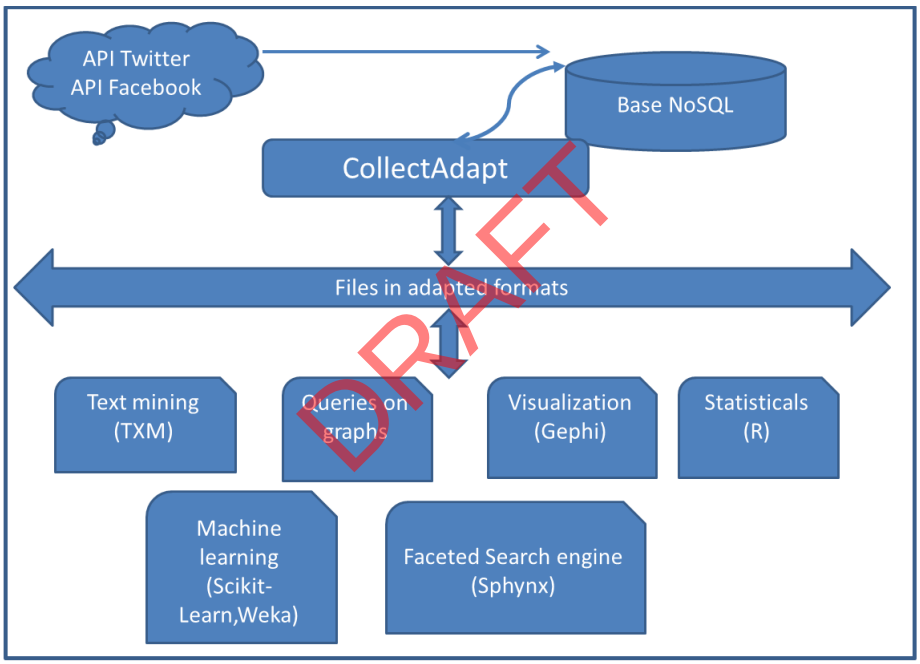

Fig. 1. Workflow and tools

[3] https://www.r-project.org/
[4] http://textometrie.ens-lyon.fr/
[5] https://gephi.org/
[6] https://neo4j.com/
[7] http://scikit-learn.org/stable/
[8] http://sphinxsearch.com/
[9] http://www.mongodb.com/

4 Applying the Methodology on the European Night of Museums Event

The analyses of the European Night of Museums (NDM16) were carried in the structuring content of the organizational and media framework in order to understand what happened during this cultural event, going beyond the display of quantitative data communicated at the time of its closure, i.e. 3000 events organized in France and in Europe, more than 2 million visitors who participated in the European Night of Museums in France.

The data acquisition stage consisted in harvesting tweets with only one hashtag, the official event hashtag #NDM16. We developed a script Python, based on Twarc (http://github.com/docnow/twarc) module proposed by Ed Summers using the streaming option of Twitter Application Programming Interface (API). In this paper, we limit the analysis to tweets in French sent during the week preceding the event and the day of the event, that is to say from 14 May to 21 May 2016 midnight. The main figures are the following: 11 264 tweets of which 3 301 original tweets (29%), 7 963 retweets (71%) sent by 4 012 participants. The specific figures are the following : more than half (56%) of organizational

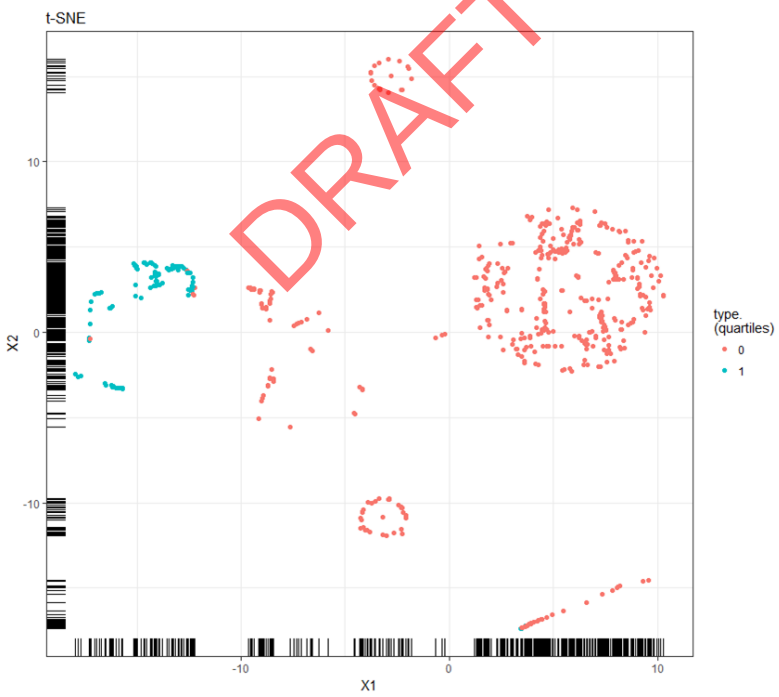

Fig. 2. t-SNE analysis

accounts and only 25% of private accounts produced original tweets. Participa-

tion in this event was largely limited to the action of retweeting (75% of tweets) the messages sent by the institutional partners. The findings of the supervised classification of original tweets sent during the event are shown in table 2. It should be noted that some tweets have been categorized twice (that is why the sum of the total percentages is greater than 100%). This table shows that the organizational accounts tweets mainly (93%) serve to promote and inform about local events planned for the occasion. This result is consistent with the organizational framework of the cultural event, the success of which is partly linked to its attendance. There is also some interaction with the public but messages of engagement are very sparse. We applied the t-SNE algorithm [12] to get a global cartography of the networks (fig 1.). Several communities are well distinguished. One community (in green) composed mainly of organizational accounts, which is more productive and interactive, and others, composed entirely of private accounts (in pink) which produced few tweets (see Section 4.3).

Table 2. Supervised classification

Classes	Percentage
Encouraging participation	4%
Interacting with the community	19%
Promoting the event and informing about it	93%
Sharing experience	0%
Not classifified	2%

4.1 Key Influential Users

Passing Accounts . We analysed the top ten passing accounts (see Section 3). A remarkable point is the non-correlation between their passing score and their number of followers. The Pearson-correlation computed on the these passing accounts (NuitdesMusées excluded) is equal to 0.15875, that is very low. Consequently, while the growth of followers of museums Twitter accounts is considered as a marketing target, the circulation of information during a cultural event is depending of key influential accounts which are not necessarily what is called 'big players' [5]. Some characteristics of these passing accounts must be pointed out. They are all organizational accounts with the exception of the account '@Mariette Escalier'. The owner of this private account is a professional in the field of cultural mediation. Ranked second after the organizational account of the event although she has less than 2 000 followers. In 5th place, the presence of the account of Alain Juppé, a French politician, mayor of the city of Bordeaux and former prime minister, shows that culture has become a political issue and plays an important role in marketing Cities and territories.

Great diversity was observed in the communicative practices of museum accounts . Some museums relay exclusively organizational accounts, while others

relay only private accounts. We also noted that the passing accounts during the event are not the same as the passing accounts of the period preceding the event. Only the 'NuitdesMusées' and 'MinistèreCultureCom" accounts are common to both these periods (see [8] for more details).

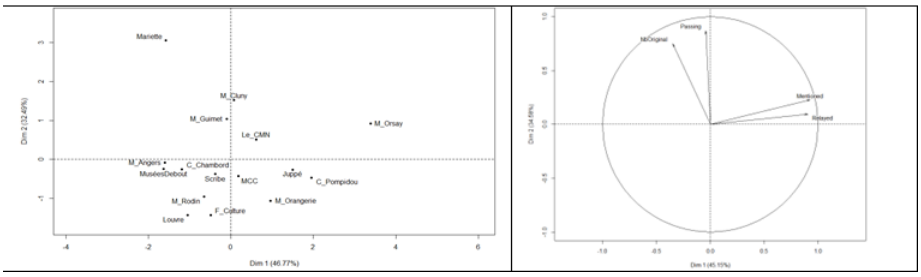

Fig. 3. Principal component analysis

In order to compare the 15 most active accounts, we performed a principal component analysis with 4 variables for each account: number of original tweets, mentioned score, relayed score and passing score (see Section 2). We did not take into account the official account of the event (@NuitdesMusees) which due to its official position, presents specific characteristics. This analysis shows the positions occupied by the 15 most active stakeholders in the 2016 NDM (Figure 1). Among them, the Musée d'Orsay is in a remarkable position, far on the right; it was mentioned and relayed many times because it organized an event (a jazz concert in the museum) which was very well attended. The success of this in situ event was relayed on Twitter which worked in this respect as a soundboard whereas the Louvre Museum, which organized conferences, was seldom mentioned or relayed despite its reputation. Similarly, the position of the private account Mariette Escalier (cultural mediator) occupies another remarkable position, in the left top corner; she is a passing account and she sent a lot of tweets, unlike most private accounts.

Integration Score. The aggregated communities around the passing account are represented in figure 2. This spatial representation realized by applying the Louvain algorithm [2] makes possible to highlight several points. The private account of the politician Alain Juppé who sought to enhance his cultural program was largely relayed by the collective who supports him (figure 3), but Alain Juppé account (bottom right) is very isolated from the rest of the network. In order to support this, we calculated a score called integration score. This score is equal to the ratio between the cardinality of the set of accounts that have retweeted the passing account, among others, and the cardinality of the set of accounts that have only retweeted the passing account. Thus, if a collective of a passing account is totally isolated from the graph of all accounts, its integration score will be equal to zero; Conversely, if a passing account is well integrated , in other words, the members of this group retweets other accounts, its integration

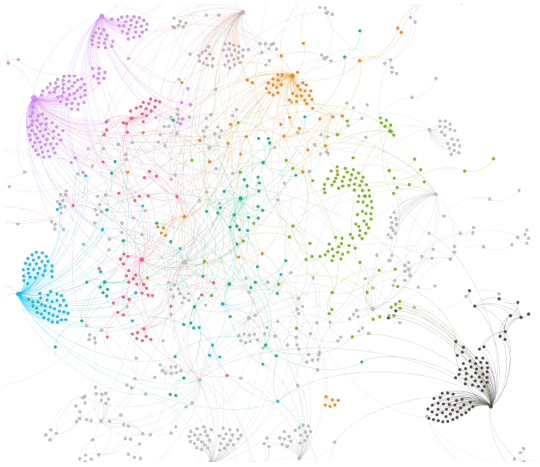

Fig. 4. Main communities

score will be equal to 1. For example, Alain Juppé integration score is equal to 0.36 while the score of Mariette Escalier is equal to 0.8. Other communities are grouped around museum accounts (on the top left) and are partially connected to the rest of the network. Nevertheless, it must pointed out that different communities did not interact. In other words, the Musée d'Orsay and the Center Pompidou, two Parisian museums share few common accounts, at least for the Night of Museums event. On the contrary, the community of the private account @Mariette Escalier (on the top right) is much more immersed in the network.

5 Conclusion

We designed and implemented a methodology to analyze forms of engagement, participation and relationship between cultural institutions, organizations and audiences. Checking this methodology on the cultural event "European Night of Museums" 2016, we first applied the t-SNE algorithm to get a global cartography of the networks. We performed a correspondence analysis that shows the specificity of one private account and the role of two museums in Paris, the Muséee d'Orsay and the Centre Pompidou, which played a major role in disseminating information. We intend to dig deeper into the instrumented methodology: more specifically, we are working on the extension of our method in order to take into account an incremental analysis of graphs. This conceptual work brings together research issues on the production and circulation of information with data mining and visualization software. Data mining is seen as an heuristic iterative and incremental process.

References

1. In: Shou-De, L., al. (eds.) 2nd Workshop on Natural Language Pro-cessing for Social Media. Association for Computational Linguistics and Dublin City University (2014)
2. Blondel, V., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics 2008(10) (2008)
3. Courtin, A., Juanals, B., Minel, J., de Saint Léger, M.: A tool-based methodology to analyze social network interactions in cultural fields: The use case museumweek. In: 6th Inter-national Conference on Social Informatics (SocInfo'14). pp. 144–156 (2014)
4. Djemili, S., Longhi, J., Marinica, C., Kotzinos, D., Sarfat, G.E.: What does twitter have to say about ideology? In: NLP 4 CMC: Natural Language Processing for Computer-Mediated Communication. pp. 16–25 (2014)
5. Espinos, A.: Museums on social media : Analyzing growth through case studies. In: Museums and the Web (2016)
6. Fletcher, A., Lee, M.: Current social media uses and evaluations in american museums. Museum Management and Curatorship (2012)
7. Foucault, N., Courtin, A.: Automatic classification of tweets for analyzing communication behavior museums. In: LREC 2016. pp. 3006–3013 (2017)
8. Juanals, B., Minel, J.: Information flow on digital social networks during a cultural event: Methodology and analysis of the european night of museums 2016 on twitter. SMS+Society Special Issue (2017)
9. Kothari, A., Magdy, W., Darwish, K., Mourad, A., Taei, A.: Detecting comments on news articles in microblogs. In: Kiciman, E., al. (eds.) 7th In-ternational Conference on Web and Social Media (ICWSM). The AAAI Press (2013)
10. Langa, L.: Does twitter help museums engage with visitors? In: iSchools IDEALS. pp. 484–495. University of Illinois, Aix-les-Thermes, France (2014)
11. Lovejoy, K., Saxton, G.D.: Information, community, and action: How nonprofit organizations use social media. Journal of Computer-Mediated Communication 17(3), 337–353 (2012)
12. Van der Maaten, L., Hinton, G.: Visualizing high-dimensional data using t-sne. Journal of Machine Learning Research 9, 2579–2605 (2008)
13. Missler, P., Romanovsky, A., Miu, T., Pal, A., Daniilakis, M., Garcia, A., Cedrim, D., Silva Sous, L.: Tacking dengue epidmeics using twitter content classification and topic modelling. In: Casteleyn S., D.P., C., P. (eds.) Current Trends in Web Engineer-ing. ICWE 2016 (SoWeMine). Springer, Cham (2016)
14. Osterman, M., Thirunarayanan, M., Ferris, E., Pabon, L., Paul, N., Berger, R.: Museums and twitter: An exploratory qualitative study of how museums use twitter for audience development and engagement. Journal of Educational Multimedia and Hypermedia 21(3), 241–255 (2012)
15. Sakaki, T., Okazaki, M., Matsuo, Y.: Tweet analysis for real-time event detection and earthquake, reporting system development. IEEE Transactions on Knowledge and Data Engineering 25(4), 919–931 (2013)
16. Shiri, A., Rathi, D.: Twitter content categorisation: A public library perspective. Journal of Information and Knowledge Management 12(4) (2013)
17. Villaespesa Cantalapiedra, H.: Measuring Social Media Success: The value of the Balanced Scorecard as a tool for evaluation and strategic management in museum. Ph.D. thesis, University of Leicester (2015)